

Universidade de São Paulo
Instituto de Física

Ensinando Máquinas a Reconstruir Matéria Escura no LHC

Carlos Hideo Yamaguchi

Orientador: Prof. Dr. Alexandre Alves

Dissertação de mestrado apresentada ao Instituto de Física
da Universidade de São Paulo, como requisito parcial para
a obtenção do título de Mestre em Ciências.

Banca Examinadora:

Prof. Dr. Alexandre Alves (Orientador) - UNIFESP

Prof. Dr. Alex Gomes Dias - UFABC

Prof. Dr. Thiago Tomei - IFT-UNESP



São Paulo

2022

FICHA CATALOGRÁFICA
Preparada pelo Serviço de Biblioteca e Informação
do Instituto de Física da Universidade de São Paulo

Yamaguchi, Carlos Hideo

Ensinando máquinas a reconstruir matéria escura no LHC.
São Paulo, 2022.

Dissertação (Mestrado) – Universidade de São Paulo. Instituto de Física. Depto. de Física Matemática.

Orientador: Prof. Dr. Alexandre Alves

Área de Concentração: Física das Partículas Elementares e Campos

Unitermos: 1. Física de partículas; 2. Aprendizado de máquina; 3. Redes neurais; 4. Matéria escura.

USP/IF/SBI-042/2022

University of São Paulo
Institute of Physics

Teaching Machines to determine dark matter in the LHC

Carlos Hideo Yamaguchi

Supervisor: Prof. Dr. Alexandre Alves

Dissertation submitted to the Physics Institute of the University of São Paulo in partial fulfillment of the requirements for the degree of Master of Science.

Examining Committee:

Prof. Dr. Alexandre Alves (Supervisor) - UNIFESP

Prof. Dr. Alex Gomes Dias - UFABC

Prof. Dr. Thiago Tomei - IFT-UNESP

São Paulo

2022

Resumo

Neutrinos, matéria escura, e partículas neutras de longas vidas médias atravessam os detetores despercebidos, carregando informação importante sobre as partículas pais e fontes de interação necessárias para reconstruir variáveis chaves como picos de ressonância em distribuições de massa invariante. Nesta dissertação, mostramos que um algoritmo de regressão de k -vizinhos mais próximos combinado com um classificador de rede neural profunda, um k NNNN, conseguem recuperar com precisão as distribuições de massa de pares W^+W^- totalmente leptônicos de um novo bóson de Higgs pesado além dos seus backgrounds do modelo padrão a partir de informações observáveis disponíveis em detetores. Também aplicamos o k NNNN para o decaimento de um par de sléptons em léptons mais neutralinos, um candidato a matéria escura, incluindo os backgrounds do modelo padrão. O resultado da regressão pode ser utilizado para treinar classificadores ainda mais poderosos em separar sinais e backgrounds no canal totalmente leptônico e garantir a seleção de bósons de Higgs e sléptons com significância estatística melhorada. O método assume conhecimento prévio da classe dos eventos e parâmetros do modelo sendo, assim, adequado para estudos pós-descoberta.

PALAVRAS-CHAVE: Fenomenologia de partículas; Aprendizado de máquina; Redes Neurais; k -Vizinhos mais próximos

Abstract

Neutrinos, dark matter, and long-lived neutral particles traverse the particle detectors unnoticed, carrying away information about their parent particles and interaction sources needed to reconstruct key variables like resonance peaks in invariant mass distributions. In this dissertation, we show that a k -nearest neighbors regressor algorithm combined with deep neural network classifiers, a k NNNN, is able to accurately recover distributions of the fully leptonic W^+W^- mass of a new heavy Higgs boson and its Standard Model backgrounds from the observable detector level information at disposal. We also applied the k NNNN to the decay of slepton pairs in lepton plus neutralino, a candidate for dark matter, along with its Standard Model backgrounds. The output of the regressor can be used to train even stronger classifiers to separate signals and backgrounds in the fully leptonic case and guarantee the selection of on-mass-shell Higgs bosons and sleptons with enhanced statistical significance. The method assumes previous knowledge of the event classes and model parameters, thus suitable for post-discovery studies.

KEYWORDS: Particle Phenomenology; Machine Learning; Neural network; k-Nearest Neighbors

Sumário

1	Introdução	5
2	Modelo Padrão	9
2.1	Setor eletrofraco e setor da QCD	11
2.2	Setor de Higgs e a quebra espontânea de simetria	12
2.2.1	Massa dos bósons	15
2.3	Setor de Yukawa e interações	17
2.3.1	Interações	18
3	Matéria Escura	21
3.1	Detecção Direta de matéria escura	22
3.2	Detecção Indireta de matéria escura	25
3.3	Busca de matéria escura em aceleradores de partículas	27
4	Aprendizado de Máquina	30
4.1	Detalhes da regressão k NN	37
4.2	Redes Neurais	39
4.2.1	Função de Ativação	41
4.2.2	Função perda	43
4.2.3	O Gradiente Descendente	45
4.2.4	Regularização	48
5	Reconstrução de ressonâncias totalmente leptônicas - Modelo Simplificado	50
5.1	Ajuste dos Regressores k NN	57
5.2	Classificação pré-regressão	62
5.3	Classificação pós-regressão	67

5.4	Melhorias na Significância Estatística do Sinal	75
6	Reconstrução de ressonâncias totalmente leptônicas - MSSM	81
6.1	Aplicação do k NNNN	82
7	Conclusão	94
A	Atualização dos pesos dado pelo backpropagation	96
B	Análise do sinal de bósons de Higgs pesado com largura diferente	98
C	Resultados extras	104
C.1	Distribuições Cinemáticas	104
C.2	Comparação entre possíveis aplicações do k NN	110
C.3	Hiperparâmetros do k NN	113
C.4	Distribuições da massa invariante para outros casos.	124

Capítulo 1

Introdução

Na procura da física além do Modelo Padrão (MP), a matéria escura tem particular interesse, pois explica diversas observações astrofísicas como, por exemplo, a curva de rotação de galáxias [1], a massa de um aglomerado de galáxias [2], a colisão de dois aglomerados de galáxias [3], lentes gravitacionais [4, 5] e a radiação cósmica de fundo em micro-ondas [6]. As partículas de matéria escura, contudo, devem interagir muito fracamente com as partículas que conhecemos do MP, podemos concluir isso a partir dos dados obtidos dos inúmeros experimentos já realizados [7, 8, 9, 10]. Diversos modelos físicos preveem partículas massivas estáveis fracamente interagentes como candidatos a matéria escura [11].

Dentre os vários experimentos já conduzidos para detectar a matéria escura de forma não-gravitacional, o Large Hadron Collider (LHC) [12] tem um grande potencial de descoberta. Nos planos de melhoria do LHC está previsto o aumento da energia de centro-de-massa para 14TeV em uma colisão próton-próton(pp) e o aumento da luminosidade integrada que, atualmente, é de 190 fb^{-1} , para 350 fb^{-1} até o final de 2024. O High-Luminosity LHC(HL-LHC), por fim, aumentará a luminosidade integrada para 3000 fb^{-1} até o final da operação do LHC[13, 14].

Os detetores do LHC como o ATLAS [15, 16] e o CMS [15, 17] podem detectar diversas partículas como fótons, elétrons, múons e hádrons, mas não partículas neutras fracamente interagentes como neutrinos e matéria escura. Isso é um problema na procura de nova física manifestando como estado escuro. A fuga dos neutrinos nos detetores previne a realização de observações importantes que poderiam beneficiar de poucos backgrounds. Por exemplo, a massa e largura do bóson de Higgs poderiam ser medidas com melhor precisão se a informação dos canais totalmente leptônicos $WW, ZZ \rightarrow \ell^+ \ell'^- \nu_l \nu_{l'}$, $l(l') = e, \mu$ fossem recuperáveis. Com exceção do canal $ZZ \rightarrow 4\ell$, contudo, dependemos dos canais semi-leptônicos ou totalmente

hadrônicos para realizar as medições com um número significativo de eventos de backgrounds. Identificar picos e limites na distribuição da massa invariante de observáveis e estados escuros ajudaria na busca de sinais de nova física, como bósons de Higgs pesados [18], produção de um par de Higgs com um Higgs invisível [19, 20], sleptons e charginos [21, 22], e decaimentos de novos gauge de bósons em neutrinos e/ou matéria escura a partir de seus backgrounds associados [23, 24], entre outras possibilidades. Outro exemplo importante onde o canal totalmente leptônico se beneficiaria de um ambiente limpo é a medição de ângulos de espalhamento dos bósons W e Z em estudos de polarização[25].

Em processos onde N_ν neutrinos são produzidos em espalhamento duro, existem $4N_\nu$ variáveis que devem ser recuperadas para reconstruir as partículas-pais. O momento transversal perdido, calculado do desbalanceamento da soma do momento transversal visível de objetos físicos reconstruídos, fornece duas restrições, apesar de não ser exatamente igual à soma do momento transversal dos neutrinos por causa de efeitos de detetores, contaminação de neutrinos e outras partículas perdidas de jatos hadrônicos, por exemplo. Restrições de massa devem providenciar informação complementar necessária para a reconstrução. O número de restrições de massa N_m , exceto $N_\nu : p_\nu^2 = 0$ de neutrinos, são dependentes do processo, e em muitos casos, não são o suficiente para recuperar o 4-momento dos neutrinos se $3N_\nu \geq N_m + 2$. Até em casos onde um número suficiente de restrições de massa existem, como sinais $t\bar{t}$ totalmente leptônicos [26, 27], a má reconstrução do momento transversal dos neutrinos, atribuição combinatória das partículas, e ambiguidades surgindo da natureza quadrática das equações não garantem uma solução para todos os eventos.

Em uma maneira independente do processo, uma abordagem para contornar a impossibilidade de recuperar o 4-momento de todas as partículas perdidas é criar variáveis cinemáticas e métodos que correlacionam com a informação perdida, por exemplo, a massa das partículas-pais. Muitos dessas variáveis são inteligentemente construídas para providenciar dicas úteis sobre as partículas em decaimento em diversas situações [28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38]. Contudo, nenhuma delas, por construção, conseguem recuperar o pico da ressonância.

Outra possível abordagem é usar um algoritmo de regressão para prever o 4-momento dos neutrinos ou alguma variável de interesse da informação observada. Uma maneira de realizar tarefas desse tipo é treinar um algoritmo para parametrizar uma função $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$, uma rede neural, por exemplo [39, 40, 41, 42]. Métodos de estimação de densidade [43] também podem ser úteis, existem outras aplicações e usos de algoritmos de regressão nas referências [44, 45,

46]. De fato, na referência [47], redes neurais regressoras foram utilizadas na estrutura do modelo padrão para reconstruir o ângulo polar do lépton definido no referencial de repouso do W em decaimentos leptônicos e semi-leptônicos num espalhamento de bósons vetoriais no LHC. As frações de bósons W polarizados transversal e longitudinalmente podem ser, então, inferidas das medições da distribuição do ângulo polar dos léptons. Uma dificuldade adicional no caso onde estamos interessados em detectar uma nova partícula, é a necessidade de inferir os parâmetros do sinal, principalmente sua massa, antes de treinar os regressores. Apesar disso, uma dedução foi conduzida em simultâneo, da regressão da variável cinemática usando algum modelo condicional, onde assumimos que foi realizada diretamente à partir dos dados.

Assumindo conhecimento prévio da ressonância do sinal, sua massa e, possivelmente, sua largura, a abordagem mais direta para reconstruir a variável de massa envolvendo neutrinos perdidos é interpolando um conjunto suporte de eventos simulados no lugar de ajustar parâmetros de alguma função que generaliza de conjunto treino a teste. Um algoritmo preciso e eficiente para uma regressão supervisionada é o algoritmo " k -nearest neighbors", conforme mostraremos. O problema dessa abordagem, como qualquer outro algoritmo supervisionado, é que precisamos conhecer o evento produzido nas colisões de antemão antes de selecionar o conjunto suporte correto para a interpolação da variável. A nossa abordagem usa o poder das redes neurais para classificar os eventos. A princípio, é possível identificar os sinais sem nenhum conhecimento prévio usando detecção de *outliers* e métodos não supervisionados. Entretanto, conforme discutido, sem conhecer os parâmetros de massa, reconstruir os picos é tarefa desafiadora, mas não impossível como mostrado na referência [48] onde a condição é que a distribuição marginal do alvo é conhecida.

Nesta dissertação mostramos como combinar redes neurais para classificação e k NN para regressão é útil para reconstruir um novo bóson de Higgs pesado decaindo em $W^+W^- \rightarrow l^+l'^- + \nu_l\nu_{l'}, l(l') = e, \nu$, um estado final totalmente leptônico com dois neutrinos escapando e os backgrounds principais do modelo padrão. Também reconstruímos o decaimento de um par de sleptons, com um estado final similar ao do Higgs pesado. Mostramos que a massa invariante predita dos léptons carregados e neutrinos pode ser usada de forma confiável como um novo atributo para limpar os backgrounds enquanto permite a seleção dos sinais, seja Higgs pesado ou sleptons.

A dissertação está organizada da seguinte forma. O capítulo 2 é uma revisão do modelo padrão. O capítulo 3 é uma introdução a matéria escura e as buscas sendo realizadas; o capítulo

4 descreve o básico de aprendizado de máquinas, na seção 4.1 detalhes do algoritmo k NN e na seção 4.2 detalhes de redes neurais; no capítulo 5 mostramos detalhes da combinação dos regressores e classificadores em identificar o bóson de Higgs pesado e os backgrounds principais do modelo padrão e na seção 5.4 resultados de melhorias na significância estatística do sinal, enquanto o capítulo 6 mostra a aplicação dos algoritmos e seus resultados para identificar um par de sleptons; o capítulo 7 é voltado para conclusões.

Capítulo 2

Modelo Padrão

O modelo padrão (SM) é uma teoria de gauge renormalizável que unifica as interações forte e eletrofraca baseado no grupo de simetria $SU(3)_C \times SU(2)_L \times U(1)_Y$, descrevendo a dinâmica de todas as partículas fundamentais conhecidas. Todas as partículas fundamentais conhecidas estão na figura 2.1. A figura também mostra propriedades como a massa, a carga elétrica e o spin das partículas.

A interação forte é descrito pelo grupo $SU(3)_C$ em que oito partículas de spin 1 são associadas aos glúons G_μ que são os bósons de gauge vetoriais. O índice C indica que os glúons carregam a carga "cor" e qualquer partícula que se transforma sob esse grupo carrega cor. A interação eletrofraca é descrito pelo grupo $SU(2)_L \times U(1)_Y$ onde os bósons de gauge vetoriais associados são o W_μ para o grupo $SU(2)_L$ e B_μ para o grupo $U(1)_Y$, o índice L indica que apenas os férmions de mão-esquerda se acoplam com os bósons enquanto o índice Y indica a hipercarga das partículas.

A lagrangiana do modelo padrão é

$$\mathcal{L}_{MP} = \mathcal{L}_{QCD} + \mathcal{L}_{EW} + \mathcal{L}_{Higgs} + \mathcal{L}_{Yukawa}. \quad (2.1)$$

Cada setor da lagrangiana será discutido nas próximas seções. Abaixo da escala eletrofraca a teoria sofre a quebra espontânea de simetria, onde detalhamos a quebra espontânea de simetria na seção 2.2. Por simplicidade escrevemos $e_{L/R} = (e, \mu, \tau)_{L/R}$, $\nu_{L/R} = (\nu_e, \nu_\mu, \nu_\tau)_{L/R}$, $u_{L/R} = (u, c, t)_{L/R}$ e $d_{L/R} = (d, s, b)_{L/R}$ para indicar a soma sobre as gerações fermiônicas.

Standard Model of Elementary Particles

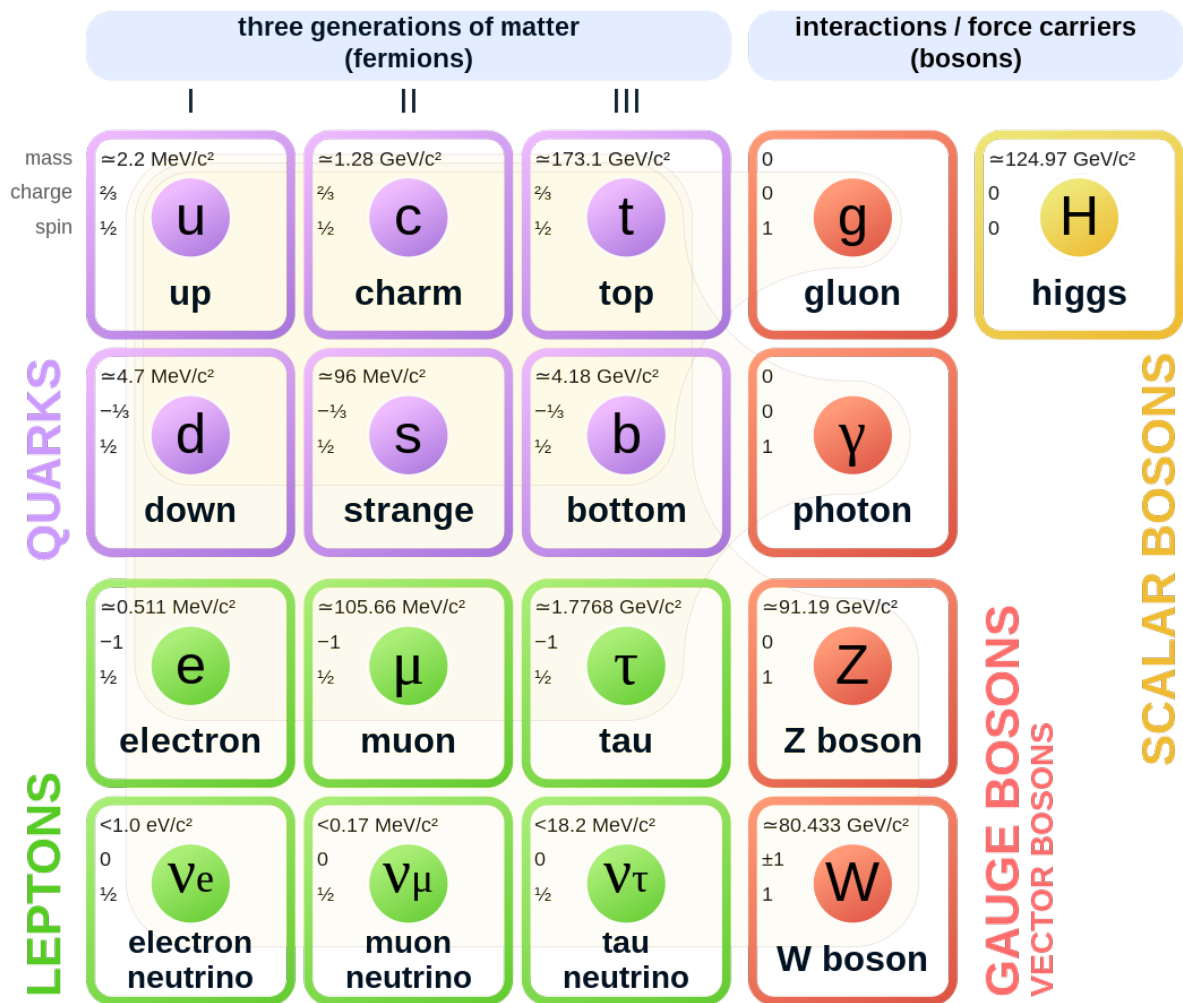


Figura 2.1: Conteúdo de partículas do modelo padrão. (Wikimedia Commons: MissMJ/CC BY 3.0)

2.1 Setor eletrofraco e setor da QCD

O setor eletrofraco é o setor responsável pela interação eletrofraca que é a descrição unificada da interação eletromagnética e fraca e é uma teoria de gauge de Yang-Mills. Esse setor descreve os bósons de gauge W^\pm , Z e o fóton A e os léptons representados pela cor verde na figura 2.1. Enquanto o setor da cromodinâmica quântica (QCD) é o setor responsável pela interação forte e é uma teoria de gauge de Yang-Mills. Esse setor descreve os glúons e quarks, representados pela cor roxa na figura 2.1, partículas fundamentais que compõe os hádrons como núcleo e nêutron.

A lagrangiana do setor eletrofraco mais da QCD é

$$\mathcal{L}_{EW+QCD} = \mathcal{L}_{EW} + \mathcal{L}_{QCD} = -\frac{1}{4}B_{\mu\nu}B^{\mu\nu} - \frac{1}{4}\text{tr}(\mathbf{W}_{\mu\nu}\mathbf{W}^{\mu\nu}) - \frac{1}{4}\text{tr}(\mathbf{G}_{\mu\nu}\mathbf{G}^{\mu\nu}) \quad (2.2)$$

$$+ i(\bar{\nu}_L, \bar{e}_L)\gamma^\mu D_\mu \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} + i\bar{e}_R\gamma^\mu D_\mu e_R + i\bar{\nu}_R\gamma^\mu D_\mu \nu_R + h.c., \quad (2.3)$$

$$+ i(\bar{u}_L, \bar{d}_L)\gamma^\mu D_\mu \begin{pmatrix} u_L \\ d_L \end{pmatrix} + i\bar{u}_R\gamma^\mu D_\mu u_R + i\bar{d}_R\gamma^\mu D_\mu d_R + h.c., \quad (2.4)$$

onde $B_{\mu\nu}$, $\mathbf{W}_{\mu\nu}$ e $\mathbf{G}_{\mu\nu}$ são os tensores de força de campo e D_μ é a derivada covariante dos dubletos e singletos fermiônicos. A equação 2.2 são os termos de gauge que contém os termos cinéticos e termos de interação entre os bósons, onde $B_{\mu\nu}$, $\mathbf{W}_{\mu\nu}$ e $\mathbf{G}_{\mu\nu}$ são definidos como

$$B_{\mu\nu} = \partial_\mu B_\nu - \partial_\nu B_\mu, \quad (2.5)$$

$$\mathbf{W}_{\mu\nu} = \partial_\mu \mathbf{W}_\nu - \partial_\nu \mathbf{W}_\mu + ig_2(\mathbf{W}_\mu \mathbf{W}_\nu - \mathbf{W}_\nu \mathbf{W}_\mu)/2, \quad (2.6)$$

$$\mathbf{G}_{\mu\nu} = \partial_\mu \mathbf{G}_\nu - \partial_\nu \mathbf{G}_\mu + ig(\mathbf{G}_\mu \mathbf{G}_\nu - \mathbf{G}_\nu \mathbf{G}_\mu). \quad (2.7)$$

As equações 2.3 e 2.4 são os termos dinâmicos dos létons e quarks, respectivamente, ou seja, contém os termos cinéticos dos férmions mais termos de acoplamento dos bósons de gauge com os férmions. A derivada covariante para os dubletos e singletos fermiônicos são

$$D_\mu \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} = \left[\partial_\mu - \frac{ig_1}{2} B_\mu + \frac{ig_2}{2} \mathbf{W}_\mu \right] \begin{pmatrix} \nu_L \\ e_L \end{pmatrix}, \quad (2.8)$$

$$D_\mu e_R = [\partial_\mu - ig_1 B_\mu] e_R, \quad (2.9)$$

$$D_\mu \nu_R = \partial_\mu \nu_R, \quad (2.10)$$

$$D_\mu \begin{pmatrix} u_L \\ d_L \end{pmatrix} = \left[\partial_\mu + \frac{ig_1}{6} B_\mu + \frac{ig_2}{2} \mathbf{W}_\mu + ig \mathbf{G}_\mu \right] \begin{pmatrix} u_L \\ d_L \end{pmatrix}, \quad (2.11)$$

$$D_\mu u_R = \left[\partial_\mu + \frac{i2g_1}{3} B_\mu + ig \mathbf{G}_\mu \right] u_R, \quad (2.12)$$

$$D_\mu d_R = \left[\partial_\mu - \frac{ig_1}{3} B_\mu + ig \mathbf{G}_\mu \right] d_R. \quad (2.13)$$

Nas derivadas covariantes o termo relacionado a B_μ é escrito como $-i\frac{g_1}{2} Y B_\mu$ onde Y é a hipercarga fraca, assim para cada dubleto e singlete fermiônico o termo é diferente. Em $SU(2)_L \times U(1)_Y$ termos de massa como $m\bar{\psi}\psi$, onde ψ são férmions, são proibidos, pois não respeitam a invariância de gauge. Assim, o setor de Higgs é responsável pela origem da massa dos bósons de gauge e a massa dos férmions são resultados da interação de Yukawa do bóson de Higgs com os férmions.

2.2 Setor de Higgs e a quebra espontânea de simetria

O mecanismo de Higgs é essencial para explicar a propriedade de massa dos bósons de gauge. No MP o mecanismo de Higgs é a quebra espontânea de simetria da simetria de gauge $SU(2)_L \times U(1)_Y$ que tem como consequência a origem da massa dos bósons de gauge W^\pm e Z . A quebra espontânea de simetria ocorre quando o campo de Higgs adquire um valor esperado do vácuo (VEV) diferente de zero. A lagrangiana do setor de Higgs é

$$\mathcal{L}_{Higgs} = (D_\mu \phi)^\dagger D^\mu \phi - V(\phi), \quad (2.14)$$

onde consideramos um campo escalar complexo ϕ com a seguinte representação

$$\phi = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix}, \quad (2.15)$$

e hipercarga fraca $Y(\phi) = 1/2$. A derivada covariante é

$$D_\mu = \left[\partial_\mu + \frac{ig_1}{2} B_\mu + \frac{ig_2}{2} \mathbf{W}_\mu \right]. \quad (2.16)$$

Para a teoria ser renormalizável e invariante sob $SU(2)_L \times U(1)_Y$ requer que o potencial seja

$$V(\phi) = -\mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \quad (2.17)$$

Se $\mu^2 < 0$ e $\lambda > 0$ não temos quebra de simetria ou um valor esperado do vácuo diferente de zero, pois $V(\phi) \geq 0$ e o mínimo está na origem e nenhum gerador é quebrado, portanto, não temos quebra de simetria. Agora se $\mu^2 > 0$ e $\lambda > 0$ existe a quebra espontânea de simetria, portanto ao minimizar obtemos

$$\left(\frac{dV}{d(\phi^\dagger \phi)} \right) \Big|_{\phi^\dagger \phi = \langle \phi^\dagger \phi \rangle} = 0 \quad (2.18)$$

$$\frac{d}{d\langle \phi^\dagger \phi \rangle} (-\mu^2 \langle \phi^\dagger \phi \rangle + \lambda \langle \phi^\dagger \phi \rangle^2) = 0 \quad (2.19)$$

$$-\mu^2 + 2\lambda \langle \phi^\dagger \phi \rangle = 0 \quad (2.20)$$

$$2\lambda \langle \phi^\dagger \phi \rangle = \mu^2 \quad (2.21)$$

$$\langle \phi^\dagger \phi \rangle = \frac{\mu^2}{2\lambda}. \quad (2.22)$$

Podemos definir $v = \sqrt{-\mu^2/\lambda}$, assim o VEV do campo escalar é

$$\langle \phi \rangle = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix}, \quad (2.23)$$

pois quebra $SU(2)_L \times U(1)_Y$ mas continua sendo invariante sob $U(1)_{EM}$ gerado pelo operador carga elétrica $Q = \frac{\sigma_3}{2} + Y$. Podemos verificar esse fato calculando o VEV que deve ser invariante, portanto pela transformação de gauge

$$\phi \rightarrow e^{ic\mathcal{G}} \phi, \quad (2.24)$$

onde \mathcal{G} é o gerador que torna o vácuo invariante, infinitesimalmente

$$(1 + ic\mathcal{G})\langle \phi \rangle = \langle \phi \rangle, \quad (2.25)$$

o que torna a condição da invariância como

$$\mathcal{G}\langle\phi\rangle = 0. \quad (2.26)$$

Agora realizamos o cálculo substituindo \mathcal{G} pelos geradores de $SU(2)_L$ e $U(1)_Y$, obtendo

$$\sigma_1\langle\phi\rangle = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} = \begin{pmatrix} v/\sqrt{2} \\ 0 \end{pmatrix} \neq 0, \quad (2.27)$$

$$\sigma_2\langle\phi\rangle = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} = \begin{pmatrix} -iv/\sqrt{2} \\ 0 \end{pmatrix} \neq 0, \quad (2.28)$$

$$\sigma_3\langle\phi\rangle = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} = \begin{pmatrix} 0 \\ -v/\sqrt{2} \end{pmatrix} \neq 0, \quad (2.29)$$

$$Y\langle\phi\rangle = 1\langle\phi\rangle = \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} \neq 0, \quad (2.30)$$

onde σ_i são as matrizes de Pauli os geradores de $SU(2)_L$. Agora utilizando Q como o gerador, obtemos

$$Q\langle\phi\rangle = \frac{1}{2}(\sigma_3 + Y)\langle\phi\rangle = \frac{1}{2} \begin{pmatrix} 1+Y & 0 \\ 0 & -1+Y \end{pmatrix} \begin{pmatrix} 0 \\ v/\sqrt{2} \end{pmatrix} = 0. \quad (2.31)$$

Como resultado obtemos três bósons de Nambu-Goldstone que serão componentes das massas dos bósons W^\pm , Z e o fóton não tem massa como consequência da invariância de VEV ao utilizar o operador Q . Podemos expandir a lagrangiana sobre o mínimo do potencial V , escrevendo

$$\phi = \exp\left(ic^a \frac{\sigma^a}{2}\right) \begin{pmatrix} 0 \\ (v+h)/\sqrt{2} \end{pmatrix}, \quad (2.32)$$

onde h é o campo do Higgs. Nessa parametrização tecnicamente deveríamos utilizar a combinação $K = \frac{1}{2}(\sigma_3 - Y)$ como o terceiro gerador, mas que é ortogonal a Q . Contudo, como Q deixa o VEV invariante e $\sigma_3 = K + Q$ o efeito em utilizar σ_3 ou K é o mesmo. Transformando para o gauge unitário

$$\phi \rightarrow \exp\left(i c^a \frac{\sigma^a}{2}\right) \phi = \begin{pmatrix} 0 \\ (v+h)/\sqrt{2} \end{pmatrix}, \quad (2.33)$$

podemos reescrever a parte escalar de \mathcal{L}_{Higgs} .

2.2.1 Massa dos bósons

Considerando o gauge unitário de $\phi = (0, v/\sqrt{2})^T$ a massa dos bósons são geradas pelo termo cinético do escalar de \mathcal{L}_{Higgs} ,

$$(D_\mu \phi)^\dagger D^\mu \phi = \frac{v^2}{8} [g_2^2 ((W_\mu^1)^2 + (W_\mu^2)^2) + (g_2 W_\mu^3 - g_1 B_\mu)] \quad (2.34)$$

$$- \mu^2 \frac{(v+h)^2}{2} - \lambda \frac{(v+h)^4}{4}. \quad (2.35)$$

O bóson vetorial W_μ^- e seu complexo conjugado são definidos como

$$W_\mu^\pm = \frac{1}{\sqrt{2}} (W_\mu^1 \mp i W_\mu^2), \quad (2.36)$$

e o termo com g_2^2 da equação 2.34 será

$$\frac{1}{2} \left(\frac{g_2 v}{2}\right)^2 W_\mu^\dagger W^\mu, \quad (2.37)$$

portanto, a massa do bóson W é

$$m_W = \frac{g_2 v}{2}. \quad (2.38)$$

O gauge de bóson neutro Z e do fóton A vem do termo $\frac{v^2}{8}(g_2 W_\mu^3 - g_1 B_\mu)$ da equação 2.34, após diagonalizar e introduzir a mistura entre W_μ^3 e B_μ , podemos escrever a lagrangiana na forma de matriz para obter

$$\frac{v^2}{8} (W_\mu^3, B_\mu) \begin{pmatrix} g_2^2 & g_1 g_2 \\ g_1 g_2 & g_1^2 \end{pmatrix} \begin{pmatrix} W^{3\mu} \\ B^\mu \end{pmatrix}, \quad (2.39)$$

onde a matriz de massa tem um dos autovalores iguais a 0 que corresponde ao fóton. Então obtemos Z_μ e A_μ em função de W_μ^3 e B_μ por uma rotação do ângulo de mistura θ_W ,

$$\begin{pmatrix} A_\mu \\ Z_\mu \end{pmatrix} = \begin{pmatrix} \cos \theta_W & \sin \theta_W \\ -\sin \theta_W & \cos \theta_W \end{pmatrix} \begin{pmatrix} B_\mu \\ W_\mu^3 \end{pmatrix} \rightarrow \begin{cases} Z_\mu = \cos \theta_W W_\mu^3 - \sin \theta_W B_\mu \\ A_\mu = \cos \theta_W B_\mu + \sin \theta_W W_\mu^3 \end{cases} . \quad (2.40)$$

Reescrevendo a equação 2.34 usando as definições 2.36 e 2.40, obtemos

$$m_W^2 W_\mu^+ W^{-\mu} + \frac{1}{2} (A_\mu, Z_\mu) \begin{pmatrix} 0 & 0 \\ 0 & m_Z^2 \end{pmatrix} \begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix}, \quad (2.41)$$

com

$$m_W = \frac{1}{2} g_2 v, \quad m_Z = \frac{1}{2} \sqrt{g_1^2 + g_2^2} v. \quad (2.42)$$

O ângulo de mistura da rotação θ_W é determinado por

$$\cos \theta_W = \frac{g_2}{\sqrt{g_1^2 + g_2^2}} = \frac{m_W}{m_Z}. \quad (2.43)$$

Utilizar a definição de A_μ com o campo do fóton que se acopla com a carga elétrica e do elétron nos termos dinâmicos dos léptons (equação 2.3) permite escrever a carga elétrica em função dos acoplamentos de gauge

$$e = \frac{g_1 g_2}{\sqrt{g_1^2 + g_2^2}}. \quad (2.44)$$

Por último, após a quebra espontânea de simetria, o potencial V é

$$V(\phi) = -\frac{\mu^4}{4\lambda} - \mu^2 h^2 + \lambda v h^3 + \frac{\lambda}{4} h^4 \quad (2.45)$$

Na equação 2.45 o terceiro e quarto termo representam as interações cúbicas e quaternárias do Higgs escalar. O segundo termo representa a massa (ao nível de árvore)

$$m_h = \sqrt{-2\mu^2} = \sqrt{2\lambda} v. \quad (2.46)$$

2.3 Setor de Yukawa e interações

A interação de Yukawa é a interação entre um campo escalar e um campo de Dirac. O setor de Yukawa contém os termos de massa dos férmions, no caso o acoplamento do Higgs com os férmions após a quebra espontânea de simetria. A lagrangiana de Yukawa é

$$\mathcal{L}_{Yukawa} = -\frac{\sqrt{2}}{v} \left[(\bar{\nu}_L, \bar{e}_L) \phi M^e e_R + \bar{e}_R \bar{M}^e \bar{\phi} \begin{pmatrix} \nu_L \\ e_L \end{pmatrix} \right], \quad (2.47)$$

$$- \frac{\sqrt{2}}{v} \left[(\bar{u}_L, \bar{d}_L) \phi M^d d_R + \bar{d}_R \bar{M}^d \bar{\phi} \begin{pmatrix} u_L \\ d_L \end{pmatrix} \right], \quad (2.48)$$

$$- \frac{\sqrt{2}}{v} \left[(-\bar{d}_L, \bar{u}_L) \phi^* M^u u_R + \bar{u}_R \bar{M}^u \phi^T \begin{pmatrix} -d_L \\ u_L \end{pmatrix} \right], \quad (2.49)$$

onde $M^i (i = e, u, d)$ são as matrizes de massa definidas como

$$M^e = \mathbf{U}_L^{e\dagger} \begin{pmatrix} m_e & 0 & 0 \\ 0 & m_\mu & 0 \\ 0 & 0 & m_\tau \end{pmatrix} \mathbf{U}_R^e, \quad (2.50)$$

$$M^u = \mathbf{U}_L^{u\dagger} \begin{pmatrix} m_u & 0 & 0 \\ 0 & m_c & 0 \\ 0 & 0 & m_t \end{pmatrix} \mathbf{U}_R^u, \quad (2.51)$$

$$M^d = \mathbf{U}_L^{d\dagger} \begin{pmatrix} m_d & 0 & 0 \\ 0 & m_s & 0 \\ 0 & 0 & m_b \end{pmatrix} \mathbf{U}_R^d, \quad (2.52)$$

onde \mathbf{U} são matrizes 3×3 unitárias e m_i são as massas das partículas definidas como

$$m_i = \frac{v}{\sqrt{2}} Y^i, \quad (2.53)$$

onde Y^i são os acoplamentos de Yukawa específicos para cada partícula que estão no espaço dos sabores.

No MP os neutrinos não têm massa, mas conhecemos que eles oscilam pelos experimentos

realizados [49, 50, 51] indicando que os neutrinos têm massa. Assim seria natural incluir um termo dedicado a massa dos neutrinos. Análogo à equação 2.49 o termo na lagrangiana seria

$$\mathcal{L}_{\text{neutrino-mass}} = -\frac{\sqrt{2}}{v} \left[(-\bar{e}_L, \bar{\nu}_L) \phi^* M^\nu \nu_R + \bar{\nu}_R \bar{M}^\nu \phi^T \begin{pmatrix} -e_L \\ \nu_L \end{pmatrix} \right], \quad (2.54)$$

onde M^ν seria

$$M^\nu = \mathbf{U}_L^{\nu\dagger} \begin{pmatrix} m_{\nu_e} & 0 & 0 \\ 0 & m_{\nu_\mu} & 0 \\ 0 & 0 & m_{\nu_\tau} \end{pmatrix} \mathbf{U}_R^\nu. \quad (2.55)$$

2.3.1 Interações

Existem diversas interações que o MP prevê como as interações cúbicas e quaternárias entre os bósons de gauge e também interações entre os bósons de gauge e os férmions. Começamos pelas interações cúbicas e quaternárias do bóson de Higgs que vem do potencial escrito na equação 2.45

$$\mathcal{L}_{HHH+HHHH} = -\lambda v h^3 - \frac{\lambda}{4} h^4. \quad (2.56)$$

As interações do bóson de Higgs com os bósons de gauge vem do termo da derivada covariante de ϕ (equações 2.34 e 2.35). Assim, a interação entre os bósons de gauge com o Higgs é

$$\mathcal{L}_{HVV} = h \left(\frac{g_2^2 v}{2} W_\mu^+ W^{\mu-} + \frac{g_2^2 v}{4 \cos \theta_W} Z_\mu Z^\mu \right), \quad (2.57)$$

$$\mathcal{L}_{HHVV} = h^2 \left(\frac{g_2^2}{4} W_\mu^+ W^{\mu-} + \frac{g_2^2}{8 \cos^2 \theta_W} Z_\mu Z^\mu \right), \quad (2.58)$$

onde V se refere a bóson de gauge vetorial, ou seja, W^\pm e Z .

As interações cúbicas e quaternárias entre os bósons de gauge W^\pm e Z vem dos termos $W_{\mu\nu} W^{\mu\nu}$ e $B_{\mu\nu} B^{\mu\nu}$ equação 2.2, assim o resultado da expansão desses termos é

$$\mathcal{L}_{WWA} = -ig_2 \sin \theta_W [(\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+)W^{-\mu}A^\nu + W_\mu^+W_\nu^-F^{\mu\nu} + h.c.], \quad (2.59)$$

$$\mathcal{L}_{WWZ} = -ig_2 \cos \theta_W [(\partial_\mu W_\nu^+ - \partial_\nu W_\mu^+)W^{-\mu}Z^\nu + W_\mu^+W_\nu^-Z^{\mu\nu} + h.c.], \quad (2.60)$$

$$\mathcal{L}_{WWWW} = -\frac{1}{2}g_2^2[(W_\mu^-W_\nu^+ - W_\nu^-W_\mu^+)W_\mu^+W_\nu^- + h.c.], \quad (2.61)$$

$$\mathcal{L}_{WWZZ} = -g_2^2 \cos^2 \theta_W (W_\mu^+Z_\nu - W_\nu^+Z_\mu)(W^{-\mu}Z^\nu - W^{-\nu}Z^\mu), \quad (2.62)$$

$$\mathcal{L}_{WWAA} = -g_2^2 \sin^2 \theta_W (W_\mu^+A_\nu - W_\nu^+A_\mu)(W^{-\mu}A^\nu - W^{-\nu}A^\mu), \quad (2.63)$$

$$\mathcal{L}_{WWAZ} = -g_2^2 \cos \theta_W \sin \theta_W (W_\mu^+A_\nu - W_\nu^+A_\mu)(W^{-\mu}Z^\nu - W^{-\nu}Z^\mu) + h.c., \quad (2.64)$$

onde $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ e $Z^{\mu\nu} = \partial^\mu Z^\nu - \partial^\nu Z^\mu$. A expansão do termo $G_{\mu\nu}G^{\mu\nu}$ é feito para obter os termos de autointeração dos glúons.

Agora a interação entre férmions e o bóson de Higgs vem da lagrangiana de Yukawa \mathcal{L}_{Yukawa} enquanto a interação entre os férmions e os bósons de gauge vem da derivada covariante dos férmions. A interação dos bósons W com os férmions, ou interação de corrente carregada é

$$\mathcal{L}_{CC} = -\frac{g_2}{2\sqrt{2}}(J_{CC}^\mu W_\mu^- + J_{CC}^{\mu\dagger} W_\mu^+), \quad (2.65)$$

onde J_{CC}^μ é conhecido como a corrente carregada fraca sendo definido como

$$J_{CC}^{\mu\dagger} = \bar{\nu}_L \gamma^\mu (1 - \gamma^5) e_L + u_L \gamma^\mu (1 - \gamma^5) V_{CKM}^\dagger d_L, \quad (2.66)$$

onde V_{CKM} é a matriz de Cabibbo-Kobayashi-Maskawa [52, 53] definido como

$$V_{CKM} = \mathbf{U}_L^u \mathbf{U}_L^{d\dagger}. \quad (2.67)$$

A matriz V_{CKM} tem como origem a incompatibilidade entre as transformações unitárias que relacionam os auto-estados fraco e de massa para os quarks. No caso dos léptons os neutrinos não têm massa e permitem uma rotação de modo que as matrizes unitárias se cancelem, mas se considerarmos a massa dos neutrinos teríamos uma matriz análoga à matriz V_{CKM} que seria chamado matriz de Pontecorvo–Maki–Nakagawa–Sakata [54] definida como

$$V_{PMNS} = \mathbf{U}_L^\nu \mathbf{U}_L^{e\dagger}. \quad (2.68)$$

A interação do bóson Z com os férmions, ou interação de corrente neutra é

$$\mathcal{L}_{NC} = -\frac{g_2}{2 \cos \theta_W} J_{NC}^\mu Z_\mu, \quad (2.69)$$

onde J_{NC}^μ é conhecido como a corrente neutra fraca sendo definido como

$$J_{NC}^\mu = \sum_{f=\ell,q} \bar{\psi}_f (v_f \gamma^\mu - a_f \gamma^\mu \gamma_5) \psi_f, \quad (2.70)$$

onde $v_f = T_3^f - 2Q_f \sin \theta_W$ e $a_f = T_3^f$ são os acoplamentos vetorial e pseudo vetorial (axial vector), respectivamente. T_3 e Q_f são o isospin fraco e a carga elétrica da partícula, respectivamente. As correntes neutras são diagonais no espaço dos sabores, portanto, não são afetadas pela mudança de base.

A interação do fóton com os férmions, ou interação eletromagnética é

$$\mathcal{L}_{em} = -e J_{em}^\mu A_\mu, \quad (2.71)$$

onde J_{em}^μ é a corrente eletromagnética definida como

$$J_{em}^\mu = \sum_{f=\ell,q} Q_f \bar{\psi}_f \gamma^\mu \psi_f. \quad (2.72)$$

Similar a corrente neutra, a corrente eletromagnética é diagonal na base dos sabores.

A interação entre o bóson de Higgs e glúons é a fusão de glúons que acontece por loops do quark top. A fusão de glúons é o método predominante para a produção do bóson de Higgs no LHC. A lagrangiana efetiva que descreve o acoplamento entre um ou dois bósons de Higgs com os glúons é obtido após integrar o quark top dado por

$$\mathcal{L}_{Higgs-Gluons} = \left(-\frac{h}{v} C_h + \frac{h^2}{2v^2} C_{hh} \right) tr(\mathbf{G}_{\mu\nu} \mathbf{G}^{\mu\nu}), \quad (2.73)$$

onde C_h e C_{hh} são os coeficientes de Wilson [55, 56] que contém as correções radiativas da QCD para os loops do quark top.

O MP experimentalmente é um grande sucesso, mas possui vários problemas como a massa dos neutrinos [57, 58], o problema da hierarquia do Higgs [59, 60] e o problema de violação CP na QCD [61], por exemplo. Um problema muito importante que o MP não consegue explicar é a matéria escura que é o foco do próximo capítulo.

Capítulo 3

Matéria Escura

Matéria escura é uma forma hipotética de matéria cuja existência vem de evidências observacionais como a curva de rotação de galáxias [1] e a massa de um aglomerado de galáxias [2], a colisão de dois aglomerados de galáxias [3], lentes gravitacionais [4, 5] e a radiação cósmica de fundo em micro-ondas [6], conforme descrito na introdução. Todas as evidências observacionais são de natureza gravitacional. Pelos dados obtidos de experimentos [7, 8, 9, 10] concluímos que devem ser fracamente interagentes com as partículas do MP, além de serem massivos e estáveis, pois tem longo tempo de vida. As partículas massivas fracamente interagente (WIMP) estáveis são candidatos a matéria escura, pois tem todas as propriedades desejadas para explicar as observações experimentais.

Uma das principais propriedades em relação à matéria escura é a sua velocidade térmica. Os termos “frio” e “quente” são utilizados para caracterizar a velocidade não relativística e relativística, respectivamente. Um exemplo de matéria escura quente são os neutrinos que são partículas estáveis, massivas (ao que tudo indica) e possuem velocidade relativística. O fato importante relacionado à matéria escura quente é que, sendo massivas e tendo velocidades relativísticas, tenderiam a espalhar grandes estruturas como as galáxias. Por outro lado, simulações de estruturas de grande escala [62] mostram um comportamento diferente do observado em experimentos [63, 64, 65] caso a matéria escura quente constituísse o total da matéria escura esperada do universo. Por esse motivo matéria escura quente pode constituir apenas uma pequena parte da matéria escura total.

Existem diversos modelos matéria escura fria massivas de interação fraca [11]. Alguns exemplos são modelos de pequeno Higgs com paridade T [66, 67] que resolvem o problema da hierarquia do Higgs. No caso os campos de Higgs são bóson de Goldstone da quebra de

simetria global na escala de energia TeV. O Higgs vira massivo pela quebra de simetria na escala eletrofraca, contudo a sua massa é protegida pela simetria global sendo livre de correções de loop na escala de energia TeV. O candidato a matéria escura é um WIMP que é a partícula mais leve com paridade T-ímpar que surge ao introduzir a paridade T.

Outro exemplo são modelos com dimensão extra universal (UED, *universal extra dimensions*) [68, 69, 70] que assumem uma ou mais dimensões espaciais extras. Nesses modelos partículas do MP podem propagar na nova dimensão e partículas com momento quantizado na nova dimensão aparecem como cópias pesadas de partículas do MP onde a partícula de Kaluza-Klein (KK) mais leve é estável com propriedades de um WIMP sendo candidato a matéria escura. Um último exemplo são modelos com áxions e partículas tipo-áxions permitem possível solução para o problema de violação CP na QCD sendo candidatas a matéria escura [61, 71].

Além de matéria escura fria existem outras possibilidades como matéria escura morna com propriedades intermediárias entre a matéria escura fria e quente [72], um exemplo são neutrinos estéreis que seriam neutrinos de mão direita que interagem apenas via gravitação [73, 74]. Também existem outras possibilidades para a matéria escura além de partículas como buracos negros primordiais [75, 76, 77] sendo buracos negros formados no período predominante radiativo pela extrema densidade de matéria presente durante a expansão inicial do universo antes da nucleossíntese do Big Bang e fluidos escuros com a hipótese que a matéria escura e a energia escura são a mesma entidade [78, 79, 80, 81].

Um grande número de experimentos permitem a busca de evidência das diversas propostas de matéria escura. Existem diversas categorias de busca de matéria escura como: (1) Detecção Direta, (2) Detecção Indireta e (3) Busca em aceleradores de partículas. A figura 3.1 mostra uma representação esquemática para as diferentes categorias de busca de matéria escura na forma de partículas (χ). Vamos descrevê-las em mais pormenores nas subseções à seguir.

3.1 Detecção Direta de matéria escura

A detecção direta da matéria escura é possível, pois halo galáctico deve possuir uma abundância de matéria escura que se estende até o sistema solar. Estudos mostram uma densidade de matéria escura no sistema solar da ordem de $\rho_0 = 0.4 \text{ GeV/cm}^3$ [82, 83, 84]. O halo galáctico pode ser considerado fixo em relação ao disco de rotação da galáxia, portanto existe uma velocidade relativa entre a matéria escura e à Terra que será a velocidade de rotação da Via-Láctea v_{rot} .

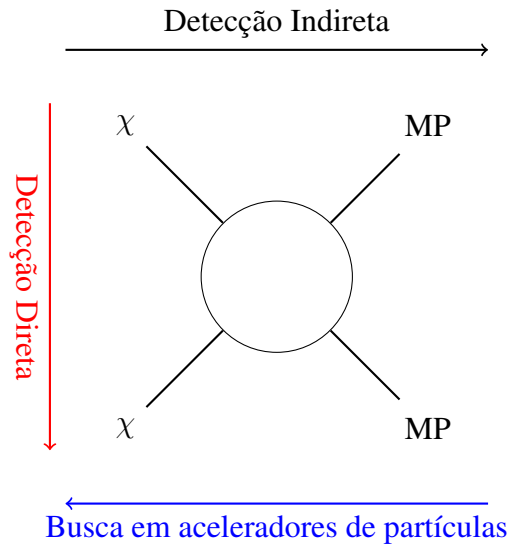


Figura 3.1: Representação esquemática das três categorias de buscas experimentais de partículas de matéria escura: Detecção direta (linha vermelha), Detecção Indireta (linha preta) e Busca em aceleradores de partículas (linha azul).

Assim uma quantidade significativa de matéria escura passa pela Terra permitindo a procura de interação de partículas de matéria escura construindo detetores com material específico que maximizam a probabilidade de interação.

A detecção direta de matéria escura envolve a interação de partículas de matéria escura com núcleons(ou elétrons) dos átomos presentes nos experimentos através de um espalhamento, conforme a linha vermelha da figura 3.1. A ideia é medir a energia do recuo dos núcleos para detectar a interação com a matéria e estimar massa da matéria escura e a seção de choque com os núcleos, σ . Existem diversos experimentos baseados em detetores com tecnologias de líquido nobre e criogênico. Por exemplo, a experiência XENON [85, 86] e PandaX [87] usam xenônio líquido e o CDMS [88] é um detetor de partículas criogênico. Um sumário de resultados presentes e futuros estão na figura 3.2 baseados na condição que a matéria escura é composta por apenas um tipo de partícula massiva de interação fraca. Os limites podem ser utilizados para impor restrições em modelos físicos.

Além das incertezas relacionadas ao modelo dos núcleos as análises são limitadas pelo conhecimento precário da densidade de matéria escura local e sua velocidade. A figura 3.3 mostra limites estabelecidos pelo experimento XENON1T para diferentes escolhas de densidade local de matéria escura e velocidade.

A sensibilidade de experimentos de detecção direta aumentam a cada geração, e eventualmente atingirá o limite onde espalhamentos de neutrinos ambientes com núcleos serão de-

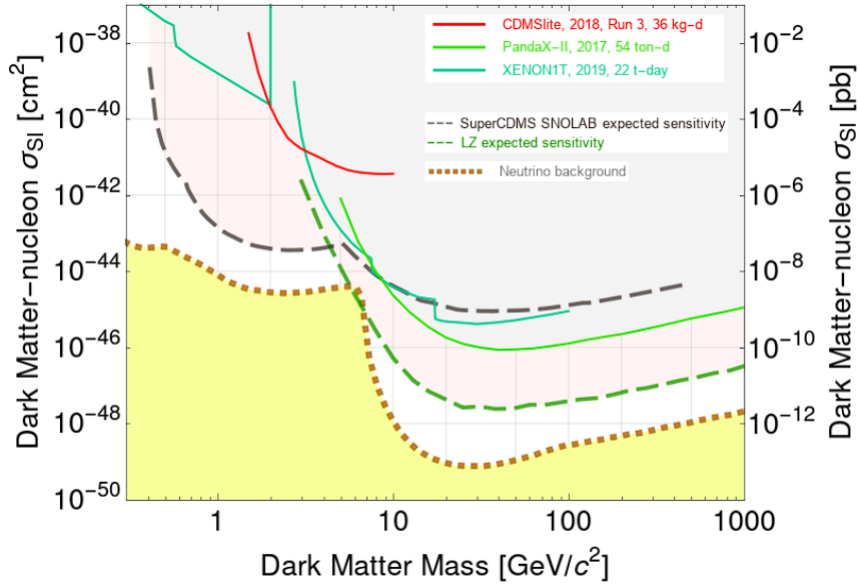


Figura 3.2: Limites de exclusão (linha sólida) dos experimentos de detecção de matéria escura no plano de seção de choque de espalhamento vs massa de matéria escura, para CDMSLite [88], PANDAX-II [87] e XENON1T [85, 86]. As linhas tracejadas correspondem a limites de futuros experimentos como SuperCDMS+SNOLAB e LZ. A região amarela representa o background de neutrinos cósmicos [89]. Figura 10 da referência [90].

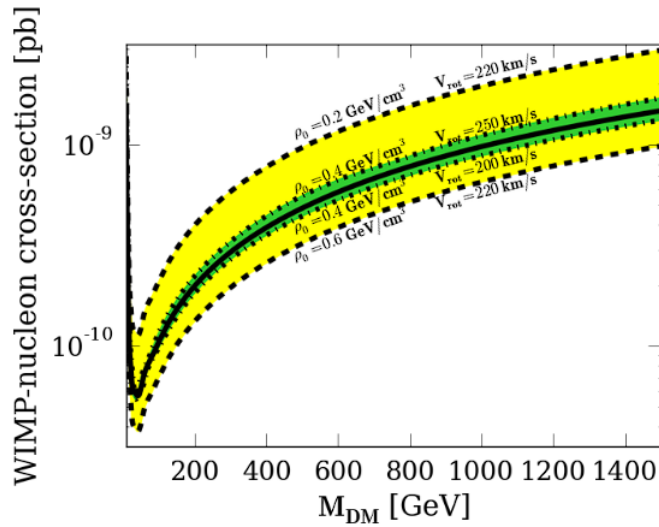


Figura 3.3: Limite superior do intervalo de confiança de 90% do experimento XENON1T da seção de choque de uma partícula massiva de interação fraca para $\rho_0 = 0.4 \text{ GeV/cm}^3$ e $v_{rot} = 220 \text{ km/s}$ (linha sólida preta). As incertezas desses valores são mostrados pela variação independente da densidade local de matéria escura (faixa amarela) e a velocidade do disco de rotação (faixa verde). Figura 3 da referência [91].

tectáveis. Apesar de interessante, esse background constituirá um problema difícil de contornar na busca de matéria escura por detecção direta.

Para partículas mais leves como áxions e partículas tipo-áxions os métodos de detecção descritos não são possíveis, pela energia de recuo ser muito pequeno, mas existem experimentos na busca de partículas tipo-axions que usam a interação do áxion com o campo eletromagnético. Neste caso, experimentos [71] usam uma cavidade eletromagnética fechada são usadas para detectar fótons produzidos na interação de áxions com o campo eletromagnético.

3.2 Detecção Indireta de matéria escura

Detecção indireta de matéria escura é baseada na produção de partículas do MP através da aniquilação ou decaimento da matéria escura detectados pelos experimentos. Os diferentes tipos de experimento de detecção indireta dependem das partículas procuradas. Os principais produtos de decaimento ou aniquilação são fótons, partículas carregadas e neutrinos. Fótons são fáceis de detectar e sua propagação não é afetada pelo meio interestelar e intergaláctico, mas a quantidade de background pode ser grande. Partículas carregadas são fortemente afetadas pelo meio interestelar e campos magnéticos e sua propagação é complexo de modelar. Os neutrinos têm uma propagação não afetada pelo meio interestelar com um baixo número de backgrounds, mas são difíceis de detectar.

Procurar por partículas produzidas indiretamente em regiões de matéria escura densa é importante, pois aumenta as chances de observar possíveis sinais. Existem diversos possíveis distribuições de matéria escura em galáxias, três perfis de densidades considerados são

$$\rho_{NFW}(r) = \rho_s \frac{r_s}{r} \left(1 + \frac{r}{r_s}\right)^{-2} \quad \text{NFW,} \quad (3.1)$$

$$\rho_{Burkert}(r) = \frac{\rho_s}{\left(1 + \frac{r}{r_s}\right) \left(1 + \left(\frac{r}{r_s}\right)^2\right)} \quad \text{Burkert,} \quad (3.2)$$

$$\rho_{Einasto}(r) = \rho_s \exp -\frac{2}{\alpha} \left[\left(\frac{r}{r_s}\right)^\alpha - 1 \right] \quad \text{Einasto,} \quad (3.3)$$

onde ρ é a densidade da galáxia, r é o raio da galáxia, ρ_s e r_s são parâmetros livres do halo da galáxia e α é um parâmetro que descreve o grau de curvatura do perfil da galáxia. Na Via Láctea o perfil NFW [92] possui parâmetros $r_s = 19.6$ kpc e $\rho_s = 0.32$ GeV/cm³ [93], o perfil

Burkert [94] possui parâmetros $r_s = 9.26$ kpc e $\rho_s = 1.57$ GeV/cm³ [83] e o perfil Einasto [95] possui parâmetros $r_s = 16.07$ kpc, $\rho_s = 0.11$ GeV/cm³ e $\alpha = 0.22$ [82]. Entre os três perfis, o perfil Einasto está de melhor acordo com simulações [96].

A figura 3.4 mostra os diferentes perfis de densidade para a Via-Láctea. Os diferentes perfis constituem um desafio na interpretação de resultados de detecção indireta, apesar de mostrarem uma densidade local de matéria escura equivalente, também mostram que a densidade de matéria escura no centro da galáxia está sujeito a uma grande incerteza.

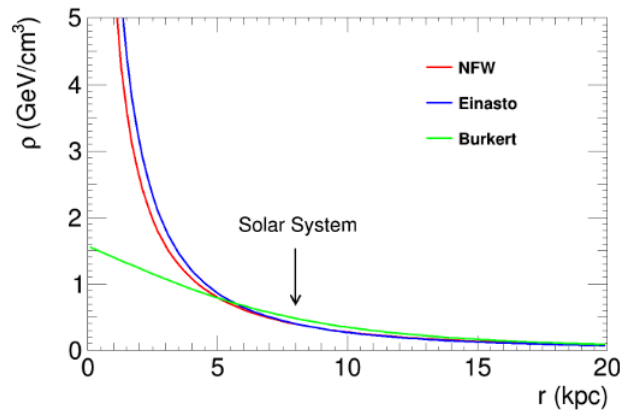


Figura 3.4: Densidade de matéria escura como função do raio a partir do centro da galáxia na Via-Láctea para os perfis NFW, Einasto e Burkert. Figura 12 da referência [90].

O estudo de detecção indireta requer entender como a matéria escura gera indiretamente partículas do MP. Consideramos a aniquilação de matéria escura, duas partículas de matéria escura podem interagir e aniquilar em partículas do MP, supondo que existe um pequeno acoplamento entre o setor de matéria escura e o setor do MP. A probabilidade de interação depende da velocidade relativa entre as partículas de matéria escura e a seção de choque de aniquilação em partículas do MP. Contudo, em muitos modelos as maiores seções de choque são relacionadas a partículas instáveis que conseqüentemente vão decair ou hadronizar. Se o produto final da aniquilação são partículas carregadas, também é necessário modelar sua propagação no meio interestelar. O decaimento é similar ao caso de aniquilação.

A detecção indireta é realizada através da procura de raios γ , de raios cósmicos e neutrinos. Exemplos de experimentos de procura de raios Gama são o HESS [97] e VERITAS [98] que fazem busca em centro de galáxias e galáxia anã. A procura de raios cósmicos é efetuada em raios cósmicos de pósitrons, antiprótons nos experimentos PAMELA [99, 100] e AMS-02 [101, 102]. A procura de neutrino é realizada em observatórios como o IceCube [103, 104]. Assim como os experimentos de detecção direta, experimentos de detecção indireta impõem restrições no

modelo físico. Por exemplo, a figura 3.5 mostra a dependência de restrições do experimento AMS-02 na escolha de perfis de matéria escura e modelos de propagação do antipróton. Observamos que a diferença de uma ordem de grandeza nas restrições são esperados, dependendo do perfil e do modelo de propagação escolhidos.

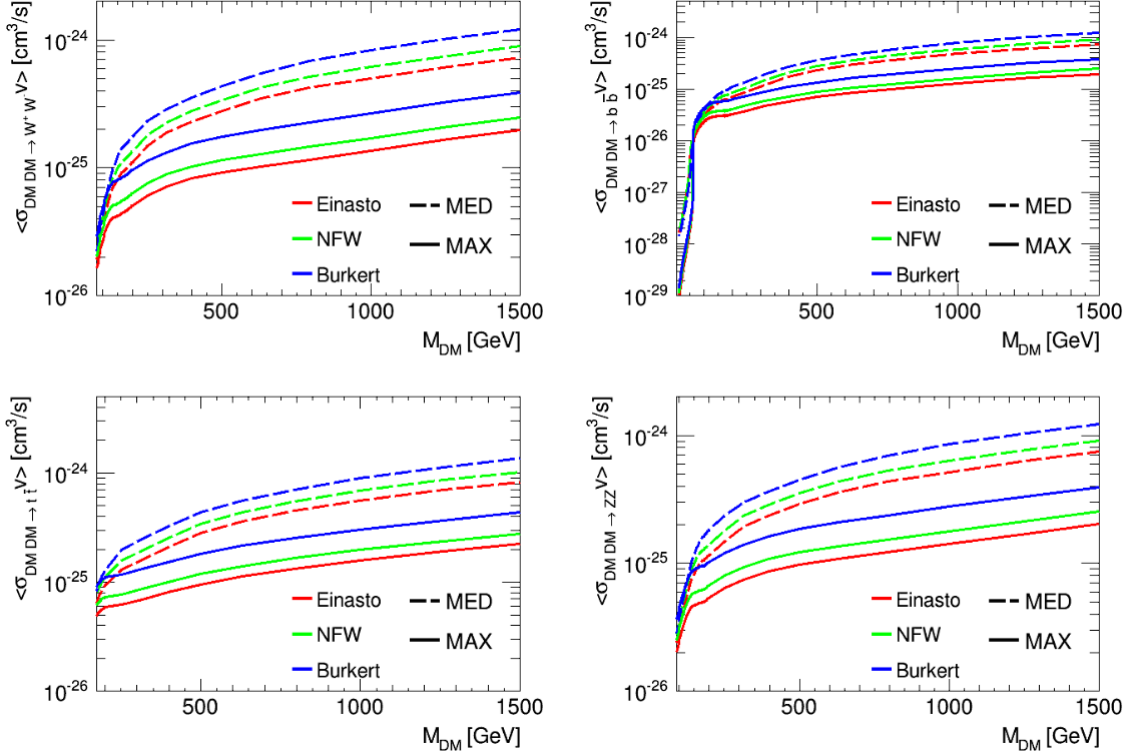


Figura 3.5: Limite superior do intervalo de confiança de 95% da seção de choque de aniquilação de matéria escura em W^+W^- , $b\bar{b}$, $t\bar{t}$ e ZZ , derivados dos dados da medição de antiprótons do experimento AMS-02 para diferentes perfis de matéria escura (Einasto, Burkert, NFW) e modelos de propagação (MED, MAX) [105, 106, 107]. Figura 1 da referência [91].

Como aniquilação e seção de choque de espalhamento são relacionados dentro de um modelo físico, a combinação de resultados de detecção direta e indireta pode definir restrições complementares em cenários de nova física, mas tais restrições estão sujeitas a grande incerteza astrofísica.

3.3 Busca de matéria escura em aceleradores de partículas

Cenários de nova física podem ser diretamente procuradas em aceleradores de partículas através da produção de novas partículas em colisores de altas energias de prótons ou elétrons e pósitrons. No contexto de matéria escura a busca de nova física é geralmente focado em estados finais com momento transversal perdido, portanto a busca de matéria escura só pode ser feita em modelos

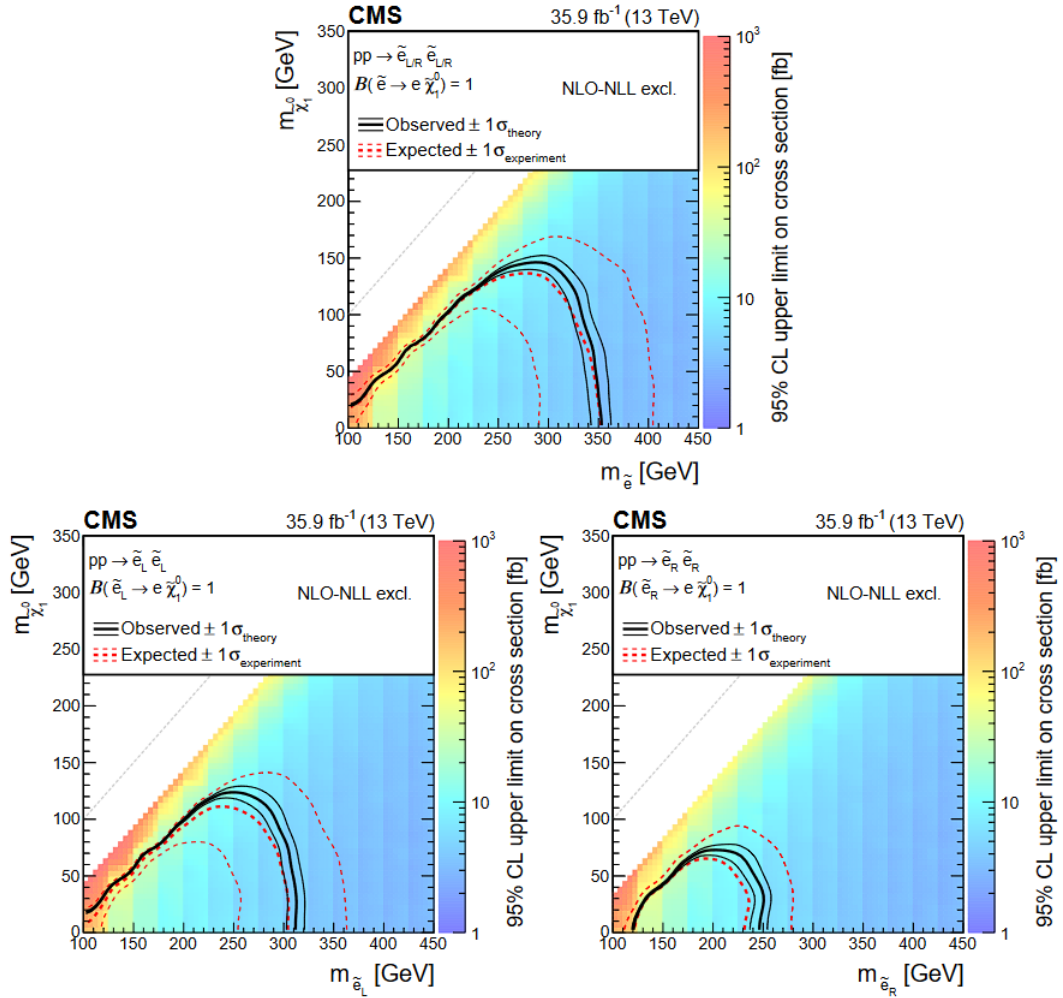


Figura 3.6: Limite superior e contorne de exclusão de intervalo de confiança de 95% para a produção de selétrons como função das massas de $\tilde{\chi}_1^0$ e \tilde{l} , assumindo uma produção de selétrons de mão esquerda e direita (linha superior), ou produção de selétrons de apenas mão esquerda (linha inferior esquerdo) ou mão direita (linha inferior direito). A região debaixo da linha espessa vermelha (sólida preta) é excluída pelo limite esperado (observado). As curvas de linha fina vermelha indicam a s regiões contendo 95% do limite da distribuição esperado sobre a hipótese de apenas background. Para selétrons de mão direita, apenas a linha esperada $+1\sigma$ (curva de linha fina vermelha) é mostrado porque nenhuma exclusão pode ser feita em -1σ . As curvas de linha fina preta mostram a mudança no limite observado devido à variação da seção de choque do sinal dentro de suas incertezas teóricas. Figura 6 da referência [21].

de matéria escura com produção associada porque a matéria escura faz parte da energia perdida. As partículas do MP associado em tais processos agem como marcadores do fenômeno de nova física, independente se for de matéria escura ou não.

As buscas de matéria escura em aceleradores são realizadas em canais de produção associada chamadas canal mono- X [108, 109, 110], onde X pode ser um jato, fóton, quark top, bóson de gauge, bóson de Higgs ou um lépton. Estudos no canal mono- X usam métodos estatísticos sofisticados como o likelihood ratio [111] e métodos estatísticos multivariados [112] e algoritmos de aprendizado de máquina. No LHC os algoritmos de aprendizado de máquina são utilizados na busca de nova física, incluindo a busca de matéria escura, nas colaborações experimentais [113, 114, 115, 116, 117].

Existem diversos modelos simples com possível matéria escura que podem estudados no LHC [110]. Enquanto modelos mais complexos resolvem problemas teóricos do MP além de possuírem candidatos a matéria escura, como o pequeno Higgs, UED, áxions e partículas tipo-áxion, introduzidos no início deste capítulo e o xSM [118, 119, 120] sendo uma extensão do MP por um singlete escalar real S (candidato a matéria escura) e de um Higgs pesado H_2 que vem do setor de Higgs estendido. Esse modelo é interessante, pois permite a existência de uma transição de fase electrofraca forte de primeira ordem condição necessária para explicar a assimetria de bárions no universo, se H_2 for do tipo-singlete [121].

Modelos supersimétricos [59, 60] são um dos principais modelos físicos complexos estudados, mais detalhes sobre esse tipo de modelo no capítulo 6. No contexto de matéria escura, o modelo supersimétrico mais simples o MSSM possui como candidato principal a matéria escura a partícula mais leve da sua teoria que é estável devido à conservação da paridade-R. No caso o neutralino mais leve é um dos candidatos a partícula mais leve do MSSM [122, 11], além de ser massivo de interação fraca, portanto um dos principais candidatos a matéria escura. Buscas de partículas supersimétricas no LHC podem ser vistos nas referências [21, 22, 123].

Capítulo 4

Aprendizado de Máquina

O aprendizado de máquina (ML, *machine learning*) é uma subárea da inteligência artificial cujo objetivo é desenvolver algoritmos que possam identificar padrões em massas de dados de qualquer natureza de modo a utilizá-los para predições de variáveis-alvo de interesse. Exemplos clássicos são a identificação de objetos em imagens [124, 125], o reconhecimento de voz [124], o diagnóstico clínico baseado em dados sobre o paciente [126]. No caso da fenomenologia de partículas em aceleradores, o exemplo mais comum de uso de ML é a identificação de um evento de colisão de altas energias baseado nos 4-momentos das partículas que passam por um detetor [39, 41, 40]. No nosso caso, essa massa de dados é representada por uma tabela onde cada coluna representa uma variável medida.

Quando se pretende identificar um tipo de evento, um rótulo, dentre um número finito de possibilidades, por exemplo, identificar um evento de sinal *versus* um evento de background, usamos algoritmos de classificação [127]. Quando pretendemos prever o valor de uma variável real, por exemplo, uma massa invariante de um conjunto de partículas, usamos um algoritmo de regressão [127]. As principais abordagens são: (1) aprendizado supervisionado e (2) não supervisionado.

O aprendizado não supervisionado usa um conjunto de dados não rotulado, relegando ao algoritmo a tarefa de encontrar uma estrutura, de descobrir similaridades e padrões dos dados. A vantagem é a capacidade de algoritmos de aprender informação sobre os dados sem nenhuma indicação formal de como é a estrutura dos dados. Exemplos clássicos destes tipos de algoritmo são o k-means [128], o DBSCAN [129] e as máquinas de Boltzmann [130]. As *Generative Adversarial Networks*, GANs [131] e os autoencoders variacionais, VAEs [132] também podem ser utilizados como detectores de outliers de maneira não-supervisionada. O k-means e DBS-

CAN são exemplos de algoritmos de análise de agrupamento de dados (*clustering* [133, 134]), enquanto máquinas de Boltzmann, GANs e VAEs são exemplos de algoritmos que usam redes neurais [124, 125, 135].

O aprendizado supervisionado, por sua vez, é a abordagem onde o conjunto de dados é rotulado, ou seja, o usuário fornece aos algoritmos informação sobre os eventos que são objetivo da classificação em uma fase de treinamento. O algoritmo visa aprender uma relação entre uma representação do conjunto de dados e os rótulos dos eventos. A relação aprendida pode ser utilizada em uma nova amostra para classificar e/ou prever sobre os dados.

Em algoritmos de classificação supervisionado o objetivo é classificar os eventos em diferentes classes. Os rótulos fornecidos pelo usuário representam as classes dos eventos. O algoritmo aprende uma relação da representação do conjunto de dados com os rótulos e irá prever os rótulos (as classes) em uma amostra chamada conjunto teste. Por exemplo, em uma classificação multi-classe com 4 classes, os rótulos fornecidos pelo usuário podem ser $\Omega = \{0, 1, 2, 3\}$ onde cada número representa uma classe, supondo um conjunto teste com N eventos, a saída da classificação será um vetor N -dimensional $y_{clf} \in \Omega$. Exemplos de algoritmos de classificação são as árvores de decisão, as Boosted Decision Trees (BDT) [136], muito usadas pelas Colaborações ATLAS e CMS, [137, 138], as florestas aleatórias [139] e as redes neurais [124, 125, 135].

Como dissemos, a principal diferença entre algoritmos de classificação e regressão são os rótulos atribuídos aos eventos. Nos algoritmos de classificação os rótulos são discretos e em algoritmos de regressão os rótulos são uma quantidade contínua relacionada ao problema. Por exemplo, podemos usar um algoritmo de regressão para prever uma distribuição, os rótulos serão valores da distribuição. Similarmente à classificação, o algoritmo de regressão aprende uma relação entre a representação do conjunto de dados com a distribuição e irá prever valores da distribuição em um conjunto teste. No caso da regressão o desempenho do algoritmo será verificado comparando a distribuição predita com a verdadeira. Exemplos de algoritmos de regressão são a regressão linear [140], k NN [141] e redes neurais [124, 125, 135].

O funcionamento de um algoritmo supervisionado é ilustrado na figura 4.1. Em ML é essencial o tratamento preliminar de dados para o processamento dos algoritmos. Por exemplo, a performance dos algoritmos melhora quando reduzimos a ordem de grandeza dos dados através

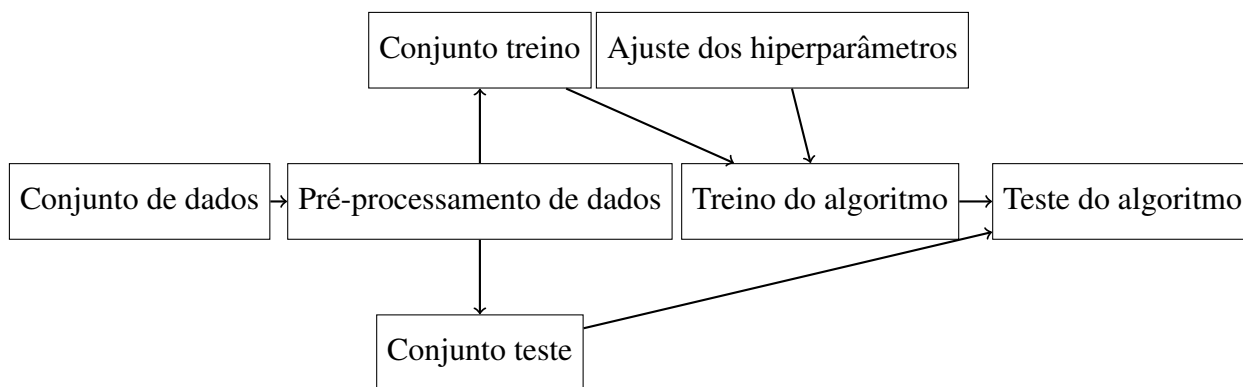


Figura 4.1: Fluxograma mostrando a aplicação de um algoritmo supervisionado

de transformações como a padronização descrita na equação

$$y = \frac{x - \mu}{\sigma}, \quad (4.1)$$

onde x é um elemento do conjunto de dados, μ a média do conjunto de dados, σ a variância e y o elemento transformado. Além disso, é preciso também filtrar os dados buscando eliminar dados perdidos, infinitos ou entradas defeituosas, do tipo NaN, *not a number*, que ocorrem, por exemplo, quando se tenta calcular uma função para uma entrada que não pertence ao seu domínio.

Os dados são separados em um conjunto-treino e um conjunto-teste. O treino do algoritmo é o processo de ajuste dos seus parâmetros de modo a cometer o menor erro possível no conjunto de validação, um subconjunto do conjunto treino com o propósito de validar o treino. Por exemplo, o treino de redes neurais pode ser realizado usando gradiente descendente e suas generalizações, detalhados na seção 4.2. Após o treino do algoritmo é realizado o teste do algoritmo, onde a saída predita pelo algoritmo usando o conjunto teste é testado em relação a uma amostra dada pelo usuário.

Existem parâmetros inerentes do pré-processamento dos dados e dos algoritmos de ML que impactam diretamente no seu desempenho. Esses parâmetros são chamados hiperparâmetros. A princípio não é possível saber como esses parâmetros influenciam no resultado, por isso é necessário fazer uma procura no espaço de hiperparâmetros de modo a maximizar o desempenho do algoritmo – essa é a fase de "tuning", ou ajuste, do algoritmo. Por exemplo, uma regressão usando o k NN possui hiperparâmetros como o número de vizinhos e uma classificação usando redes neurais possui o número de nós e o número de camadas ocultas, o significado desses hiperparâmetros serão discutidos nas seções 4.1 e 4.2, respectivamente.

Após a fase de "tuning", o treino do algoritmo é realizado usando os melhores hiperparâmetros obtidos no processo de ajuste, ou seja, usando aqueles que resultaram em menor erro durante esta fase. Após o treino do algoritmo, realizamos um teste independente usando o conjunto teste – comparando a predição do algoritmo e os rótulos dos eventos ou a variável de regressão verdadeiros. A verificação do desempenho do algoritmo depende do seu uso. Usando como exemplo uma classificação multi-classe de um evento de colisão de altas energias com 4 classes com rótulos $\Omega = \{A, B, C, D\}$ onde A representa o sinal e B,C,D representam diferentes backgrounds, podemos definir métricas como a precisão, ou *accuracy*, a matriz de confusão ou a curva característica de operação do receptor [142], por exemplo. A métrica curva característica de operação do receptor será discutido no capítulo 5. O desempenho do algoritmo pode ser verificado ao comparar y_{clf} com um vetor y_{true} que contém os rótulos do conjunto teste com N eventos, como na tabela 4.1.

Evento	1	2	3	4	5	6	7	8	9	...	N
Predito (y_{clf})	A	C	A	A	D	D	B	B	C	...	A
Verdadeiro (y_{true})	A	C	A	B	D	D	D	B	C	...	C

Tabela 4.1: Exemplo da comparação entre o vetor predito (y_{clf}) e vetor verdadeiro (y_{true}) para uma classificação multi-classe.

No exemplo da tabela 4.1 o quarto evento possui um rótulo verdadeiro igual a B, mas o rótulo predito pelo algoritmo foi A. Assim a porcentagem de eventos classificados corretamente é a *accuracy* definida como,

$$\text{accuracy} = 100 * \frac{\text{N}^\circ \text{ de classificações corretas}}{\text{N}^\circ \text{ total de classificações}}. \quad (4.2)$$

A matriz confusão permite a visualização do desempenho do algoritmo, cada linha da matriz representa a quantidade de eventos pertencentes a classe da linha enquanto cada coluna representa a quantidade de eventos preditos para a classe da coluna. Usando o exemplo acima com $N = 220$ eventos, uma possível matriz confusão para um conjunto teste é

		Classe Predita			
		A	B	C	D
Classe Verdadeira	A	50	1	3	6
	B	5	40	4	1
	C	3	3	25	9
	D	10	5	5	50

As classes A, B, C e D tem 60, 50, 40, 70 eventos, respectivamente. Enquanto as classes A, B, C e D tem 68, 49, 37 e 66 eventos preditos, respectivamente. A classe A tem 50 eventos preditos como A, 1 evento predito como B, 3 eventos preditos como C e 6 eventos preditos como D; a classe B tem 5 eventos preditos como A, 40 eventos preditos como B, 4 eventos preditos como C e 1 evento predito como D; a classe C tem 3 eventos preditos como A, 3 eventos preditos como B, 25 eventos preditos como C e 9 eventos preditos como D; a classe D tem 10 eventos preditos como A, 5 eventos preditos como C, 5 eventos preditos como C e 50 eventos preditos como D. A partir dos elementos da diagonal principal que são a quantidade de eventos classificados corretamente para cada classe podemos calcular a acurácia, obtendo uma acurácia de

$$\text{accuracy} = 100 \times \frac{50 + 40 + 25 + 50}{220} = 100 \times \frac{165}{220} = 75\%. \quad (4.3)$$

Em uma classificação de um evento de colisão de altas energias conhecer o desempenho da separação de sinal e backgrounds é importante, podemos considerar o sinal como positivo e os backgrounds como negativo, no caso juntamos as classes B,C,D como uma única classe como se fosse uma classificação binária entre sinal e backgrounds. A partir da matriz confusão podemos calcular verdadeiro positivo (VP), verdadeiro negativo (VN), falso positivo (FP), falso negativo (FN), positivo (P) e negativo (N) como

$$\text{VP} = 50, \quad (4.4)$$

$$\text{VN} = 40 + 4 + 1 + 3 + 25 + 9 + 5 + 5 + 50 = 142, \quad (4.5)$$

$$\text{FP} = 5 + 3 + 10 = 18, \quad (4.6)$$

$$\text{FN} = 1 + 3 + 6 = 10, \quad (4.7)$$

$$\text{P} = \text{VP} + \text{FP} = 50 + 18 = 68, \quad (4.8)$$

$$\text{N} = \text{VN} + \text{FN} = 142 + 10 = 152 \quad (4.9)$$

onde VP são os eventos do sinal preditos como positivo, VN são os eventos dos backgrounds preditos como negativos, FP são os eventos dos backgrounds preditos como positivo e FN são os eventos de sinal preditos como negativo, P são todos os eventos preditos como positivo e N são todos os eventos preditos como negativos. Observamos que só podemos definir VP, VN,

FP, FN, P e N para uma classificação binária ou quando queremos separar as classes em dois conjuntos diferentes.

Algoritmos de regressão podem ser avaliados com métricas como o erro quadrático médio (MSE, *mean squared error*), erro absoluto médio (MAE, *mean absolute error*) e a divergência de Kullback–Leibler (KL) [143], por exemplo. Consideramos uma regressão de uma distribuição de massa invariante onde \hat{y}_i é o i -ésimo valor predito pelo algoritmo e y_i o valor verdadeiro correspondente, então as métricas definidas sobre a amostra será

$$\text{MSE}(y, \hat{y}) = \frac{1}{N_{ev}} \sum_{i=0}^{N_{ev}-1} (y_i - \hat{y}_i)^2, \quad (4.10)$$

$$\text{MAE}(y, \hat{y}) = \frac{1}{N_{ev}} \sum_{i=0}^{N_{ev}-1} |y_i - \hat{y}_i|, \quad (4.11)$$

onde N_{ev} é o número total de eventos da amostra.

Como a regressão prediz a distribuição de variável cinemática, podemos definir também a divergência KL da distribuição predita Q em relação à distribuição verdadeira P , como

$$D_{KL}(P||Q) = \sum_{i=1}^{N_{ev}} P(x_i) \log \left(\frac{P(x_i)}{Q(x_i)} \right), \quad (4.12)$$

onde $x_i \in N_{ev}$. Existe a possibilidade da divergência KL divergir quando $Q(x_i) = 0$ e $P(x_i) \neq 0$. Podemos contornar o problema usando a divergência de Jensen-Shannon (JSD) baseada na divergência KL, definida como

$$\text{JSD}(P||Q) = \frac{1}{2} D_{KL}(P||M) + \frac{1}{2} D_{KL}(Q||M), \quad (4.13)$$

onde $M = \frac{1}{2}(P + Q)$.

Os algoritmos de ML podem ter um resultado longe do desejado, causado por problemas de subajuste ou sobreajuste.

O subajuste ocorre quando o algoritmo não consegue capturar a relação entre o conjunto treino e os valores-alvos (ou a estrutura por trás dos dados, no aprendizado não supervisionado), errando muitas predições no conjunto treino e no conjunto teste. Isso ocorre quando o modelo é muito rígido. Por exemplo, quando tenta-se ajustar uma reta a um conjunto cuja relação funcional entre as entradas e as saídas é não-linear.

O sobreajuste, por outro lado, ocorre quando o algoritmo aprende a relação entre os dados

e os rótulos do conjunto treino, mas não generaliza, não consegue prever corretamente os rótulos no conjunto teste. Isso ocorre quando o modelo é tão complexo que aprende inclusive as flutuações aleatórias do conjunto treino, mas não as encontra, naturalmente, no conjunto teste. Algumas técnicas para impedir esses problemas serão discutidos em seções posteriores.

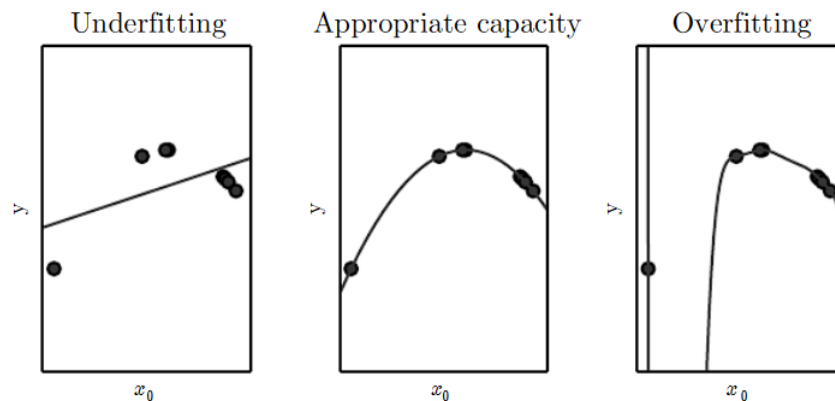


Figura 4.2: A predição de três modelos para uma função quadrática. O lado esquerdo mostra o ajuste de uma função linear sofrendo subajuste (*underfitting*). O centro mostra o ajuste de uma função quadrática que não sofre subajuste ou sobreajuste (*appropriate capacity*). O lado direito mostra o ajuste uma função polinomial de 9 grau sofrendo sobreajuste (*overfitting*). Figura 5.2 da referência [135].

Por exemplo, a figura 4.2 mostra o subajuste e o sobreajuste para uma função quadrática $y(x)$, onde a capacidade (*capacity*) do algoritmo se refere a habilidade de ajustar uma grande variedade de funções. Quando ocorre o subajuste o algoritmo não conseguiu capturar a curvatura presente nos dados, no caso de sobreajuste a solução passa por todos os pontos, mas não captura a estrutura correta dos dados, o lado esquerdo do gráfico deveria ser decrescente conforme a capacidade ideal, mas observamos um comportamento crescente.

Um algoritmo simples tem a tendência de ter uma pequena diferença entre o erro do conjunto teste e treino, mas deve existir uma complexidade mínima do algoritmo para capturar a estrutura correta dos dados, conforme ilustrado na figura 4.3. No caso de redes neurais, a complexidade escala rapidamente com o número de nós e camadas internas, tornando o sobreajuste um problema comum neste tipo de algoritmo.

O aprendizado de máquina tem várias implementações em diferentes linguagens de programação. A linguagem de programação escolhida foi o Python pela abundância de bibliotecas voltadas para o aprendizado de máquina, tornando o Python e suas bibliotecas a ferramenta principal do projeto.

As principais bibliotecas utilizadas são o NumPy [144] que adiciona o suporte a arrays multidimensionais e uma abundância de funções e rotinas para trabalhar com arrays. O SciPy [145]

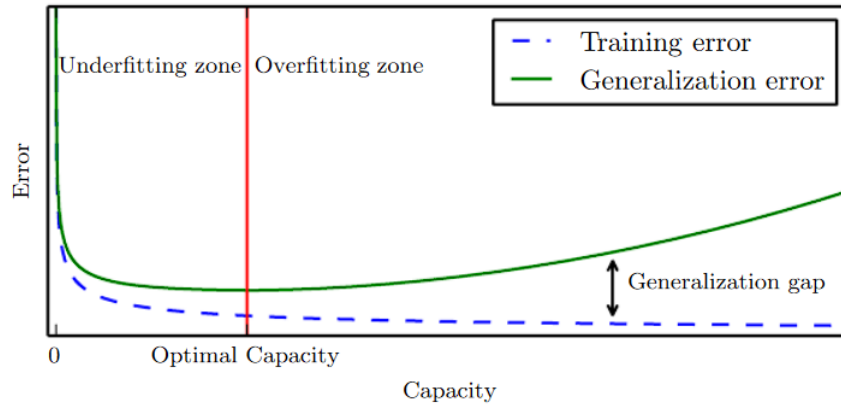


Figura 4.3: Relação entre capacidade e erro. O erro do conjunto treino (*Training error*) tem um comportamento diferente do erro do conjunto teste (*Generalization error*). O lado esquerdo do gráfico mostra ambos os erros elevados sendo o regime de subajuste. Aumentar a capacidade faz com que o erro do treino diminua, mas a diferença entre o erro do treino e teste aumente. Eventualmente, o tamanho da diferença supera a diminuição do erro do treino e entramos no regime de sobreajuste, onde a capacidade é maior que a capacidade ideal (*optimal capacity*). Figura 5.3 da referência [135].

é uma coleção de funções e algoritmos matemáticos criado em cima da biblioteca NumPy. O Matplotlib [146] serve para criar a visualização dos dados sejam elas estáticas, animadas ou interativas. O Pandas [147] é um conjunto de ferramentas para manipulação e análise de dados.

O Scikit-learn [148] é uma biblioteca de ML que inclui diversos algoritmos para regressão, classificação e clustering como o k -Nearest Neighbors, Random Forest, Gradient Boosting Machine, entre outros algoritmos. O Tensorflow [149] e o Keras [150] são as bibliotecas para a implementação de redes neurais profundas.

Vamos, agora, detalhar melhor o funcionamento dos algoritmos que usamos em nossas análises, começando pelo algoritmo de regressão.

4.1 Detalhes da regressão k NN

O regressor k -Nearest Neighbors [141] é um algoritmo simples e efetivo para interpolação. Primeiramente, definimos um conjunto de dados suporte $\mathcal{S} = \{(\mathbf{X}_i, F(\mathbf{X}_i)), i = 1, \dots, N_s\}$, exemplares que serão usados para prever o valor da função de interesse. Segundo, definimos uma distância $Dist(\mathbf{X}, \mathbf{Y})$, para decidir os exemplares de \mathcal{S} mais próximos de um novo ponto \mathbf{X}_{new} , onde \mathbf{X} , em nosso caso, é um vetor no \mathbb{R}^n . Terceiro, escolhemos quantos vizinhos de \mathbf{X}_{new} serão usados para computar $F(\mathbf{X}_{new})$, o alvo da regressão, conforme a média ponderada

$$F(\mathbf{X}_{\text{new}}) = \frac{\sum_{m=1}^k F(\mathbf{X}_m) / \text{Dist}(\mathbf{X}, \mathbf{X}_m)}{\sum_{m=1}^k 1 / \text{Dist}(\mathbf{X}, \mathbf{X}_m)}. \quad (4.14)$$

Substituindo $\text{Dist}(\mathbf{X}, \mathbf{Y}) = 1$ na fórmula 4.14 corresponde a uma média aritmética para F . A escolha entre a média aritmética e a média ponderada será decidido na procura de hiperparâmetros.

A princípio, uma vez escolhida a distância, o número de vizinhos k usado para calcular $\mathcal{T}_F(\mathbf{X}_{\text{new}})$ é o único hiperparâmetro do algoritmo. Tal tipo de modelo não possui parâmetros a serem ajustados, ao contrário de uma rede neural, que possui pesos e vieses, daí motivo de não precisarmos de uma fase de treino. Entretanto, o número de vizinhos, k , a função-distância entre os vizinhos, e possivelmente outros hiperparâmetros devem ser ajustados para obter um bom regressor minimizando alguma função erro. Todos os $F(\mathbf{X}_m)$, $m = 1, \dots, N_s$ são conhecidos, portanto, estamos trabalhando com um algoritmo supervisionado.

Em nosso caso, a função alvo da regressão, F , é a massa invariante leptônica ($\ell^+ \ell'^- \nu_{\ell} \bar{\nu}'_{\ell}$), $M_{\ell\ell\nu\nu}$. A entrada da função será a informação observada obtida dos 4-momento dos elétrons e múons, p_e e p_{μ} , respectivamente. A representação escolhida foi a energia e o momento dos léptons carregados mais, informações de alto nível construídas a partir das informações de baixo nível: $\mathbf{X} = (f_{ij}(p_i, p_{\bar{i}}), i = 1, \dots, N_{ev}, j = 1, \dots, M)$ representando N_{ev} eventos com M informações, onde teremos $M = 21$.

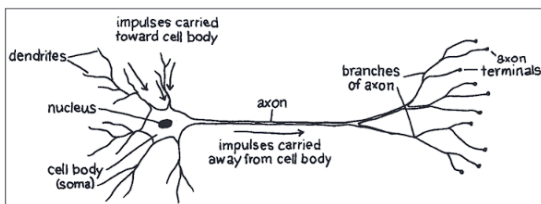
Se o número de dimensões for elevado, modelos baseados em distância como o k NN requerem uma quantidade de cálculos em cada iteração que cresce exponencialmente com a dimensão do conjunto de dados, tornando-os computacionalmente proibitivos. Esse é o problema conhecido como *maldição da dimensionalidade*. Uma possível solução é a redução da dimensionalidade [151], processo que consiste em projetar os dados do espaço de alta dimensão em um espaço de baixa dimensão, tornando a redução da dimensionalidade um passo importante antes de treinar e testar o algoritmo. Alguns exemplos de redução da dimensionalidade são a Análise de Componentes Principais (PCA) [152], *t-Distributed Stochastic Neighbor Embedding* (TSNE) [153] e o *Uniform Manifold Approximation and Projection* (UMAP) [154].

Escolhemos efetuar uma transformação linear dos dados utilizando a Análise de Componentes Principais e procurando pelo vizinho mais próximo no espaço transformado nas primeiras $P < M$ variáveis que melhor explicam a variância dos dados, $\mathbf{X}^{pca} = \mathcal{T}_P(\mathbf{X})$, onde $P \in 1, 2, 3, 4, 5, 6, 7, 8$. O motivo de não utilizar métodos como o TSNE e o UMAP é pelo tempo elevado de computação se comparado ao PCA por serem métodos não-lineares.

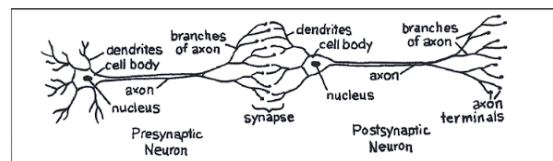
No nosso caso estamos reconstruindo uma variável que em observações experimentais é representada como um histograma, temos a escolha de prever o valor ou o bin que o evento pertence. Observamos que não existe uma diferença no resultado em ambas as abordagens mais detalhes no apêndice C.2. Assim escolhemos a abordagem onde o regressor irá prever o bin do evento.

4.2 Redes Neurais

Redes neurais são modelos não-lineares compostos por camadas de nós (neurônios) análogo a uma rede neural biológica. Em uma rede neural biológica, as células do sistema nervoso são os neurônios conectados um para o outro por axônios e dendritos e a região de conexão entre os axônios e dendritos são chamados sinapses, as conexões são ilustrados na figura 4.4. A força das conexões sinápticas frequentemente mudam em resposta a estimulação externa, essas mudanças são responsáveis pelo aprendizado em organismos vivos.



(a) Neurônio, a unidade funcional do sistema nervoso. O neurônio possui processos chamados dendritos (*dendrites*) que recebem sinais e um axônio (*axon*) que transmite sinais para outro neurônio.



(b) Neurônios transmitem informação para outro neurônio. Informação passa do axônio do neurônio pré-sináptico (neurônio que transmite informação) para os dendritos do neurônio pós-sináptico (neurônio que recebe informação).

Figura 4.4: Exemplo de um neurônio (esquerda) e rede neural biológica (direita). Figura 2.1 e 2.2 da ref [155].

As redes neurais simulam o mecanismo biológico através do uso de nós no lugar de neurônios. Cada nó é uma unidade de processamento linear que recebe um vetor do \mathbb{R}^n , análogo aos dendritos que recebem informação, e devolve um vetor no \mathbb{R}^m , análogo à transmissão de informação pelos axônios. Como exemplo a entrada é um vetor no \mathbb{R}^3 e a saída um vetor no \mathbb{R} , a estrutura pode ser observada na figura 4.5, onde $x_i, i = 1, 2, 3$ representam as componentes do vetor de entrada, $\omega_0 x_0$ ($x_0 = 1$) é o viés, ω_i o peso associado a cada componente do vetor de entrada cujo análogo biológico são a força das conexões sinápticas, ϕ a função de ativação [156, 157] e \hat{y} a saída dada pela equação

$$\hat{y} = \phi \left(\sum_{i=0}^m \omega_i x_i \right). \quad (4.15)$$

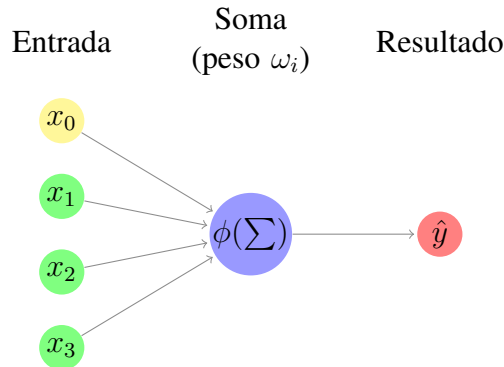


Figura 4.5: A figura mostra a estrutura de um nó com três entradas $x_i, i = 1, 2, 3$, três pesos $\omega_i, i = 1, 2, 3$ e um viés x_0 . O resultado \hat{y} é dado pela equação 4.15 usando a função de ativação ϕ . O análogo biológico é a entrada ser os dendritos, os pesos serem a força das conexões sinápticas, a soma ser o núcleo do neurônio, a função de ativação ser o axônio que define o resultado do nó.

A rede neural é uma função que usa a camada de entrada e devolve uma saída que usa os pesos dos nós como parâmetros intermediários. O aprendizado ocorre pela atualização dos pesos, análogo ao estímulo externo que muda a força das conexões sinápticas em uma rede neural biológica. No caso, o estímulo externo será definido pelo treino da rede neural. A saída da rede neural depende do seu objetivo que pode ser uma classificação ou regressão no aprendizado supervisionado. A garantia que a rede neural de uma camada oculta pode sempre representar uma função adequada ao seu objetivo é dado pelo teorema da aproximação universal [158, 159], para redes com mais de uma camada oculta existem variações do teorema [160, 161]. Contudo, ajustar tal função a uma determinada tarefa de aprendizado é difícil.

A figura 4.6 é uma visualização de uma rede neural que possui a camada de entrada com quatro componentes, duas camadas intermediárias que chamamos camadas ocultas, a primeira camada oculta com quatro nós, a segunda camada oculta com três nós e uma camada de saída com dois nós. Todas as camadas transmitem o resultado para a próxima camada que recebe o resultado como entrada, análogo ao neurônio pré-sináptico e pós-sináptico. Por exemplo, na figura cada nó da camada oculta 2 ($h_i^{(2)}, i = 1, 2, 3$) recebe o resultado da camada oculta 1 e transmite o resultado para a próxima camada que no caso é a camada de saída.

Avaliar o objetivo da rede neural pode ser alcançado fazendo o “forward pass”, processo de cálculo dos valores da camada de saída a partir dos valores da camada de entrada que atravessam toda a rede. Os valores da camada de saída são utilizados em uma função perda que quantifica

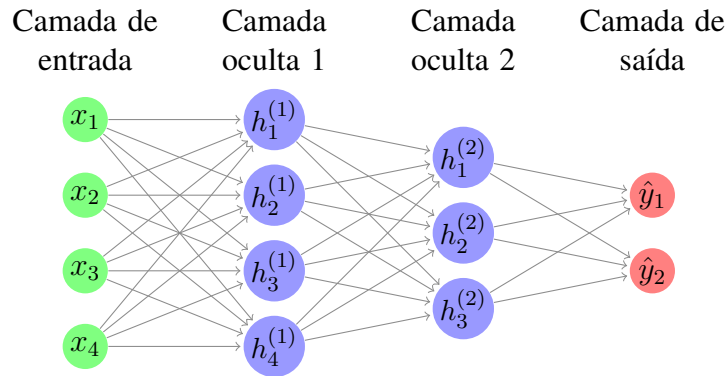


Figura 4.6: A figura mostra um exemplo de rede neural com uma camada de entrada, duas camadas ocultas e uma camada de saída. A camada de entrada tem quatro nós, a primeira camada oculta tem quatro nós, a segunda camada oculta tem três nós e a camada de saída tem dois nós.

o erro do objetivo. O “backward pass” é a atualização dos pesos através de algoritmos como o gradiente descendente [162, 163] que implementa o chamado “backpropagation” [135, 156, 157], processo do cálculo do gradiente da função perda em relação aos pesos que começa na camada de saída e retroativamente faz a regra da cadeia até a camada de entrada. O processo de treino, portanto, consiste em diversas iterações (“forward pass” mais “backward pass”) no conjunto de dados até a obtenção de um resultado satisfatório dado pelo cálculo da função perda ou do número de iterações desejados pelo usuário. Em geral, separamos um conjunto teste específico para o processo de aprendizado que leva o nome de conjunto de validação. O conjunto de validação é utilizado para o cálculo da função perda no processo de treino. O cálculo da atualização dos pesos dado pelo backpropagation está no apêndice A.

4.2.1 Função de Ativação

O resultado da saída de um nó é determinado pela sua função de ativação ϕ que possui a tarefa de decidir a intensidade da saída do nó. Seu análogo biológico, o axônio responsável pela transmissão de informação, pode ter saída nula ou não. Os neurônios artificiais, contudo, não precisam ser totalmente ativados ou desativados. O papel da função de ativação também será introduzir não-linearidade permitindo à rede neural aprender padrões complexos não-lineares. Exemplos de funções de ativação são

$$\phi(x) = x \quad \text{Linear,} \quad (4.16)$$

$$\phi(x) = \frac{1}{1 + e^{-x}} \quad \text{Sigmóide,} \quad (4.17)$$

$$\phi(x) = \tanh x \quad \text{Tanh,} \quad (4.18)$$

$$\phi(x) = \max(0, x) \quad \text{ReLU,} \quad (4.19)$$

$$\phi_k(x) = e^{x_k} / \sum_{i=1}^K e^{x_i} \quad \text{SoftMax,} \quad (4.20)$$

mostrados na figura 4.7. Com exceção da função de ativação linear (eq. 4.16) todas as funções são não-lineares e diferenciáveis, com exceção de ReLU, a *Rectified Linear Unity*, que não é diferenciável em $x = 0$. A escolha da função de ativação depende se é utilizada em uma camada oculta ou camada de saída. Em camadas ocultas a escolha é baseado no desempenho da rede neural que é dependente do conjunto de dados. Note que, com exceção da ReLU, as outras funções nunca são desligadas totalmente.

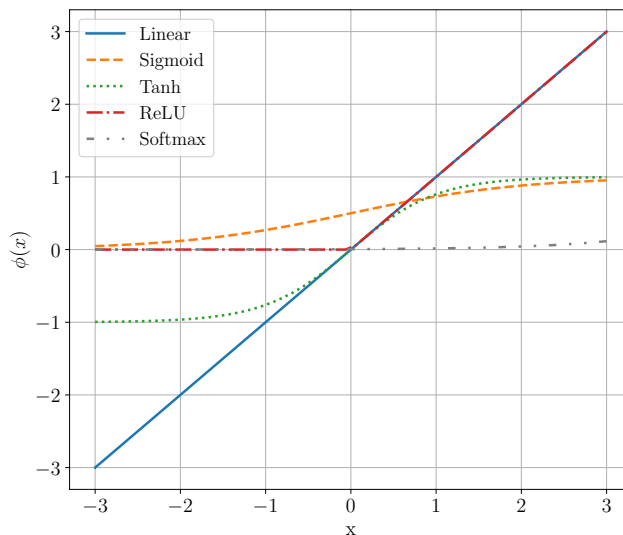


Figura 4.7: Cinco exemplos de funções de ativação. Com exceção da função linear, todas as funções são não lineares.

No aprendizado supervisionado, a camada de saída pode ter como resultado uma saída com valores reais (regressão) ou um conjunto de probabilidades (classificação). Assim a função de ativação para a camada de saída é definido pelo seu objetivo. Por exemplo, uma classificação binária com classes $\Omega \in \{0, 1\}$ usa a função de ativação Sigmóide (eq. 4.17) [157], cuja saída pode

ser interpretado como uma distribuição de probabilidade entre 0 e 1. Assim podemos interpretar os eventos com valores próximos de 0 como eventos da classe 0 e eventos com valores próximos de 1 como sendo eventos da classe 1.

Na classificação multi-classe a camada de saída usa a função de ativação SoftMax(eq. 4.20) [157]. A camada de saída tem um nó por classe em uma classificação multi-classe. Por exemplo, em uma classificação com 4 classes $\Omega \in A, B, C, D$ a camada de saída possui 4 nós uma para cada classe. A saída de cada nó será a probabilidade do evento ser da classe correspondente a do nó, a garantia que a soma das probabilidades seja igual a 1 é dado pela função de ativação SoftMax.

4.2.2 Função perda

A função perda é uma métrica do erro cometido na predição da rede neural (ou qualquer outro algoritmo), agregando os erros sobre o conjunto de dados teste e calculando uma média para obter um valor que representa a proximidade da rede neural com os resultados verdadeiros. A busca do estado ideal é equivalente a encontrar os parâmetros (pesos e vieses) que minimizam a função perda. Assim a função perda ajuda a transformar o treino de uma rede neural como um problema de otimização. A busca dos melhores parâmetros não pode ser feita analiticamente, mas pode ser aproximada com algoritmos de otimização como o gradiente descendente como explicamos na seção anterior.

A escolha da função perda depende da aplicação em questão. Exemplos de função perda em uma regressão é o erro quadrático médio, o erro absoluto médio, a divergência KL e a divergência JSD, definidas nas equações 4.10, 4.11, 4.12 e 4.13, respectivamente.

Para as classificações, utilizar a *accuracy* como função perda (que deve ser maximizada) é possível, mas como é uma função entre valores discretos a atualização dos pesos em uma rede neural pode não ser adequada. Por exemplo, em uma classificação com 4 classes $\Omega \in A, B, C, D$ em um conjunto desbalanceado onde a classe A tem 2000 eventos e as classes B, C e D tem 100 eventos cada um, uma acurácia de 95% no treino e a maioria da predição correta ser da classe A indica que a atualização dos pesos será enviesado para o acerto da classe A, ou seja, será enviesado em relação à classe predominante. Utilizar uma função perda em relação às probabilidades do evento ser da classe X que será um valor contínuo permite uma atualização correta dos pesos em relação a todas as classes. Uma função perda muito utilizada nas classificações é a entropia cruzada.

Podemos usar a estimação da máxima verossimilhança (*maximum likelihood*) [157, 164] para deduzir a entropia cruzada. A estimação da máxima verossimilhança é um método para estimar os parâmetros de uma distribuição de probabilidade, considerando dados observados. A estimação é feita maximizando uma função verossimilhança que nas condições do modelo estatístico os dados observados são os prováveis.

Por exemplo, uma classificação multi-classe de um evento de colisão de altas energias com k classes e N eventos com n_1, n_2, \dots, n_k onde n_i é o número de eventos da classe i com $\sum_{i=1}^k n_i = N$. Os dados são $D = (x_1, y_1), \dots, (x_n, y_n)$ onde $x \in \mathbb{R}^n$ e $y \in 1, 2, \dots, k$ são o alvo de interesse. Um modelo de rede neural $f(x, \theta)$ com parâmetros θ prediz a probabilidade de y para dado x , onde x e y são os eventos verdadeiros do conjunto treino. Pela estimativa da máxima verossimilhança queremos encontrar θ que maximiza a função verossimilhança $V(\theta)$. Assumindo uma distribuição multinomial e que cada evento dos dados D são independentes, temos

$$V(\theta) = \frac{n!}{n_1!n_2! \cdots n_k!} \prod_{i=1}^n f(x_i, \theta)^{y_i}, \quad (4.21)$$

fazendo o logaritmo de ambos os lados e usando propriedades do logaritmo obtemos

$$\log V(\theta) = \log n! - \log n_1!n_2! \cdots n_k! + \sum_{i=1}^n y_i \log f(x_i, \theta), \quad (4.22)$$

como $\log n! - \log n_1!n_2! \cdots n_k!$ não depende de θ podemos ignorar obtendo

$$\log V(\theta) = \sum_{i=1}^n y_i \log f(x_i, \theta), \quad (4.23)$$

em vez de maximizar o lado direito, podemos minimizar o seu valor negativo

$$-\log V(\theta) = - \sum_{i=1}^n y_i \log f(x_i, \theta). \quad (4.24)$$

Assim a função perda entropia cruzada $L(f(x_i, \theta), y_i)$ é definida como

$$L(f(x_i, \theta), y_i) = - \sum_{i=1}^n y_i \log f(x_i, \theta). \quad (4.25)$$

4.2.3 O Gradiente Descendente

A ideia do gradiente descendente [165] é minimizar a função de interesse (a função perda da rede neural, no nosso caso) atualizando os parâmetros θ da função (os parâmetros da rede neural) indo no sentido contrário do gradiente calculado em respeito aos parâmetros, considerando que o gradiente aponta no sentido de maior crescimento da função de interesse. O algoritmo para quando alcança a convergência da função de interesse ou pela condição dada pelo usuário, o número de passos, por exemplo. Considerando o gradiente descendente no passo t , a atualização dos parâmetros é dada pelas equações

$$v_t = \alpha \nabla_{\theta} F(\theta_t), \quad (4.26)$$

$$\theta_{t+1} = \theta_t - v_t, \quad (4.27)$$

onde α é uma constante que define o tamanho do passo, chamado taxa de aprendizado (*learning rate*). Observamos que dependendo do tamanho de α , o gradiente descendente tem dificuldades na convergência, pois pode divergir completamente se α é muito grande ou aproximar-se do mínimo desejado muito lentamente quando α é muito pequeno. O ajuste de α é realizado na fase de "tuning" do algoritmo.

O cálculo do gradiente descendente possui problemas como o alto custo computacional além de dificuldades de convergência do algoritmo. O alto custo computacional nos leva a separar o conjunto de dados no processo de treino em lotes. Em vez de realizar uma iteração para cada elemento do conjunto de dados é feito uma iteração para o lote que contém N_{ev} eventos, ou seja, em vez de N_{ev} iterações temos apenas uma iteração contendo N_{ev} eventos. Assim uma iteração é definido como o "forward pass" mais "backward pass" no lote. Portanto, o treino da rede neural é definido pelo número de épocas, onde uma época se refere ao treino no conjunto de dados, onde uma época contém $(N_{ev}/\text{tamanho do lote})$ iterações.

As dificuldades relacionadas ao gradiente descendente podem ser amenizadas [165]. Técnicas baseadas no *momento* servem para amenizar o possível efeito de oscilação do gradiente descendente ao incluir um termo na atualização dos parâmetros, assim a atualização no passo t é

$$v_t = \alpha \nabla_{\theta} F(\theta_t) - \beta v_{t-1}, \quad (4.28)$$

$$\theta_t = \theta_t - v_t, \quad (4.29)$$

onde $\beta \in [0, 1)$ determina o decaimento exponencial da taxa da contribuição dos gradientes anteriores. O decaimento exponencial vem da multiplicação de β na atualização. A contribuição dos gradientes anteriores garantem o sentido do momento diminuindo o efeito a oscilação.

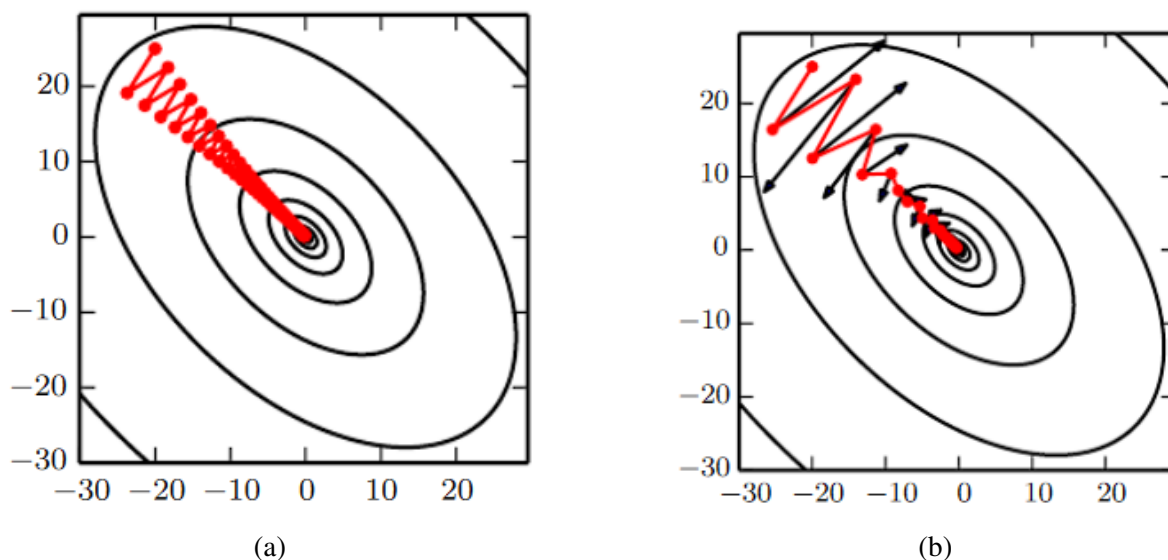


Figura 4.8: Comparação do gradiente descendente (lado esquerdo) e do gradiente descendente com momento (lado direito) para uma função perda quadrática (linha contornada) com mínimo na origem e uma taxa de aprendizado fixo. O caminho vermelho sobre os contorno indica a trajetória do gradiente descendente. As flechas pretas no lado direito indicam a direção e sentido do passo que seria realizado pelo gradiente descendente sem momento. Figura 4.6 e 8.5 da referência [135].

A figura 4.8 mostra uma comparação do gradiente descendente e do gradiente descendente com momento para uma função perda quadrática. A figura 4.8a mostra o caso sem momento onde a convergência é lenta pela oscilação do gradiente. Enquanto a figura 4.8b mostra o caso de gradiente descendente com momento onde a convergência é mais rápida e o efeito de oscilação é amenizado pelo uso do momento.

Outra técnica para generalizar o gradiente descendente é ter uma taxa de aprendizado adaptativo. Em vez de usar um α fixo podemos escolher um α_t que depende do passo/tempo. Uma implementação é criar um cronograma para a taxa de aprendizado, onde o usuário escolhe como α_t deve evoluir, em redes neurais, podemos escolher o cronograma baseado no número

de iterações ou no número de épocas.

Os outros métodos adaptam a taxa de aprendizado em função dos gradientes e momentos. Por exemplo, o algoritmo ADAM [163], usa uma taxa de aprendizado α fixo e dois momentos m e n , onde m representa a média móvel dos gradientes e n a média móvel do gradiente quadrado com parâmetros $\beta_1, \beta_2 \in [0, 1)$ que controlam a taxa de decaimento exponencial. As médias móveis são a estimativas do primeiro momento(média) e do segundo momento(variância descentralizada) do gradiente. A inicialização na origem m e n são enviesados em direção a origem, a correção do viés é realizado resultando nas estimativas com correção de viés \hat{m} e \hat{n} . A adaptação da taxa de aprendizado vem da multiplicação de α com $\hat{m}/\sqrt{\hat{n}}$ na atualização dos parâmetros.

No nosso caso, aplicaremos o algoritmo em uma rede neural $f(x_i, \theta)$ com predição y_i para minimizar a função perda $L(f(x_i, \theta), y_i)$, onde x_i e y_i são os dados do lote e a predição do lote, respectivamente. Na prática, podemos ter problemas com a precisão numérica dos gradientes, para impedir uma divisão por zero, definimos ϵ uma constante para impedir a divisão por zero. Assim o funcionamento do algoritmo ADAM pode ser observado no algoritmo 1. Existem outros algoritmos com taxa de aprendizado adaptativo como Adagrad [166] e RMSProp [167].

Algorithm 1 Algoritmo ADAM com parâmetros padrões

Require: $\alpha = 0.001$: Tamanho do passo

Require: $\beta_1 = 0.9, \beta_2 = 0.999; \beta_1, \beta_2 \in [0, 1)$: Taxas de decaimento exponencial para estimativa dos momentos

Require: $\epsilon = 10^{-8}$: Constante para estabilização numérica

Require: $L(f(x_i, \theta), y_i)$: Função perda com parâmetros θ

Require: θ_0 : Vetor de parâmetros iniciais

$m_0 \leftarrow 0$ (Inicializa vetor do primeiro momento)

$n_0 \leftarrow 0$ (Inicializa vetor do segundo momento)

$t \leftarrow 0$ (Inicializa passo do tempo)

while θ_t não converge **do**

$t \leftarrow t + 1$

$g_t \leftarrow \nabla_{\theta} L_t(f(x_i, \theta_{t-1}), y_i)$ (Gradiente com respeito ao parâmetros θ no tempo t)

$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t$ (Atualiza a estimativa do primeiro momento enviesado)

$n_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2$ (Atualiza a estimativa do segundo momento enviesado)

$\hat{m}_t \leftarrow m_t / (1 - \beta_1^t)$ (Correção do viés no primeiro momento)

$\hat{n}_t \leftarrow n_t / (1 - \beta_2^t)$ (Correção do viés no segundo momento)

$\theta_t \leftarrow \theta_{t-1} - \alpha \cdot \hat{m}_t / (\sqrt{\hat{n}_t} + \epsilon)$ (Atualização dos parâmetros)

end while

return θ_t

4.2.4 Regularização

A facilidade de construir uma rede neural com um modelo complexo/capacidade elevada é um dos motivos do sobreajuste ser um problema comum. A regularização é um dos principais métodos para reduzir o sobreajuste sendo projetados para reduzir o erro no conjunto teste/validação, ao custo de um possível aumento no erro do conjunto treino. Existem diversas técnicas de regularização, uma delas consiste em limitar a capacidade do modelo ao penalizar a função perda. Usando as definições utilizadas anteriormente, a penalização (de um lote) é descrita pela equação

$$\hat{L}(f(x_i, \theta), y) = L(f(x_i, \theta), y_i) + \lambda \Omega(\theta), \quad (4.30)$$

onde \hat{L} é a função perda penalizada, λ , $\lambda > 0$ é o hiperparâmetro que controla a contribuição relativa do termo da penalidade e Ω é a penalidade dos parâmetros. Usualmente $\Omega = \|\omega_i\|_2^2$, onde ω_i são os pesos da rede neural e leva o nome de regularização de Tikhonov ou regularização L_2 , obtendo

$$\hat{L}(f(x_i, \theta), y_i) = L(f(x_i, \theta), y_i) + \lambda \|\omega_i\|_2^2. \quad (4.31)$$

A vantagem de usar uma rede neural com regularização é a possibilidade de usar um modelo complexo e ajustar o parâmetro λ conforme o problema específico, criando uma grande flexibilidade na implementação de uma rede neural. No treino da rede neural a atualização dos pesos definida pelo gradiente descendente sem regularização é

$$\omega_i \leftarrow \omega_i - \alpha \frac{\partial L}{\partial \omega_i}, \quad (4.32)$$

onde α é a taxa de aprendizado. Na regularização L_2 , a atualização dos pesos é obtida calculando o gradiente de \hat{L} , obtemos

$$\omega_i \leftarrow \omega_i(1 - \alpha\lambda) - \alpha \frac{\partial L}{\partial \omega_i}, \quad (4.33)$$

com a diferença sendo a multiplicação do peso ω_i pelo termo de decaimento $(1 - \alpha\lambda)$, diminuindo o vetor do peso a cada passo.

Existem outras regularizações similares como a regularização L_1 onde $\Omega = \|\omega_i\|_1$ que possuem um uso específico em redes neurais esparsas [156], ou seja, uma rede neural com

componentes ω_i nulos. Comparando a regularização L_1 e L_2 , a regularização L_2 na maioria das aplicações possui um desempenho melhor que a regularização L_1 [124]. Portanto, usamos a regularização L_2 .

Uma técnica de regularização envolve salvar os parâmetros toda vez que o erro no conjunto de validação melhora, antecipando um possível sobreajuste. Quando o algoritmo termina o treino, o modelo retorna os parâmetros salvos, em vez dos últimos parâmetros obtidos. O algoritmo termina quando não obtemos um novo parâmetro que melhora o erro no conjunto de validação para um número pré-determinado de épocas. Essa técnica leva o nome de *early stopping*, pois o treino tende a ser menor do previsto.

Outra técnica de regularização utilizado é o dropout [168] que previne adaptações complexas no conjunto de treino, diminuindo o sobreajuste. A técnica envolve omitir aleatoriamente a informação transmitida pelo nó de uma camada oculta para a próxima, a omissão é decidido por uma probabilidade p que é um hiperparâmetro dado pelo usuário.

No próximo capítulo explicamos como os algoritmos de classificação e regressão são usados na identificação mais precisa de bósons de Higgs pesados decaindo em pares de W totalmente leptônicos.

Capítulo 5

Reconstrução de ressonâncias totalmente leptônicas - Modelo Simplificado

Temos dois objetivos principais: (1) é mostrar que a ressonância pode ser reconstruída de forma confiável, e (2) ao usar a ressonância podemos melhorar o desempenho da *accuracy* de classificadores de ML (e outras métricas), e mais importante, da significância estatística do sinal comparado ao classificador sem a massa $M_{\ell\ell\nu\nu}$ regredida. O valor da significância estatística depende do número de eventos de sinal que, por sua vez, é dependente do modelo. Como estamos mais interessados em mostrar o ganho da significância estatística, fixaremos o número de eventos do sinal para ilustrar o método utilizado. A única suposição é que os léptons e neutrinos do sinal de um novo escalar pesado provém predominantemente do decaimento em WW com interferências negligenciáveis com os backgrounds correspondentes do MP.

Também não estamos interessados em estudar um modelo específico de um novo bóson de Higgs, ao invés disso, assumiremos um modelo simplificado que prevê as interações H_2 -glúons e H_2 -bósons W , onde H_2 é o novo Higgs pesado. Existem, contudo, vários modelos onde tais interações ocorrem. Por exemplo, o 2HDM [169], um modelo com dois dubletos de Higgs que possui diversas motivações como: (1) previsão de áxions que permitem uma possível solução para o problema de violação CP na QCD [61], (2) mecanismo de bariogênese [170, 171] que o modelo padrão não consegue explicar [172], (3) é o setor escalar mínimo para modelos supersimétricos [59] os quais oferecem solução para o chamado *problema da hierarquia do Higgs*. Nesse caso, porém, é de se esperar contribuições também do canal $H_2 \rightarrow ZZ$. Outro exemplo são os modelos 331 [173, 174] onde, ao invés da inclusão de outro dubleto de escalares, o dubleto de $SU(2)_L$ do modelo padrão é promovido a um triplete. De qualquer forma, léptons

oriundos de ZZ são facilmente identificáveis. Por exemplo, é possível suprimir a contribuição do canal ZZ vetando pares de léptons carregados cuja massa invariante sejam próxima da massa do Z .

De uma forma simplificada, escolhemos um modelo de Higgs pesado com acoplamento em WW similar ao do Higgs padrão, porém com acoplamentos em ZZ suprimidos¹. A lagrangiana é

$$\mathcal{L}_{H_2} = \mathcal{L}_{MP} + V(h, H_2) + g_{H_2} m_W (W_\mu^+ W^{-\mu}) H_2 - \left(\frac{H_2}{v} C_h - \frac{H_2^2}{2v^2} C_{hh} \right) \text{tr}(\mathbf{G}_{\mu\nu} \mathbf{G}^{\mu\nu}), \quad (5.1)$$

onde $V(h, H_2)$ representa o potencial escalar do modelo, g_{H_2} a constante de acoplamento entre H_2 e os bósons W e C_h, C_{hh} são os coeficientes de Wilson.

A produção dominante do H_2 é via fusão de glúons e assumimos acoplamentos iguais aos do modelo padrão. O processo de colisão simulado foi

$$gg \rightarrow H_2 \rightarrow W^+ W^- \rightarrow \ell^\pm \ell^\mp + \cancel{E}_T, \quad (5.2)$$

onde $\ell(\ell') = e, \mu$ e o *missing energy* é todo devido aos neutrinos que escapam da detecção. Assumiremos três diferentes valores de massas do Higgs pesado: 1 TeV, 1.5 TeV, 2 TeV e largura de 1% e 10% da massa ($\Gamma_H/m_H = 10\%$). A análise dos resultados considerou apenas os casos com largura de 10%, o apêndice B tem uma discussão comparando os resultados entre as diferentes larguras.

Consideramos os seguintes backgrounds do canal leptônico [175] para a nossa análise:

$$pp \rightarrow W^+ W^- \rightarrow \ell^\pm \ell^\mp + \cancel{E}_T, \quad (5.3)$$

$$pp \rightarrow ZZ(\gamma^*) \rightarrow \ell^\pm \ell^\mp + \cancel{E}_T, \quad (5.4)$$

$$pp \rightarrow t\bar{t} \rightarrow b\bar{b} + \ell^\mp \ell^\pm + \cancel{E}_T, \quad (5.5)$$

onde a eq. 5.3 é o background irreduzível dominante, eq. 5.4 é o background irreduzível subdominante e eq. 5.5 o background redutível dominante. Todas as simulações de eventos

¹De qualquer forma, é fácil vetar eventos com um bóson- Z , basta eliminar todos os pares de léptons carregados cuja massa invariante dos léptons esteja próxima da massa do Z .

partônicos foram feitas através do MadGraph5 2.7.3 [176] ao nível de árvore. A hadronização dos quarks é realizado pelo Pythia8.302 [177], enquanto efeitos dos detetores [178] são simulados com o Delphes3.4.2 [179] usando a configuração(card) do detetor CMS.

O conjunto de dados gerados contém 3.9 milhões de eventos de background(1.3 milhão de eventos para cada background) e 400 mil eventos de cada sinal. O conjunto de dados para a regressão contém 300 mil eventos de cada background e 150 mil de cada sinal e a proporção de eventos separados foram 20% teste e 80% para a busca em rede. Cada classificador recebeu 500 mil eventos de cada background e 250 mil de sinal e a proporção de eventos separados foram 70% treino, 20% teste e 10% no conjunto validação para os classificadores. A tabela 5.1 mostra como foi feito a separação dos eventos para uso nos algoritmos para um caso.

Número de eventos	k NN		NN ₁ /NN ₂		
	Teste	Ajuste	Treino	Teste	Validação
Porcentagem	20%	80%	70%	20%	10%
Nº de Eventos Sinal	30000	120000	175000	50000	25000
Nº de Eventos Background	60000	240000	350000	100000	50000
Nº de Eventos Total	210000	840000	12250000	350000	175000

Tabela 5.1: Tabela mostrando o número de eventos para uso nos algoritmos de ML. O total significa o número de eventos para um caso que contém o sinal mais, três backgrounds.

Eventos partônicos foram usados para obter a distribuição da massa $M_{\ell\ell\nu\nu}$ verdadeira, visto que o 4-momento dos neutrinos estão disponíveis. Estamos explicitamente assumindo que toda a energia perdida é causada pelo escape dos neutrinos produzidos no espalhamento duro, mas não devido à má reconstrução dos momentos observados ou do momento transversal de outras partículas. Contudo, o momento dos léptons e a energia perdida que alimenta os algoritmos são os eventos partônicos que incluem os efeitos de hadronização e efeitos de detetores simulados. Portanto, parte da diferença entre a distribuição verdadeira e a predita pelo k NN podem ser explicadas dessa maneira. Usar eventos partônicos não é um demérito, pois existem trabalhos de aplicação de ML que tentam precisamente descontar efeitos de detetor e hadronização e recuperar a informação ao nível de párons [180, 181].

As simulações realizadas utilizam o LHC como base. O LHC é o maior acelerador de partículas do mundo que realiza colisões próton-próton com energia de centro de massa de $\sqrt{s} = 13$ TeV construído pelo *European Organization for Nuclear Research* (CERN). O LHC consiste em um anel de ímãs supercondutores de 27 km com estruturas para aumentar a energia da partícula ao longo do caminho. Dentro do LHC existem diversos detetores como o

ATLAS [16] e CMS [17].

O detetor *Compact Muon Solenoid* (CMS) é um detetor multiuso que consiste em um solenoide supercondutor com 13 m de comprimento e 6 m em diâmetro, que fornece um campo magnético axial de 3.8 T. No volume do solenoide existem diversos sistemas para detecção de partículas. Trajetórias de partículas-carregadas são medidos por um detetor de píxel de silício e rastreadores de *strip* que cobrem uma região de $0 < \phi \leq 2\pi$ no ângulo azimutal e $|\eta| < 2.5$, onde a pseudo-rapidez η é definido como $-\ln[\tan(\theta/2)]$, com θ sendo o ângulo polar da trajetória da partícula com respeito a direção do feixe no sentido anti-horário. Um calorímetro eletromagnético de cristal chumbo tungstênio (ECAL) e um calorímetro hadrônico de latão e cintilador cercam o volume de rastreamento. Os calorímetros provêm medidas de energia e direção dos elétrons e jatos hadrônicos. Múons são detectados por detetores de ionização a gás embutidos no jugo de retorno de fluxo de aço fora do solenoide. O detetor é quase hermético, permitindo medições de balanço do momento transverso (p_T) no plano perpendicular da direção do feixe. Um sistema de gatilhos com dois níveis seleciona eventos de interesse para análise física. Os gatilhos têm uma cobertura da pseudo-rapidez de léptons de $|\eta| < 2.5(2.4)$ para elétrons(múons) que limita $\Delta R_{\ell\ell} > 0.4$, onde $\Delta R_{\ell\ell} = \sqrt{(\Delta\eta_{\ell\ell})^2 + (\Delta\phi_{\ell\ell})^2}$, $(\Delta\eta_{\ell\ell})$ representa a diferença na pseudo-rapidez e $(\Delta\phi_{\ell\ell})$ a diferença no ângulo azimutal ϕ dos léptons carregados. Uma descrição mais detalhada do detetor CMS, junto a definição do sistema de coordenadas utilizado e variáveis cinemáticas relevantes pode ser visto na referência [17].

O detetor ATLAS é um detetor multiuso com uma geometria cilíndrica simétrica frente-trás e uma cobertura de ângulo sólido próximo de 4π . O detetor de rastreamento interno (ID), que cobre uma região de pseudo-rapidez $|\eta| < 2.5$, consiste em um detetor de píxel de silício, um detetor *microstrip* de silício e um rastreador transição-radiação. O detetor interno é cercado por um fino solenoide supercondutor fornecendo um campo magnético de 2 T, e segmentado por um calorímetro eletromagnético de chumbo/argônio líquido (LAr). Um calorímetro hadrônico de aço/telha cintilante cobre a região central. As regiões frontal e do fim da capsula são cobertos por calorímetros eletromagnéticos e hadrônicos de LAr com aço, cobre e tungstênio como materiais absorventes. Um sistema de espectrômetro de múon que incorpora ímãs toroidais supercondutores cercam os calorímetros. Três camadas de câmaras de fio de precisão cobrem uma região de $|\eta| < 2.7$ para o rastreamento dos múons, enquanto câmaras rápidas são utilizados para desencadear na região de $|\eta| < 2.4$. Uma descrição mais detalhada do detetor ATLAS pode ser visto na referência [16].

Adotamos os seguintes cortes cinemáticos para aceitação de eventos com dois léptons carregados de cargas opostas e energia perdida

$$p_{T,\ell} > 20 \text{ GeV}, \quad |\eta_\ell| < 2.4, \quad \Delta R_{\ell\ell} > 0.4, \quad M_{\ell\ell} > 30 \text{ GeV}, \quad \cancel{E}_T > 40 \text{ GeV}, \quad |\Delta\eta_{\ell\ell}| < 3.0, \quad (5.6)$$

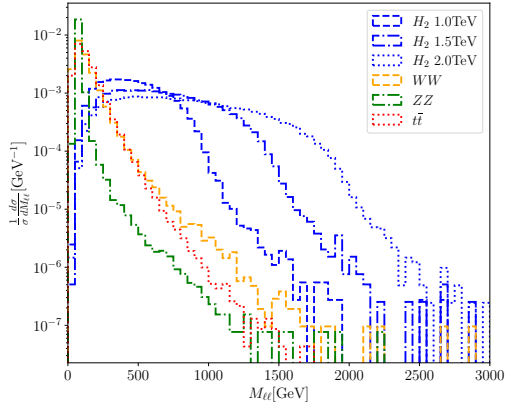
onde $p_{T,\ell}$ e η_ℓ são o momento transverso e pseudo-rapidez dos léptons, respectivamente, enquanto $M_{\ell\ell}$, \cancel{E}_T e $\Delta\eta_{\ell\ell}$ são a massa invariante dos léptons carregados, a energia perdida (*missing energy*) e a distância no plano $\eta \times \phi$ do evento. Os cortes η e $\Delta R_{\ell\ell}$ são impostos pelos limites experimentais dos detetores do LHC. O corte $|\Delta\eta_{\ell\ell}|$ foi imposto para suprimir backgrounds de eventos de fusão de bósons fracos e $M_{\ell\ell}$ para suprimir backgrounds de baixa massa de léptons de $Z\gamma^*$. O número de eventos na tabela 5.1 são o número de eventos pré-corte, portanto o número de eventos real é menor, pois cada classe tem um comportamento diferente após o corte cinemático.

A figura 5.1 mostra algumas distribuições das informações utilizadas para representar os eventos e prever suas classes e $M_{\ell\ell\nu\nu}$, o restante está na figura C.1 do apêndice.

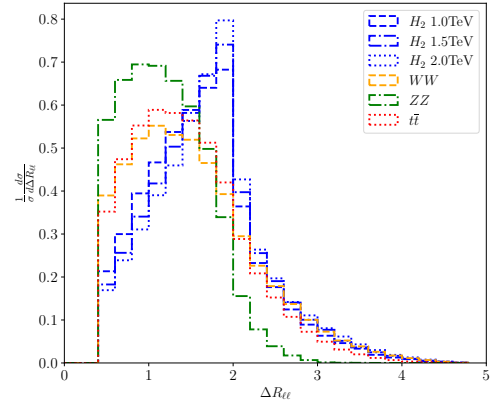
Além da energia e das componentes do 3-momento dos léptons carregados, incluímos os momentos transversos dos léptons, η e $\Delta R_{\ell\ell}$ e as seguintes variáveis na representação dos eventos:

- \cancel{E}_T , Energia perdida (Missing Energy),
- $M_{\ell\ell}$, Massa invariante dos léptons,
- $|\Delta\phi_{\ell\ell}| = |\max(\phi_1, \phi_2) - \min(\phi_1, \phi_2)|$, onde ϕ_1 e ϕ_2 são os ângulos azimutais dos léptons,
- *b*-tag, o número de jatos marcado como jatos bottom [182] para suprimir eventos de $t\bar{t}$,
- $\sqrt{\hat{s}}(0) = \sqrt{E_{\ell\ell}^2 - p_{T,\ell\ell}^2} + \cancel{E}_T$, proposto na ref. [31],
- $\cos\theta^* = \tanh\left(\frac{\Delta\eta_{\ell\ell}}{2}\right)$, proposto na ref. [183],
- $M_{Reco} = \sqrt{M_{\ell\ell}^2 + 2p_{T,\ell\ell}\sqrt{M_{\ell\ell}^2 + p_{T,\ell\ell}^2} - p_{T,\ell\ell}}$ onde $p_{T,\ell\ell}$ é o momento transverso do par de léptons carregados,
-

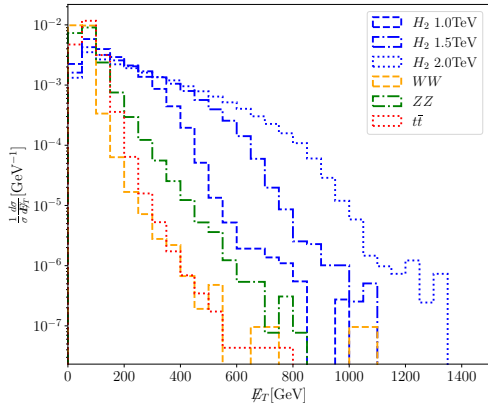
$$\text{Higgsness} = \underset{p_\nu, p_{\bar{\nu}}}{\text{argmin}} \left[\frac{(M_{\ell^+\ell^-\nu\bar{\nu}}^2 - m_H^2)^2}{\delta_H^4} + \min \left(\frac{(M_{\ell^+\nu}^2 - m_W^2)^2}{\delta_W^4}, \frac{(M_{\ell^-\bar{\nu}}^2 - m_W^2)^2}{\delta_W^4} \right) \right], \quad (5.7)$$



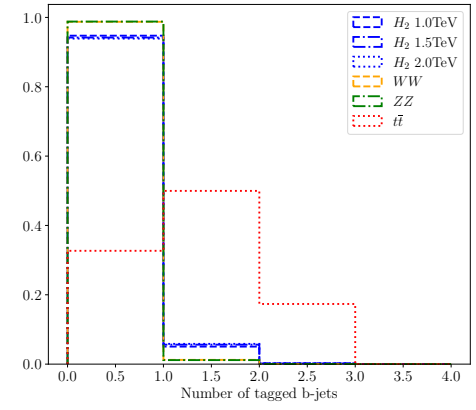
(a) Distribuição da massa invariante dos léptons $M_{\ell\ell}$



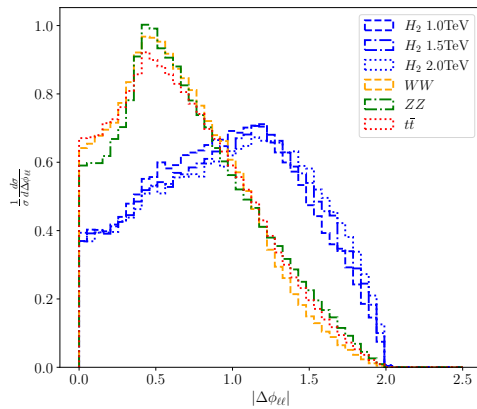
(b) Distribuição de $\Delta R_{\ell\ell}$



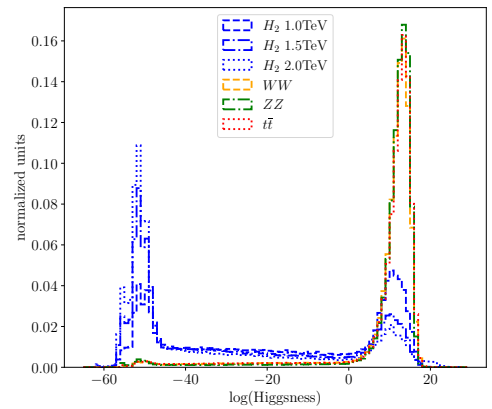
(c) Distribuição de missing energy E_T



(d) Distribuição do b -tag.



(e) Distribuição de $\Delta\phi_{\ell\ell}$



(f) Distribuição de Higgsness

Figura 5.1: Distribuições cinemáticas de algumas informações de alto nível dos backgrounds e sinais para representar os eventos nos algoritmos de ML. O azul representa os sinais H_2 com largura de 10%, laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

proposto na ref. [184].

A variável $\cos \theta^*$ foi proposta para ajudar na medição de spin de sleptons em um modelo supersimétrico [183]. A variável $\sqrt{\hat{s}}(0)$ é proposta como uma variável global, independente do modelo e inclusiva para medidas em uma nova escala de massa em eventos com missing energy nos colisores de hádrons [31]. Além destas variáveis, usamos também o b -tag dos eventos, pois permite uma separação do background $t\bar{t}$ em relação às outras componentes de background e dos sinais. A variável Higgsness tem como origem denunciar a presença do bóson de Higgs decaindo em $W^\pm W^{*\mp} \rightarrow \ell^+ \ell'^- + \nu_\ell \bar{\nu}_{\ell'}$. A ideia é procurar o 4-momento de um evento que minimiza a equação 5.7, onde δ_H e δ_W seriam parâmetros que representam incertezas experimentais, mas no nosso caso tratamos como parâmetros livres com o intuito de maximizar o discernimento entre as classes.

Em geral, as distribuições apresentam o comportamento dos backgrounds diferente dos sinais. Os backgrounds WW e $t\bar{t}$ apresentam um comportamento mais parecido entre si na maioria das distribuições cinemáticas com exceção do b -tag que apresenta uma diferença no comportamento de WW e $t\bar{t}$. As distribuições $M_{\ell\ell}$, $\Delta R_{\ell\ell}$, \cancel{E}_T , M_{Reco} , $\cos \theta^*$ e $\sqrt{\hat{s}}(0)$ são distribuições nos quais o ZZ apresenta um comportamento diferente dos outros backgrounds. Por exemplo, nas distribuições $M_{\ell\ell}$ e M_{Reco} o ZZ é o único que possui um pico, nas distribuições o \cancel{E}_T , $\cos \theta^*$ e $\sqrt{\hat{s}}(0)$ o ZZ possui uma distribuição com comportamento diferente das outras classes e $\Delta R_{\ell\ell}$ o ZZ apresenta um comportamento similar, mas com um máximo maior e um decaimento mais rápido se comparado aos outros backgrounds.

As distribuições $\Delta R_{\ell\ell}$, $\Delta\phi_{\ell\ell}$ e Higgsness apresentam um comportamento do sinal completamente diferente dos backgrounds, em $\Delta R_{\ell\ell}$ os sinais tem um pico em $\Delta R_{\ell\ell} = 2$, $\Delta\phi_{\ell\ell}$ mostra o pico dos backgrounds em 0.5 e o pico dos sinais em 1.5 e em Higgsness os sinais possuem dois picos um em -50 e outro que é menos pronunciado em 10 que é o mesmo lugar do pico dos backgrounds. Enquanto as distribuições $M_{\ell\ell}$, \cancel{E}_T , p_{T1} , p_{T2} , M_{Reco} , $\sqrt{\hat{s}}(0)$ são similares em relação à diferença entre background e sinal, onde os sinais têm um máximo menor e um decaimento mais lento nas distribuições se comparado aos backgrounds. Por último as distribuições η_1 , η_2 e $\cos \theta^*$ os sinais diferem dos backgrounds, mas possuem um comportamento similar que dificulta na separação entre as classes. Assim as informações propostas teoricamente permitem separar os eventos de backgrounds dos sinais.

5.1 Ajuste dos Regressores k NN

Para cada classe, construímos uma função para a distribuição $M_{\ell\nu\nu}$ conforme a equação 4.14. O número de eventos usados para a regressão estão na tabela 5.1. Para garantir que o tamanho do conjunto de dados não tenha influência nos resultados, separamos 80% dos dados para a busca dos melhores hiperparâmetros. A busca dos melhores hiperparâmetros do algoritmo k NN foi realizada com um *grid search* no espaço de hiperparâmetros

$$k \in [1, 5], \quad Dist \in \{\text{Minkowsky, Manhattan, Chebyshev, Canberra, Braycurtis}\}, \\ P \in [1, 8], \quad \text{weight} \in \{\text{uniform, weighted}\}, \quad (5.8)$$

onde o k é o número de vizinhos, $Dist$ é a função distância entre os vizinhos, $weight$ é a opção da média ponderada ou aritmética na equação 4.14 e P a dimensão do espaço latente do PCA. O melhor conjunto de hiperparâmetros são aqueles com o menor erro quadrático médio(MSE) entre a distribuição verdadeira $M_{\ell\nu\nu}$ e a predita. Os melhores hiperparâmetros obtidos estão na tabela 5.2 para cada tipo de processo.

Classe	k	Distância	Peso	Dimensão do espaço latente
WW	1	Chebyshev	Uniform	2
ZZ	1	Minkowski	Uniform	1
$t\bar{t}$	1	Canberra	Uniform	8
H_2 1 TeV(10%)	1	Manhattan	Uniform	1
H_2 1.5 TeV(10%)	1	Canberra	Uniform	3
H_2 2 TeV(10%)	1	Chebyshev	Uniform	6

Tabela 5.2: Tabela mostrando os melhores hiperparâmetros encontrados pela busca em rede para a predição da distribuição $M_{\ell\nu\nu}$.

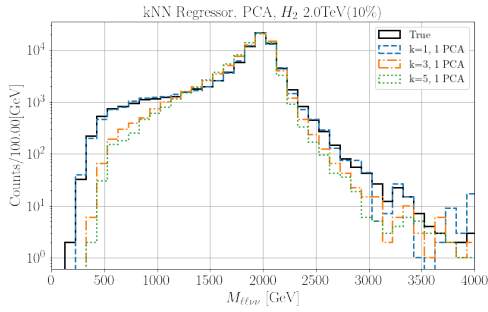
Na figura 5.2 mostramos alguns resultados da busca de hiperparâmetros para k , P e $Dist$ para mostrar a qualidade da regressão do k NN para o sinal de H_2 com massa 2 TeV. Os outros casos estão nas figuras C.10, C.11, C.7, C.8, C.9 do apêndice C. Todos os casos apresentaram um comportamento similar.

O regressor k NN é robusto à variação dos parâmetros enquanto mantém boa precisão nas predições. Em geral, para todos os backgrounds e sinais, o vizinho mais próximo do novo ponto no espaço latente da transformação do PCA é a predição mais precisa do nosso alvo. Testamos algumas alternativas ao k NN como o gradient boosting machine (GBM) [185] e re-

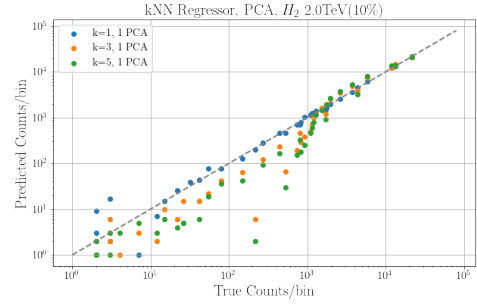
des neurais regressoras. A abordagem de vizinhos mais próximos mostrou ser a superior em aproximar a distribuição da massa verdadeira. Por outro lado, notamos que as redes neurais apresentam um poder de generalização melhor em todas as classes comparado aos outros algoritmos, especialmente o algoritmo k NN que é totalmente dependente da classe do evento. Por exemplo, treinar uma rede neural com o background WW pode ser útil para obter $M_{\ell\ell\nu\nu}$ de outras classes, especialmente os backgrounds. Infelizmente, para eventos de sinal, os regressores treinados com eventos WW não são capazes de generalizar, retornando distribuições muito diferentes das verdadeiras, tendo especial dificuldade em identificar os picos ressonantes. Além disso, de forma geral, notamos que as redes neurais que treinamos não são competitivas, em termos de generalização, comparadas às variáveis simples que correlacionam com a massa da ressonância, como $\sqrt{\hat{s}}(0)$ [31] e outras variáveis de massa transversa. A vantagem significativa de algoritmos com bom desempenho generalizado é ser agnóstico em relação às outras classes, dependendo menos do conhecimento prévio dos diferentes eventos.

O número de dimensão do PCA onde a representação dos dados originais é projetado mostrou uma variação significativa. Enquanto o background ZZ e o Higgs H_2 de 1 TeV, o menor valor de MSE foi obtido com um espaço latente 1-dimensional, o background WW obteve seu melhor desempenho em um espaço 2-dimensional, o Higgs H_2 de 1.5 TeV com um espaço 3-dimensional, o Higgs H_2 de 2 TeV com dimensão 6 PCA e o $t\bar{t}$ com dimensão 8 PCA. A escolha da métrica não tem impacto no desempenho dos algoritmos, consequência do peso uniforme ter um desempenho melhor que o peso ponderado em todos os casos, implicando que a predição é uma média aritmética dos vizinhos mais próximos para dado ponto projetado no espaço das componentes principais dos eventos. Também testamos transformações não-lineares para o espaço latente como o TSNE, mas com ganhos pequenos custando um alto tempo computacional.

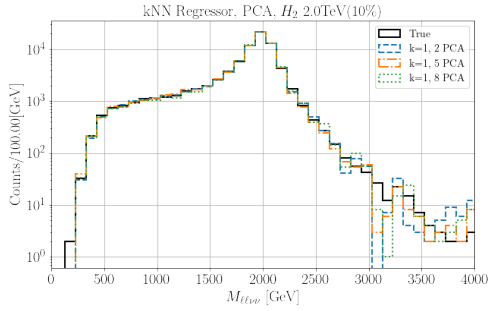
Nas figuras 5.3 e 5.4 mostramos a reconstrução da massa invariante predita e verdadeira para os backgrounds e sinais de H_2 com largura de 10%, respectivamente. Conforme observamos, os regressores funcionam muito bem. A largura das ressonâncias tem pouca influência na precisão da regressão mais detalhes no apêndice B.



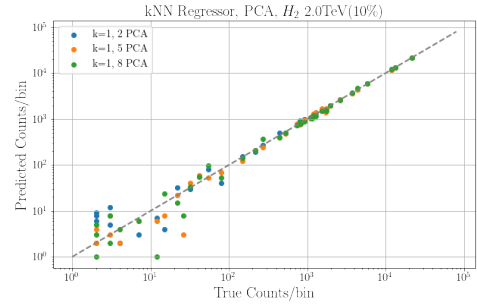
(a) Comparação entre as distribuições da massa invariante do sinal variando o número de vizinhos k . Os valores descritos na legenda são os hiperparâmetros: o k e o número de dimensão do PCA



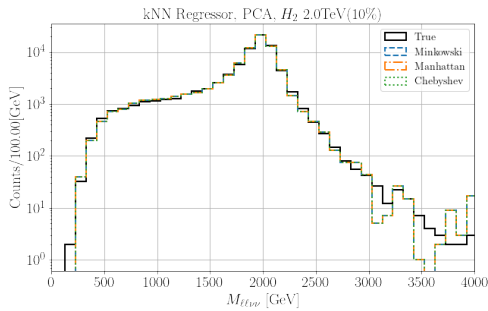
(b) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de vizinhos k .



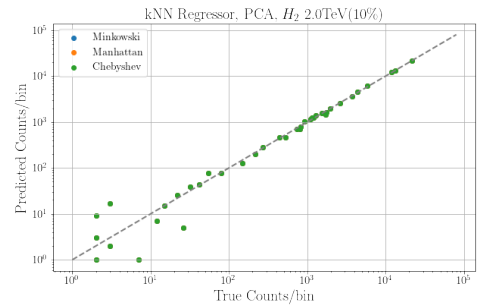
(c) Comparação entre as distribuições da massa invariante do sinal variando o número de dimensão do PCA.



(d) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de dimensão do PCA.

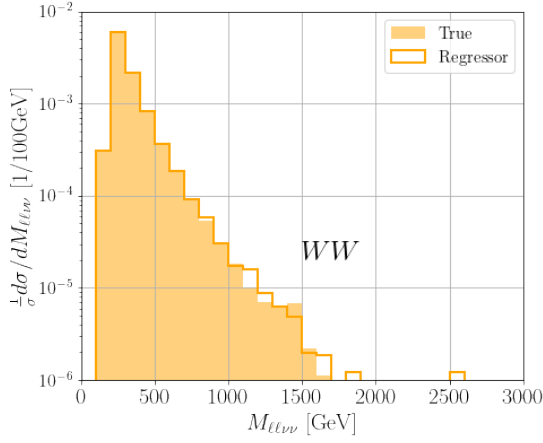


(e) Comparação entre as distribuições da massa invariante do sinal variando a distância utilizada. A legenda mostra as métricas utilizadas.

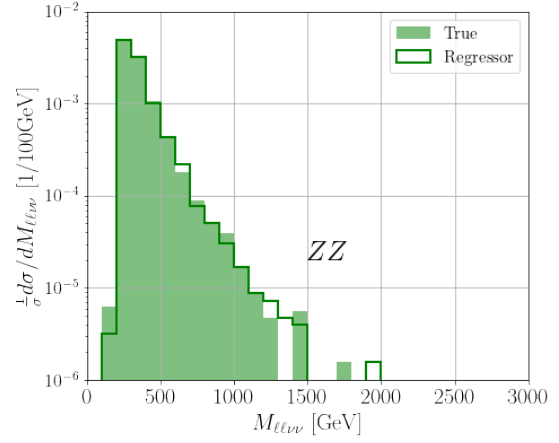


(f) Comparação do número de eventos dentro de um certo bin para os resultados obtidos variando a distância utilizada.

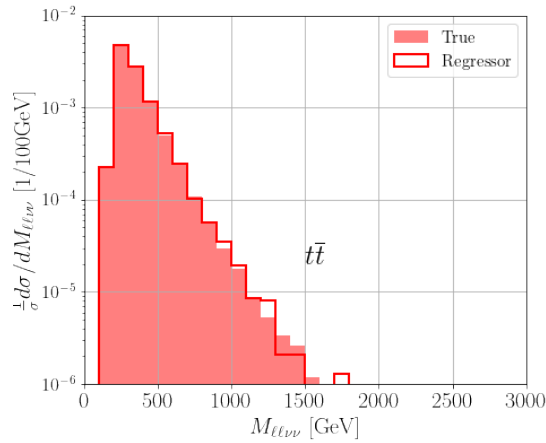
Figura 5.2: Os gráficos 5.2a, 5.2c, 5.2e mostram a comparação da distribuição da massa invariante do sinal variando os hiperparâmetros com a real (linha preta). Os gráficos 5.2b, 5.2d, 5.2f mostram a comparação do número de eventos para dado bin entre o evento verdadeiro (eixo x) e o predito pela regressão (eixo y), idealmente todos os pontos estariam na reta $y = x$ (linha pontilhada e cinza).



(a) Predição de $M_{\ell\ell\nu\nu}$ realizada pela regressão (linha laranja) com a distribuição verdadeira (fundo laranja) da massa invariante $M_{\ell\ell\nu\nu}$ do background WW .

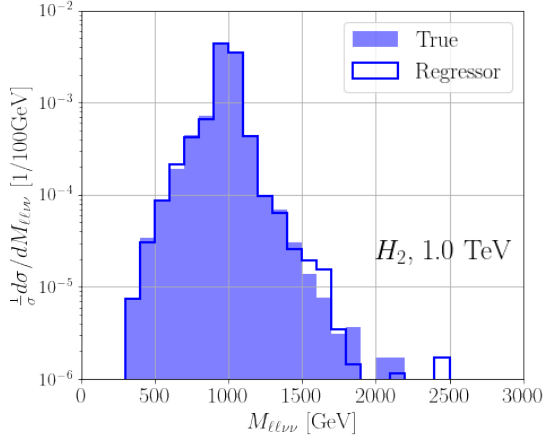


(b) Predição de $M_{\ell\ell\nu\nu}$ realizada pela regressão (linha verde) com a distribuição verdadeira (fundo verde) da massa invariante $M_{\ell\ell\nu\nu}$ do background ZZ .

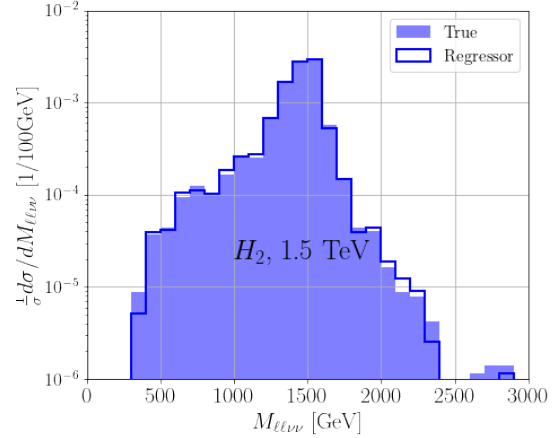


(c) Predição de $M_{\ell\ell\nu\nu}$ realizada pela regressão (linha vermelho) com a distribuição verdadeira (fundo vermelho) da massa invariante $M_{\ell\ell\nu\nu}$ do background $t\bar{t}$.

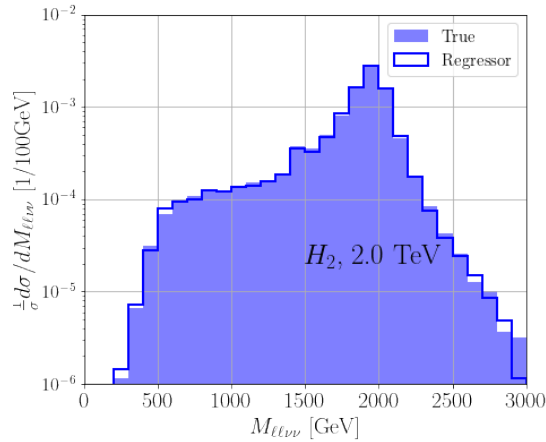
Figura 5.3: Comparação da distribuição de $M_{\ell\ell\nu\nu}$ predita pela regressão com a distribuição verdadeira da massa invariante $M_{\ell\ell\nu\nu}$, onde 5.3a mostra a comparação do background WW , 5.3b do background ZZ e 5.3c do background $t\bar{t}$.



(a) Predição de $M_{\ell\ell\nu\nu}$ realizada pela regressão (linha azul) com a distribuição verdadeira (fundo azul) da massa invariante $M_{\ell\ell\nu\nu}$ para o sinal de 1 TeV e largura de 10%.



(b) Predição de $M_{\ell\ell\nu\nu}$ realizada pela regressão (linha azul) com a distribuição verdadeira (fundo azul) da massa invariante $M_{\ell\ell\nu\nu}$ para o sinal de 1.5 TeV e largura de 10%.



(c) Predição de $M_{\ell\ell\nu\nu}$ realizada pela regressão (linha azul) com a distribuição verdadeira (fundo azul) da massa invariante $M_{\ell\ell\nu\nu}$ para o sinal de 2 TeV e largura de 10%.

Figura 5.4: Comparação da distribuição de $M_{\ell\ell\nu\nu}$ predita pela regressão com a distribuição verdadeira da massa invariante $M_{\ell\ell\nu\nu}$, onde 5.4a mostra a comparação para o sinal de 1 TeV e largura de 10%, 5.4b do sinal de 1.5 TeV e largura de 10% e 5.4c do sinal de 2 TeV e largura de 10%

5.2 Classificação pré-regressão

O k NN construído de certa classe só permite a reconstrução da distribuição de eventos pertinentes àquela classe. Por exemplo, se usamos eventos de sinal em um k NN treinado com background, o k NN encontra o valor da distribuição do background mais próximo do evento de sinal. Para prever corretamente as classes, precisamos conhecer a classe dos eventos com a melhor precisão possível e também a massa da ressonância.

A classificação dos eventos foi realizado com uma rede neural (NN) [124, 125, 135] baseado nas mesmas informações da regressão. A quantidade de dados utilizados estão na tabela 5.1 para ambas as classificações, totalizando 1.75 milhões de eventos para cada classificador. A nossa aplicação de redes neurais são classificações multi-classe, portanto usamos a entropia cruzada como função perda, regularização L_2 , *early stopping* e a possibilidade de usar dropout como técnicas de regularização. O espaço de hiperparâmetro está na tabela 5.3.

Hiperparâmetro	Espaço do hiperparâmetro
Número de camadas	1, 2, 3, 4
Número de nós	32, 64, 96, 128, 160, 256, 384, 512, 1024
Regularização L2	$[1^{-8}, 1^{-4}]$
Taxa de aprendizado inicial	$[1^{-8}, 1^{-2}]$
Uso de dropout	False, True
Taxa de dropout	$[0.0, 0.9]$
Tamanho dos lote	32, 64, 96, 128, 160, 256, 384, 512, 1024
Inicialização dos pesos	Uniform, Normal
Função de ativação	ReLU, Tanh, Softplus

Tabela 5.3: Espaço dos Hiperparâmetros da rede neural.

Utilizamos o `Tensorflow` [149] e `Keras` [150] para construir as redes neurais de classificação multi-classe. A busca da melhor arquitetura e hiperparâmetros foi feito com 30 runs do `Hyperopt` [186, 187] em um conjunto de dados separado apenas para a busca dos melhores hiperparâmetros. O `Hyperopt` é um pacote do Python para a busca dos melhores hiperparâmetros que usa algoritmos de otimização bayesiana como o *Tree-structured Parzen Estimator Approach* (TPE) [188] que foi o algoritmo utilizado.

Uma taxa de aprendizado inicial foi ajustado seguido de uma agenda reduzindo pela metade cada dez épocas. O treino é interrompido se nenhuma melhoria no desempenho no conjunto de validação for observado sobre 20 épocas até um máximo de 100 épocas ser atingido. O modelo com melhor desempenho no conjunto de validação é salvo para o treino dos algoritmos. Começamos mostrando os resultados e análise para os sinais de H_2 com largura de 10%.

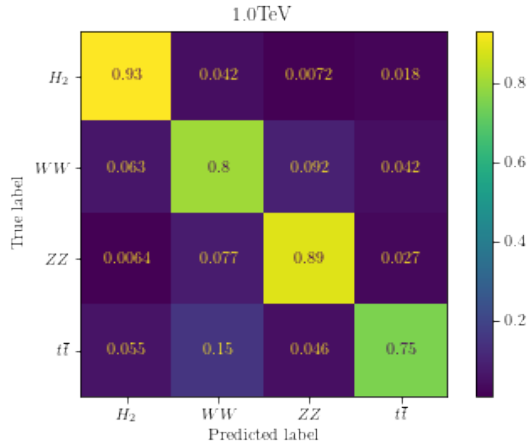
Os melhores hiperparâmetros e arquitetura da rede neural são mostrados na tabela 5.4. A tabela mostra que o sinal de H_2 com 1 TeV possui um modelo com arquitetura menos complexa, uma regularização mais rígida com um valor de regularização L2 maior que os outros casos e o uso de uma camada de dropout com 3% de probabilidade enquanto os outros casos não possuem a camada de dropout mostrando uma maior dificuldade em separar as classes comparando com os sinais de maior massa.

Hiperparâmetro e arquitetura	1 TeV	1.5 TeV	2 TeV
Regularização L2	5.5×10^{-6}	4.7×10^{-8}	3.3×10^{-8}
Learning rate inicial	2×10^{-3}	8.2×10^{-3}	9.6×10^{-3}
Tamanho dos lotes	160	160	128
Taxa de dropout	0.03	–	–
Inicialização dos pesos	normal	uniform	uniform
Função de ativação	tanh	ReLU	tanh
Número de camadas e nós	(64,32,16,8)	(160,80,40,20)	(128,64,32,16)
Total de parâmetros	4188	20544	13748

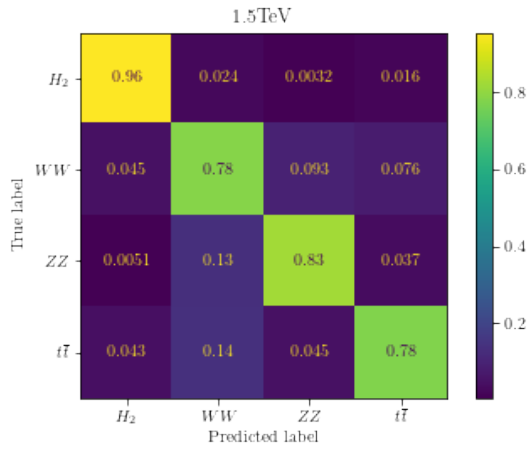
Tabela 5.4: Hiperparâmetros e arquitetura da rede neural para classificação dos sinais de 1, 1.5, 2 TeV dos bósons de Higgs. Os casos de 1.5 e 2 TeV não foi necessária camadas de dropout.

Na figura 5.5 mostramos a matriz confusão do classificador da rede neural(NN_1) treinado para reconhecer sinais de H_2 com largura de $\Gamma_H = 10\%$ contra eventos de WW , ZZ e $t\bar{t}$ e o resultado dos scores está na figura 5.6. Conforme esperado, eventos de WW e $t\bar{t} \rightarrow W^+W^- + b\bar{b}$ são confundidos pelo classificador, com cerca de 14% de $t\bar{t}$ considerados como WW e a quantidade de eventos de WW considerados como $t\bar{t}$ aumentam quanto maior a massa de H_2 . Conforme dito anteriormente, o comportamento de WW e $t\bar{t}$ são similares nas figuras 5.1 e C.1 e isso se reflete na matriz confusão. Em contrapartida, cerca de 1/3 de eventos de $t\bar{t}$ não possuem b -jatos marcados, o discriminante principal contra a produção do par WW . Essa similaridade pode ser resumida pela figura 5.6 onde todos os casos apresentam distribuições de scores de WW e $t\bar{t}$ se sobrepondo.

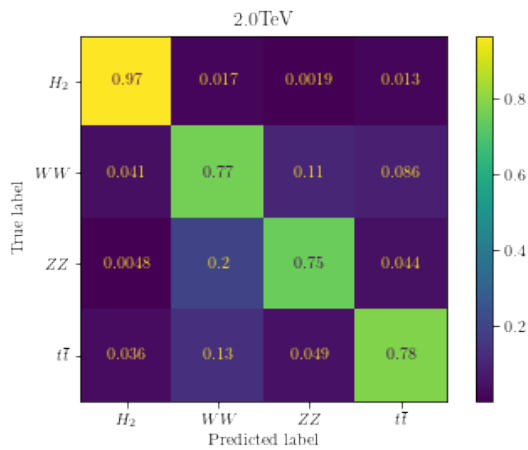
Entretanto, a classe mais confundida foi o ZZ , onde 7.7%, 13% e 20% foram classificados como WW para os casos de H_2 de 1 TeV, 1.5 TeV e 2 TeV, respectivamente. Por outro lado, cerca de 3 a 4% dos eventos de sinais foram classificados como backgrounds. A quantidade de $ZZ(\gamma^*)$ classificados de maneira errada podem ser explicados com a inclusão de distribuições como $|\Delta\phi_{\ell\ell}|$ e Higgsness como informações de alto nível. Conforme visto nas figuras 5.1e e 5.1f, enquanto essas informações são poderosas em discernir sinais de backgrounds, os backgrounds são todos similares. Por exemplo, no caso de sinal de H_2 com 2 TeV retirar Higgsness



(a) Matriz confusão da classificação NN_1 para o sinal de 1 TeV e largura de 10%.



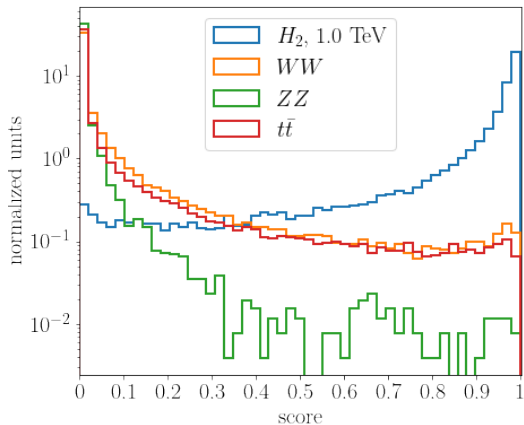
(b) Matriz confusão da classificação NN_1 para o sinal de 1.5 TeV e largura de 10%.



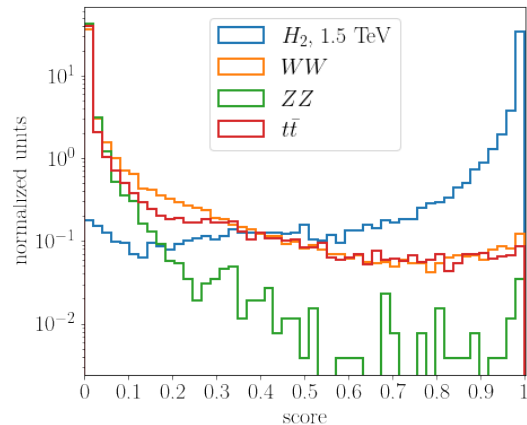
(c) Matriz confusão da classificação NN_1 para o sinal de 2 TeV e largura de 10%.

Figura 5.5: Matriz confusão do segundo classificador NN_1 para os sinais de H_2 com 1.5 TeV 5.5a, 1.5 TeV 5.5b e 2 TeV 5.5c.

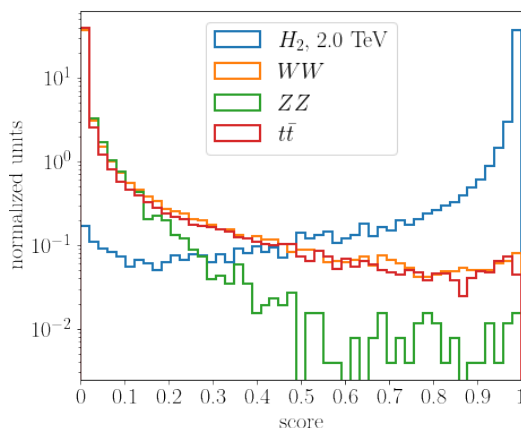
fez a taxa de verdadeiro positivo diminuir de 97% para 94%, enquanto a proporção de eventos de ZZ considerados como WW foi de 19% para 6%.



(a) Scores da classificação NN_1 para o sinal de 1 TeV e largura de 10%.



(b) Scores da classificação NN_1 para o sinal de 1.5 TeV e largura de 10%.

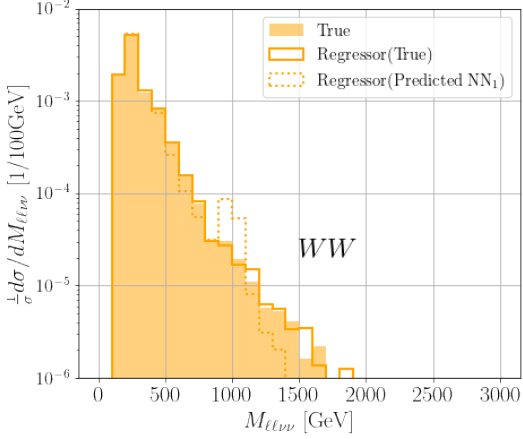


(c) Scores da classificação NN_1 para o sinal de 2 TeV e largura de 10%.

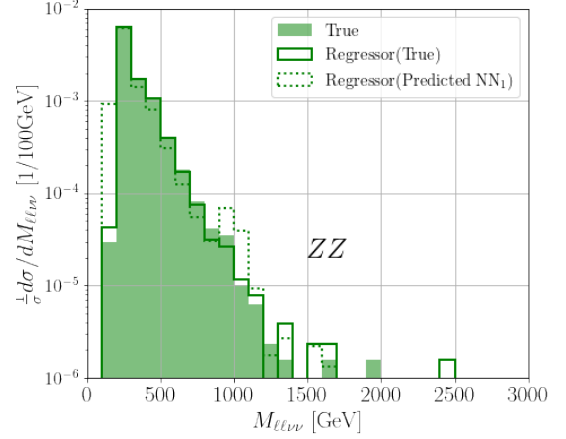
Figura 5.6: Probabilidade de cada classe ser sinal utilizando a rede neural como classificador. Valores de score próximos de 1 representam eventos com probabilidade de serem sinais e valores de score próximos de 0 representam eventos com probabilidade de serem background. Em todos os painéis, a linha azul representa o sinal, a linha laranja representa o background WW , a linha verde representa o background ZZ e a linha vermelha representa o background $t\bar{t}$.

Com o classificador NN_1 em mãos, podemos reconstruir a massa $M_{\ell\ell\nu\nu}$ dos eventos. Enfatizamos ser necessário conhecer a classe do evento antes da regressão visto que a variável alvo só pode ser corretamente estimado quando interpolado sobre um conjunto de dados apropriado do algoritmo k NN. Em outras palavras, o regressor não generaliza de uma classe a outra. Se apresentamos uma instância nunca constatados pelo regressor, a falta de correlações necessárias irá apresentar um resultado sem sentido. Por exemplo, no referencial de repouso do

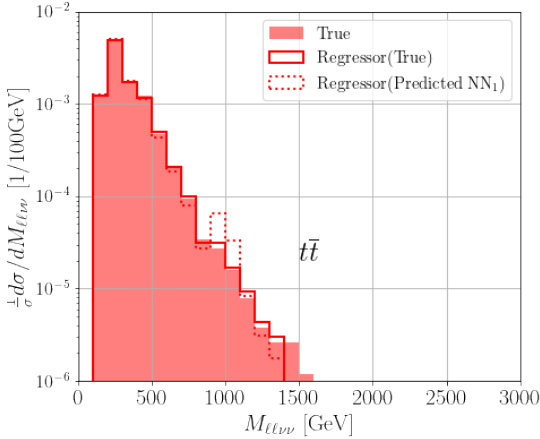
Higgs, a soma da energia dos léptons carregados e a energia dos neutrinos é igual à massa do Higgs, $E_{\ell\ell}^* + E_{\nu\nu}^* = m_H$. Nesse caso, o regressor só consegue aprender a simples relação $E_{\nu\nu}^* = m_H - E_{\ell\ell}^*$ para recuperar a informação perdida a partir do evento observado se é treinado com eventos de sinal com m_H conhecido.



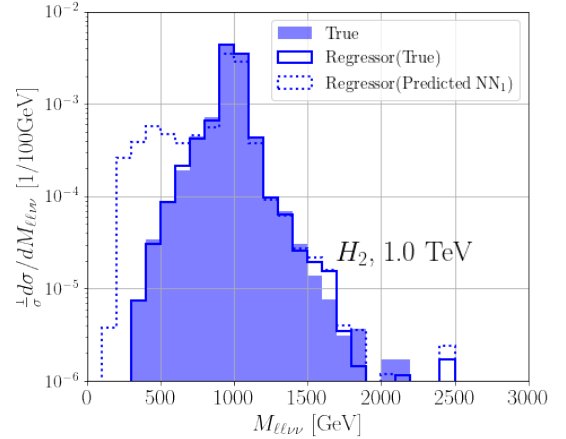
(a) Distribuição $M_{\ell\nu\nu}$ do background WW .



(b) Distribuição $M_{\ell\nu\nu}$ do background ZZ .



(c) Distribuição $M_{\ell\nu\nu}$ do background $t\bar{t}$.



(d) Distribuição $M_{\ell\nu\nu}$ do sinal H_2 de 1 TeV.

Figura 5.7: Gráficos mostrando a distribuição $M_{\ell\nu\nu}$ verdadeira(fundo), com a regredida a partir dos dados reais(linha sólida) e a regredida com os dados identificados com NN_1 (linha pontilhada). O azul representa o sinal de 1 TeV, laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

Na figura 5.7, mostramos a massa $M_{\ell\nu\nu}$ predita para os eventos classificados com uma rede neural para o H_2 de 1 TeV. Novamente, os resultados para outras massas e larguras são idênticas, ver apêndices C e B. Notamos contaminação dos eventos de sinais nas caudas das distribuições de WW e $t\bar{t}$. O que é esperado dado que 1.7% e 1.4% de eventos de sinais são classificados como WW e $t\bar{t}$, respectivamente. Apenas 0.19% de eventos de H_2 foram classificados como ZZ , por isso não observamos um pico na cauda da distribuição de ZZ . O oposto é que 4.1%

de WW e 3.7% de $t\bar{t}$ de eventos são considerados como evento de sinal, populando os bins de baixa massa na distribuição de H_2 acima da distribuição verdadeira. Na prática, se alguém está interessado em identificar o bóson de Higgs, exigir um valor de score acima de 0.5(ou maior) é eficiente para atenuar a contaminação da distribuição dos backgrounds permitindo uma estimativa confiável na região da ressonância.

Uma maneira de contornar as contaminações para melhorar o desempenho nas estimativas de massa é apresentado na próxima seção.

5.3 Classificação pós-regressão

Como podemos se livrar da contaminação nas distribuições dos backgrounds e sinal? Na referência [189], uma ensemble de classificadores foi usado para aumentar a *accuracy* da classificação de eventos de bóson de Higgs com um desempenho similar ao de uma rede neural profunda [190]. Usamos a mesma ideia para aumentar o desempenho do nosso classificador empilhando outra rede neural em cima da primeira classificação descrita na seção anterior. Para um review de métodos de ensemble, veja [191].

Mostramos um fluxograma do algoritmo proposto do começo ao fim na figura 5.8. O conjunto de dados \mathbf{X} , composto pelas informações descritos no início do capítulo são separados em vários subconjuntos para treinar/validar o classificador e a regressão. Dois subconjuntos são usados para treinar o primeiro classificador, descrito como NN_1 na figura 5.8 e o regressor kNN . Nesse esquema, o regressor recebe o conjunto de dados mais a saída de NN_1 para decidir quais dados devem usados para calcular $M_{\ell\nu\nu}$ para dado evento. Após esse estágio, o algoritmo produziu duas peças importantes que serão anexadas a \mathbf{X} : o vetor de scores \mathbf{p} e $M_{\ell\nu\nu}$ resultando em uma nova representação \mathbf{X}' . Essa nova representação é então usada para treinar uma segunda rede neural, NN_2 . A combinação do kNN com classificadores de redes neurais, chamamos o algoritmo de $kNNNN$. Notamos que a saída de NN_2 é a saída final do algoritmo $kNNNN$.

Nas figuras 5.9 e 5.10 mostramos a matriz confusão e os scores de NN_2 , respectivamente. A separação de classes foi melhorado após a segunda classificação. Para confirmar a melhoria calculamos a *accuracy* e a assimetria do score definido como

$$\frac{N(\text{score} > 0.5) - N(\text{score} < 0.5)}{N(\text{score} > 0.5) + N(\text{score} < 0.5)}, \quad (5.9)$$

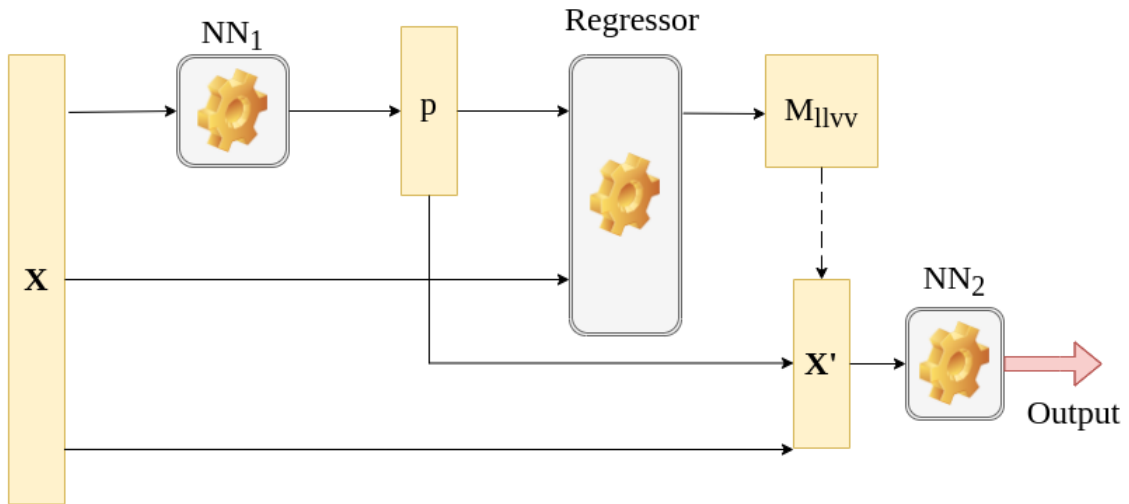


Figura 5.8: Fluxograma mostrando a aplicação dos algoritmos no projeto. A linha pontilhada indica que NN_2 pode ser utilizado sem a massa predita. No texto NN_2 irá se referir a NN_2 com a massa predita, mas existem figuras onde utilizamos NN_2 sem a massa predita e NN_2 com a massa predita a diferença pode ser identificados na legenda da figura.

onde N é o número de eventos da classe.

Também calculamos a razão de verossimilhança positivo e negativo, definido na referência [192]

$$LR_+ = \frac{\text{Sensitividade}}{1 - \text{Especificidade}} = \frac{\text{Sensitividade}}{\text{Taxa de Falso Positivo}}, \quad (5.10)$$

$$LR_- = \frac{1 - \text{Sensitividade}}{\text{Especificidade}} = \frac{\text{Taxa de Falso Negativo}}{\text{Especificidade}}. \quad (5.11)$$

Essas métricas têm o objetivo de medir o quão efetivo o classificador é em prever as classes em um problema binário. Sensitividade é a razão entre o número de eventos classificados corretamente como positivos (VP) e o número total de eventos classificados como positivo ($P=VP+FP$), mede a capacidade do classificador em identificar a classe positiva, eventos de H_2 . Especificidade, por outro lado, é a razão entre o número de eventos corretamente classificados como negativos (VN) e o número total de eventos classificados como negativos, os backgrounds ($N=VN+FN$). Para usar essas métricas, consideramos todos os backgrounds como uma única classe. Analogamente à sensibilidade, a especificidade mede o quão competente o classificador é em identificar instâncias negativas.

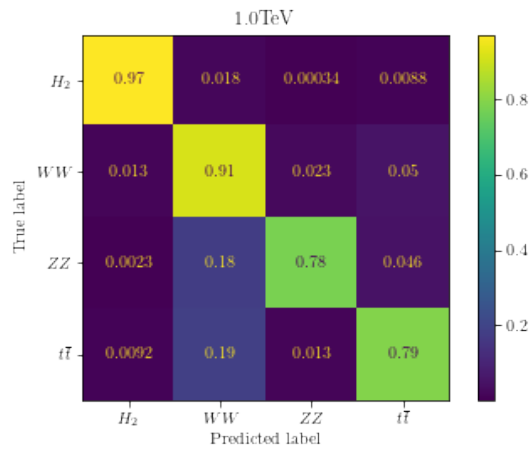
A partir da sensibilidade e especificidade também podemos definir a *receiver-operator curve* (ROC) como um gráfico que mostra a variação do limiar de discriminação para um classificador binário, onde o eixo- x representa a taxa de falso positivo (1-Especificidade) e o eixo-

y representa a sensibilidade. Assim podemos interpretar o eixo- x como a aceitação do sinal/background e o eixo- y como a rejeição do background/sinal. No classificador multi classe uma possível generalização da curva ROC é realizar a abordagem (*one-vs-one*) [193], onde a curva ROC é calculada para todas as combinações de duas classes, onde se exclui as outras classes como se fosse uma classificação binária. No nosso caso, fizemos o sinal H_2 pelos backgrounds, obtendo H_2 vs WW , H_2 vs ZZ e H_2 vs $t\bar{t}$. Assim, temos gráficos mostrando como aceitação do sinal H_2 pela rejeição do background específico. Por último, a área debaixo da curva ROC é definido como a área sobre a curva (*AUC, area under curve*). Quanto mais próximo à área é igual a 1, mais próximo o classificador é do ideal, ou seja, é uma quantificação da curva ROC que indica a proximidade com um algoritmo ideal.

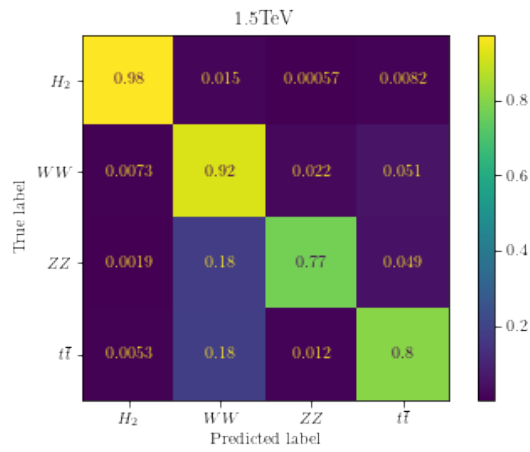
Para os sinais, LR_+ resume o quão mais provável os sinais são corretamente preditos sinais do que backgrounds serem preditos erradamente como sinais. Por outro lado, LR_- resume quantas vezes é menos prováveis sinais serem erradamente preditos como backgrounds do que eventos de backgrounds corretamente preditos como background. Um melhor classificador deve, portanto, maximizar LR_+ e minimizar LR_- . Na comparação de dois classificadores, por exemplo, NN_1 e NN_2 , se $LR_+(NN_2) > LR_+(NN_1)$ e $LR_-(NN_2) < LR_-(NN_1)$, então NN_2 é melhor que NN_1 na confirmação de ambos positivos e negativos. Quando a desigualdade da primeira condição é mantido, mas a segunda desigualdade muda, então NN_2 é melhor que NN_1 na confirmação de positivos, mas pior na de negativos. O contrário acontece quando as condições do caso anterior são trocadas, NN_2 é melhor que NN_1 na confirmação de negativos, mas pior na de positivos.

Na tabela 5.5, mostramos a *accuracy*, a assimetria e a razão de probabilidade positivo e negativo e a área sobre a curva para os três bósons de Higgs investigados. Todas as métricas apontam uma melhoria significativa de NN_2 sobre NN_1 , mas o ganho mais pronunciado é o caso de 1 TeV. Massas menores apresentam atributos menos discerníveis do background, então aproveitam mais de uma ensemble de classificadores que usam informações distintas como o score das classificações e a massa $M_{\ell\ell\nu\nu}$. A melhoria é mais significativa para os sinais e background WW comparado a $ZZ(\gamma^*)$ e $t\bar{t}$, comportamento verificado na figura 5.11, a diferença entre a matriz confusão de NN_1 e NN_2 .

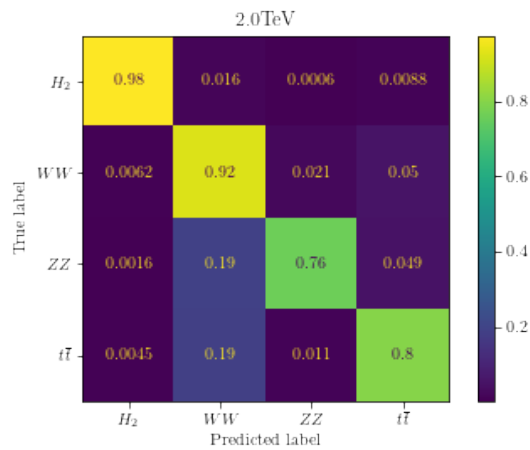
Primeiramente, queremos que as diagonais na figura 5.11 sejam positivo, implicando que NN_2 aumenta a taxa de verdadeiro positivo comparado a NN_1 . Ao mesmo tempo, entradas não-diagonais negativas implicam uma classificação errada menor entre as classes. Em geral,



(a) Matriz confusão da classificação NN_2 para o sinal de 1 TeV e largura de 10%.

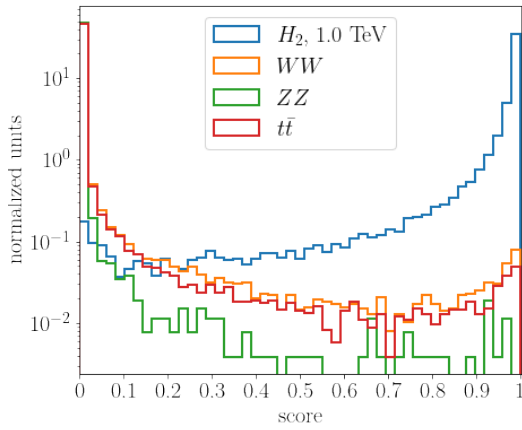


(b) Matriz confusão da classificação NN_2 para o sinal de 1.5 TeV e largura de 10%.

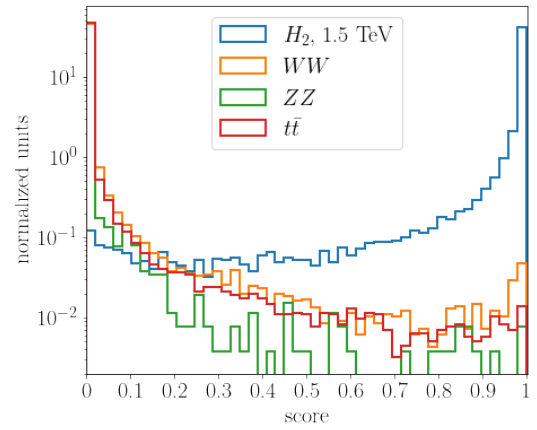


(c) Matriz confusão da classificação NN_2 para o sinal de 2 TeV e largura de 10%.

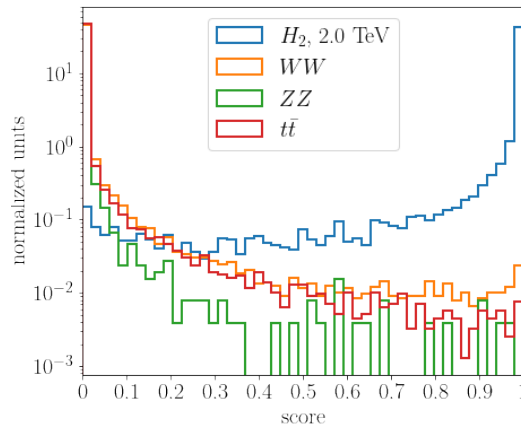
Figura 5.9: Matriz confusão do segundo classificador NN_2 para os sinais de H_2 com 1 5.9a, 1.5 5.9b e 2TeV 5.9c.



(a) Scores da classificação NN_2 para o sinal de 1 TeV e largura de 10%.



(b) Scores da classificação NN_2 para o sinal de 1.5 TeV e largura de 10%.



(c) Scores da classificação NN_2 para o sinal de 2 TeV e largura de 10%

Figura 5.10: Probabilidade de cada classe ser sinal utilizando a rede neural como classificador. Valores de score próximos de 1 representam eventos com probabilidade de serem sinais e valores de score próximos de 0 representam eventos com probabilidade de serem background. Em todos os painéis, a linha azul representa o sinal, a linha laranja representa o background WW , a linha verde representa o background ZZ e a linha vermelha representa o background $t\bar{t}$.

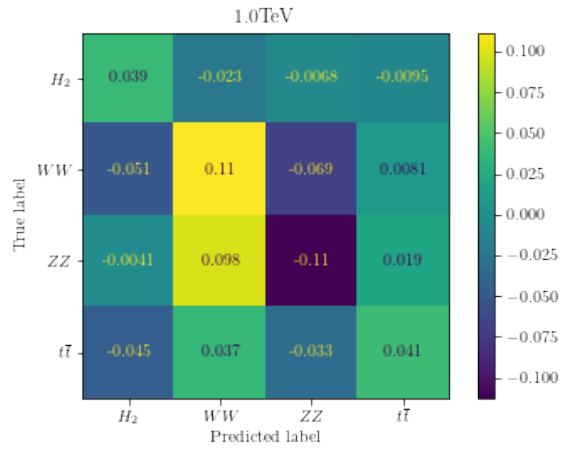
Métrica	1 TeV		1.5 TeV		2 TeV	
	NN ₁	NN ₂	NN ₁	NN ₂	NN ₁	NN ₂
<i>accuracy</i>	81.1%	87%	81.3%	87.6%	80.6%	87.5%
assimetria, H_2	0.821	0.933	0.887	0.944	0.917	0.941
assimetria, WW	-0.903	-0.978	-0.933	-0.988	-0.939	-0.991
assimetria, ZZ	-0.992	-0.996	-0.994	-0.997	-0.994	-0.999
assimetria, $t\bar{t}$	-0.916	-0.984	-0.934	-0.992	-0.947	-0.994
LR_+	28.225	74.371	45.299	86.010	61.330	82.360
LR_-	0.603	0.581	0.589	0.578	0.582	0.578
$AUC(H_2 \times WW)$	0.9813	0.9977	0.9915	0.9988	0.9940	0.9991
$AUC(H_2 \times ZZ)$	0.9965	0.9994	0.9975	0.9996	0.9975	0.9996
$AUC(H_2 \times t\bar{t})$	0.9822	0.9984	0.9929	0.9992	0.9950	0.9993

Tabela 5.5: Comparação do desempenho das métricas obtidas das classificações NN₁ e NN₂ para identificar o bóson de Higgs nas três massas consideradas.

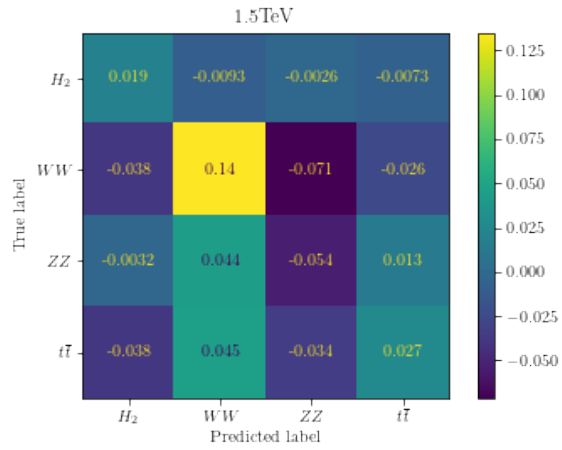
considerar o resultado das três massas de H_2 , vemos uma melhoria de NN₂ comparado a NN₁. Exceto a classe $ZZ(\gamma^*)$ nos casos de 1 e 1.5 TeV, todas as diagonais são positivas com uma grande melhoria na classificação de WW . Além disso, o sinal de 1 TeV se beneficiou mais de NN₂ do que as massas mais pesadas. Essa é uma característica positiva do k NNNN, ajudando nos casos com sinais difíceis. Em relação às entradas não-diagonais, observamos uma clara tendência, a classe $ZZ(\gamma^*)$ são identificados com maior precisão nos modelos cujo objetivo é separar Higgs mais pesados. Em contraste, outras classes são confundidas em menor proporção em NN₂. A consequência de ter as classes WW e $ZZ(\gamma^*)$ mais precisas é a identificação errada dos eventos de $t\bar{t}$ como WW .

A hipótese da significância estatística do sinal depende aproximadamente da fração de eventos de background rejeitados, $r_B = 1 - \epsilon_B$ por uma taxa fixa de aceitação de sinal, ϵ_s , $\epsilon_s/\sqrt{1 - r_B}$. Na figura 5.12) mostramos as curvas ROC de H_2 de 1 TeV e $\Gamma_H/m_H = 0.1$ contra os backgrounds na abordagem "one-vs-one" [193]. Observamos que para uma taxa fixa de aceitação de H_2 , a rejeição dos backgrounds aumenta na classificação NN₂ comparado a NN₁, ou seja, a rejeição é maior quando incluímos a massa $M_{\ell\nu\nu}$ regredida na representação dos dados. O AUC que resume a curva ROC, pode ser constatada nas últimas linhas da tabela 5.5 para às três massas de H_2 . O AUC aumenta de NN₁ para NN₂ em todos os casos, seja para outras massas e larguras.

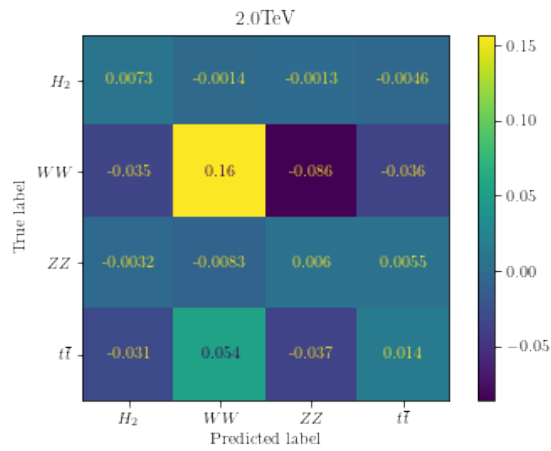
Após a segunda classificação, usando os scores das classes de NN₁ e a massa predita pelo regressor, a segunda rede neural fornece uma predição mais precisa para informar o regressor que conjunto de dados usar para a regressão. O resultado é que a contaminação das outras



(a) Diferença entre a matriz confusão de NN_1 com NN_2 para o sinal de 1 TeV e largura de 10%.

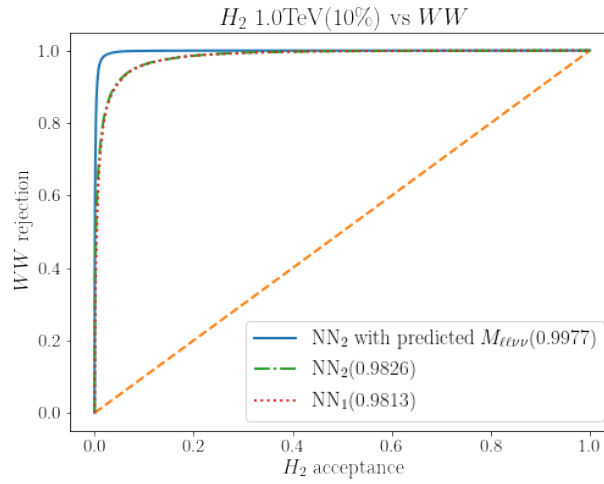


(b) Diferença entre a matriz confusão de NN_1 com NN_2 para o sinal de 1.5 TeV e largura de 10%.

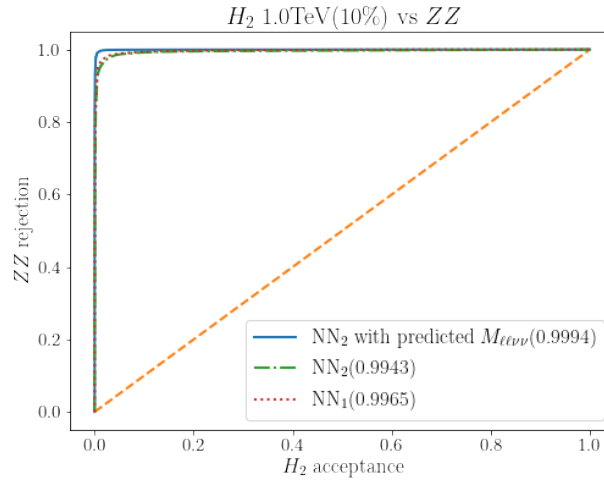


(c) Diferença entre a matriz confusão de NN_1 com NN_2 para o sinal de 2 TeV e largura de 10%.

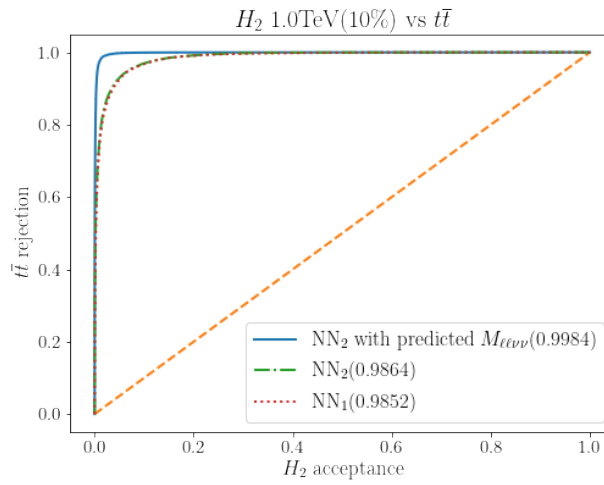
Figura 5.11: Diferença entre a matriz confusão de NN_1 com o NN_2 , ou seja, $NN_2 - NN_1$.



(a)

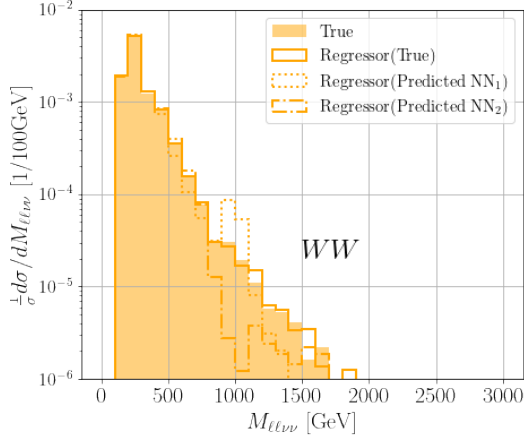


(b)

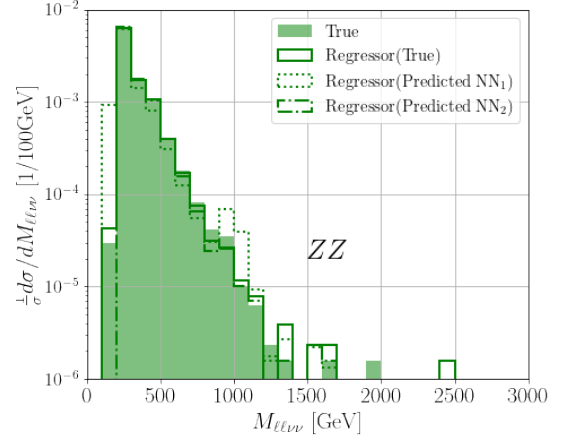


(c)

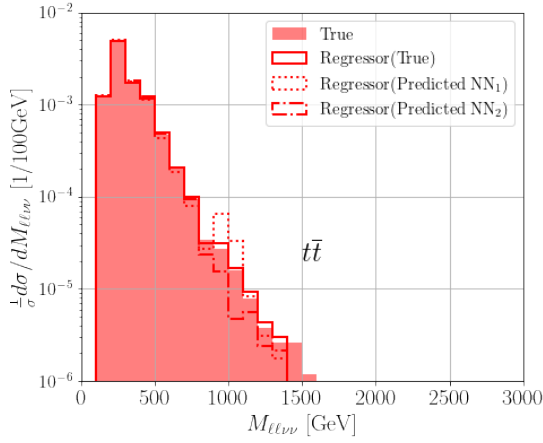
Figura 5.12: As curvas ROC do H_2 de 1 TeV (com $\Gamma_H/m_H = 0.1$) contra as classes de background WW , ZZ , e $t\bar{t}$. Nas legendas, em parêntesis, mostramos o AUC correspondente da primeira classificação, NN_1 e da segunda classificação, NN_2 , com e sem incluir a massa predita.



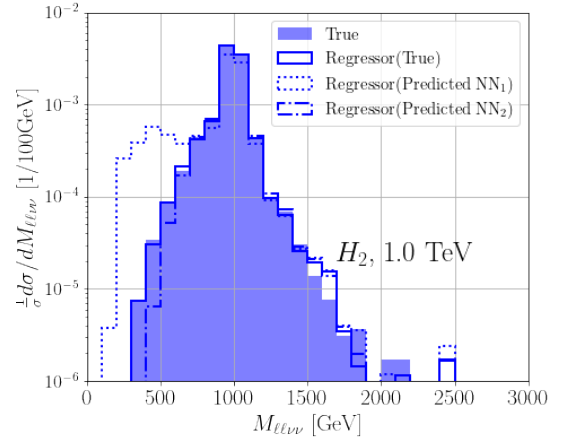
(a) Distribuição $M_{\ell\nu\nu}$ do background WW .



(b) Distribuição $M_{\ell\nu\nu}$ do background ZZ .



(c) Distribuição $M_{\ell\nu\nu}$ do background $t\bar{t}$.



(d) Distribuição $M_{\ell\nu\nu}$ do sinal H_2 de 1 TeV.

Figura 5.13: Gráficos mostrando a distribuição $M_{\ell\nu\nu}$ verdadeira(fundo), com a regredida a partir dos dados reais(linha sólida), a regredida com os dados identificados com NN_1 (linha pontilhada) e a regredida com os dados identificados com NN_2 (linha semi-pontilhada). O azul representa o sinal de 1 TeV, laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

classes é reduzido e a predição de $M_{\ell\nu\nu}$ melhora. Mostramos a massa invariante $M_{\ell\nu\nu}$ após a segunda classificação do sinal de H_2 com 1 TeV na figura 5.13, as outras massas estão no apêndice C.

5.4 Melhorias na Significância Estatística do Sinal

Agora que estabelecemos um algoritmo para predizer a massa $M_{\ell\nu\nu}$, queremos investigar se ele pode aumentar a significância estatística de descoberta [194] do sinal. A significância estatística é calculada de acordo com

$$N_\sigma = \frac{\epsilon_{cut}^{(S)} \times N_S}{\sqrt{\sum_i \epsilon_{cut}^{(i)} \times N_{B_i} + (\epsilon_B \times \sum_i \epsilon_{cut}^{(i)} \times N_{B_i})^2}}, \quad (5.12)$$

onde N_S e N_{B_i} , $i = WW, ZZ, t\bar{t}$ denotam o número de eventos de sinais e backgrounds, respectivamente; $\epsilon_{cut}^{(S)}$ e $\epsilon_{cut}^{(i)}$, $i = WW, ZZ, t\bar{t}$ denotam a eficiência dos cortes (ambos nas variáveis cinemáticas e na saída dos scores) de sinal e background, respectivamente; finalmente, ϵ_B representa a incerteza sistemática da taxa dos backgrounds assumindo, por simplicidade, uma incerteza comum para todas as fontes de background.

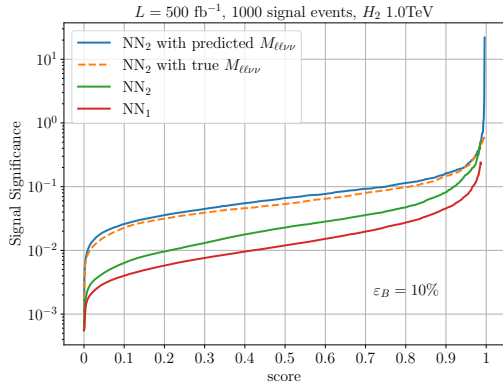
As seções de choque de produção de WW , ZZ e $t\bar{t}$, em Leading Order, são 102.8, 14.15 e 674.1 picobarn, respectivamente. Os branching ratios para $W \rightarrow \ell\nu$, $Z \rightarrow \ell^+\ell^-$, $Z \rightarrow \nu\bar{\nu}$ e $t \rightarrow bW^-$ são 10.68%, 3.37%, 20% e 100%, respectivamente. A eficiência dos cortes foi calculada para os cortes dados na Eq. 5.6, assumindo uma luminosidade integrada de 500 fb^{-1} e estimamos 2.35×10^6 , 1.91×10^5 , 1.53×10^7 eventos para WW , ZZ e $t\bar{t}$, respectivamente, totalizando 1.8×10^7 eventos de background no LHC de 13TeV. Incluir correções de NLO QCD devem aumentar por algumas dezenas de por cento esses números, tanto os backgrounds quanto os sinais, porém não esperamos que mudem o comportamento dos classificadores e regressores desde que o treino dos algoritmos seja realizado com eventos que incluam tais correções. Fixamos o número de eventos de sinal em 1000 para todas as massas para fins de ilustração. A real seção de choque de produção do sinal é dependente do modelo específico, no caso relacionado à massa e à largura de H_2 .

Conforme discutido anteriormente, estamos interessados em mostrar o aumento da significância do sinal que o algoritmo proposto é esperado produzir ao incluir a massa $M_{\ell\ell\nu\nu}$ predita na representação dos dados. O ganho da significância é definido como

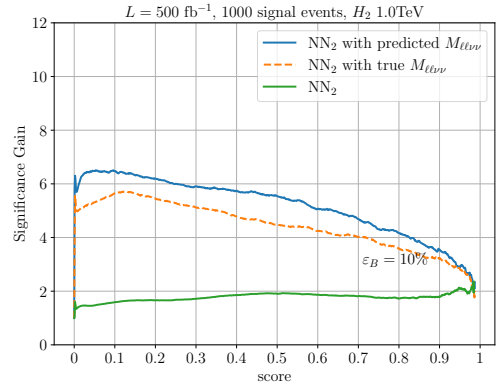
$$\text{Significance Gain} = \frac{N_\sigma(k\text{NNNN})}{N_\sigma(\text{NN}_1)}, \quad (5.13)$$

onde $N_\sigma(k\text{NNNN})$ é a significância estatística após usar NN_2 para a classificação com ou sem a inclusão da massa $M_{\ell\ell\nu\nu}$ predita como informação. Além disso, também queremos verificar se as massas preditas causam uma subestimação ou uma sobrestimação da significância estatística comparado ao que obteríamos se soubéssemos a distribuição de $M_{\ell\ell\nu\nu}$ verdadeiro.

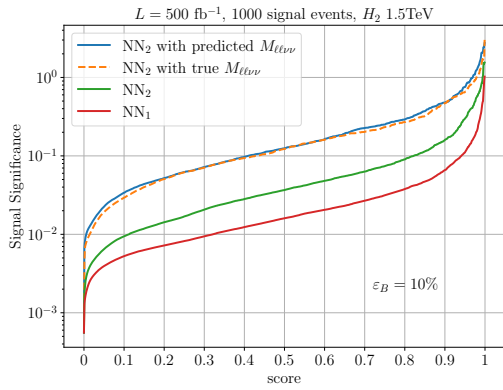
Na figura 5.14, mostramos nos gráficos da esquerda a significância estatística assumindo uma incerteza sistemática $\epsilon_B = 10\%$ na taxa dos backgrounds, para um novo bóson de Higgs de 1, 1.5 e 2 TeV, da linha superior a inferior, respectivamente. Para aumentar a significância,



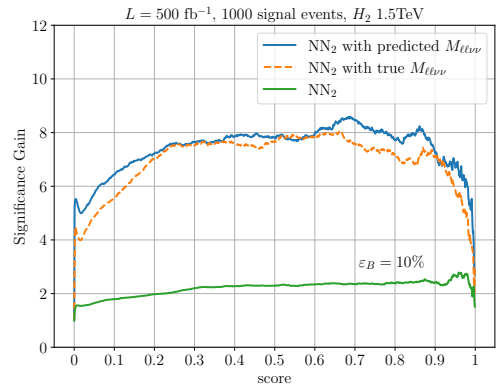
(a) Significância estatística de descoberta para o sinal H_2 de massa 1 TeV.



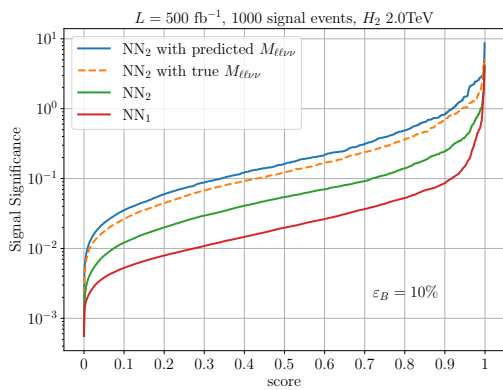
(b) Ganho da significância estatística de descoberta para o sinal H_2 de massa 1 TeV.



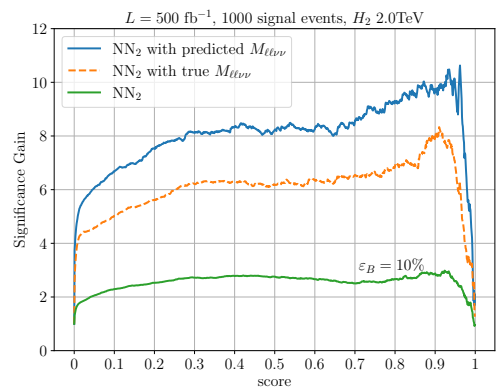
(c) Significância estatística de descoberta para o sinal H_2 de massa 1.5 TeV.



(d) Ganho da significância estatística de descoberta para o sinal H_2 de massa 1.5 TeV.



(e) Significância estatística de descoberta para o sinal H_2 de massa 2 TeV.



(f) Ganho da significância estatística de descoberta para o sinal H_2 de massa 2 TeV.

Figura 5.14: A significância estatística do k NNNN algoritmo como uma função do corte na saída de scores (gráficos da esquerda) e do ganho da significância, eq.5.13, comparado a classificação NN_1 (gráficos da direita) para os sinais de H_2 com 1 TeV (primeira linha), 1.5 TeV(segunda linha e 2 TeV(terceira linha). A incerteza sistemática é fixo em 10%.

cortamos o score de sinal representado pelo eixo horizontal do gráfico. O "1st NN" e "2nd NN" representam a significância de NN_1 e NN_2 sem incluir a predição $M_{\ell\ell\nu\nu}$, respectivamente. Nos três sinais de H_2 a significância estatística de NN_2 com ou sem a massa predita é melhor que NN_1 , o ganho de NN_2 sem a massa predita é em torno de 2 e o ganho de NN_2 com a massa predita é em torno de 6, 8 e 10 para massas de 1, 1.5 e 2 TeV, respectivamente. O que torna NN_2 com massa predita um ótimo método para selecionar eventos de H_2 . Entretanto, existe uma sobrestimação na ordem de 20 a 25% de NN_2 com massa predita se comparado a NN_2 com massa verdadeira nos casos de H_2 de 1 e 2 TeV dependendo do corte dos scores que denuncia uma possível quantidade insuficiente de backgrounds simulados.

Em geral, utilizar os scores de NN_1 e a massa predita trazem melhorias significativas que podem ser utilizados para melhor selecionar eventos de sinal. A garantia que os resultados obtidos são válidos vem de NN_2 com massa verdadeira que mostrou um comportamento próximo ou igual a NN_2 com massa predita. Resultados similares foram observados quando variamos a largura do bóson de Higgs para $\Gamma_H/m_H = 1\%$. Os conjuntos de treino/teste/validação foram aleatoriamente separados cinco vezes para avaliar a robustez dos resultados, e pequenas variações foram observadas na validação cruzada.

A importância da massa reconstruída com k NNNN é confirmado de uma análise da importância da variável usando o pacote SHAP (*SHapley Additive exPlanations*) [195]. O SHAP usa valores de Shapley da teoria de jogos para explicar a saída do modelo através das variáveis o que possibilita interpretar o modelo. Os modelos utilizados são redes neurais para uma classificação supervisionada, assim a interpretação depende do rótulo utilizado para cada classe definido na tabela 5.6.

	H_2	WW	ZZ	$t\bar{t}$
Rótulo	0	1	2	3

Tabela 5.6: Rótulo utilizado para cada classe em NN_1 e NN_2 .

Na figura 5.15 mostramos a hierarquia da importância da variável para o caso de H_2 de 1 TeV com 1000 eventos, ou seja, os valores de SHAP de cada variável para 1000 eventos do conjunto teste. A cor representa o valor da variável, um valor vermelho representa um valor alto e um valor azul representa um valor baixo. Por exemplo, um valor vermelho para a distribuição M_{Reco} , figura C.1f, indica um valor próximo de 50/60 enquanto um valor azul indica um valor próximo de 10/20. O valor de SHAP indica o impacto da variável na saída do modelo, um valor de SHAP positivo indica uma contribuição positiva, ou seja, a variável impacta o modelo com

a soma de um valor positivo, enquanto um valor negativo indica uma contribuição negativa, ou seja, uma subtração no lugar de soma. A saída do modelo para um evento é interpretado como a soma dos valores de SHAP para todas as variáveis que será um valor entre 0 e 3, onde os valores 0,1,2,3 representam as classes conforme a tabela 5.6. Por exemplo, a saída do modelo para um evento tem a soma dos valores de SHAP próximo de 0 indica que o evento é sinal, o análogo acontece para as outras classes.

As variáveis mais importantes aparecem nas linhas superiores e a sua importância decresce nas linhas inferiores, assim observamos que $M_{\ell\ell\nu\nu}$ predito é a variável mais importante, seguido dos scores obtidos de NN_1 de WW e $t\bar{t}$. Interessante observar que valores vermelhos de $M_{\ell\ell\nu\nu}$ predito tem a maior contribuição negativa, uma possível interpretação é que valores próximos do pico da massa invariante $M_{\ell\ell\nu\nu}$ ajudam os eventos a serem classificados como sinal enquanto para valores azuis, ou eventos longe do pico, o oposto acontece. Um comportamento similar é observado até a variável scores ZZ onde o valor da variável ser baixo ou alto não influencia mais no impacto do modelo. Os outros sinais mostraram um resultado similar com a massa $M_{\ell\ell\nu\nu}$ predita como a mais importante.

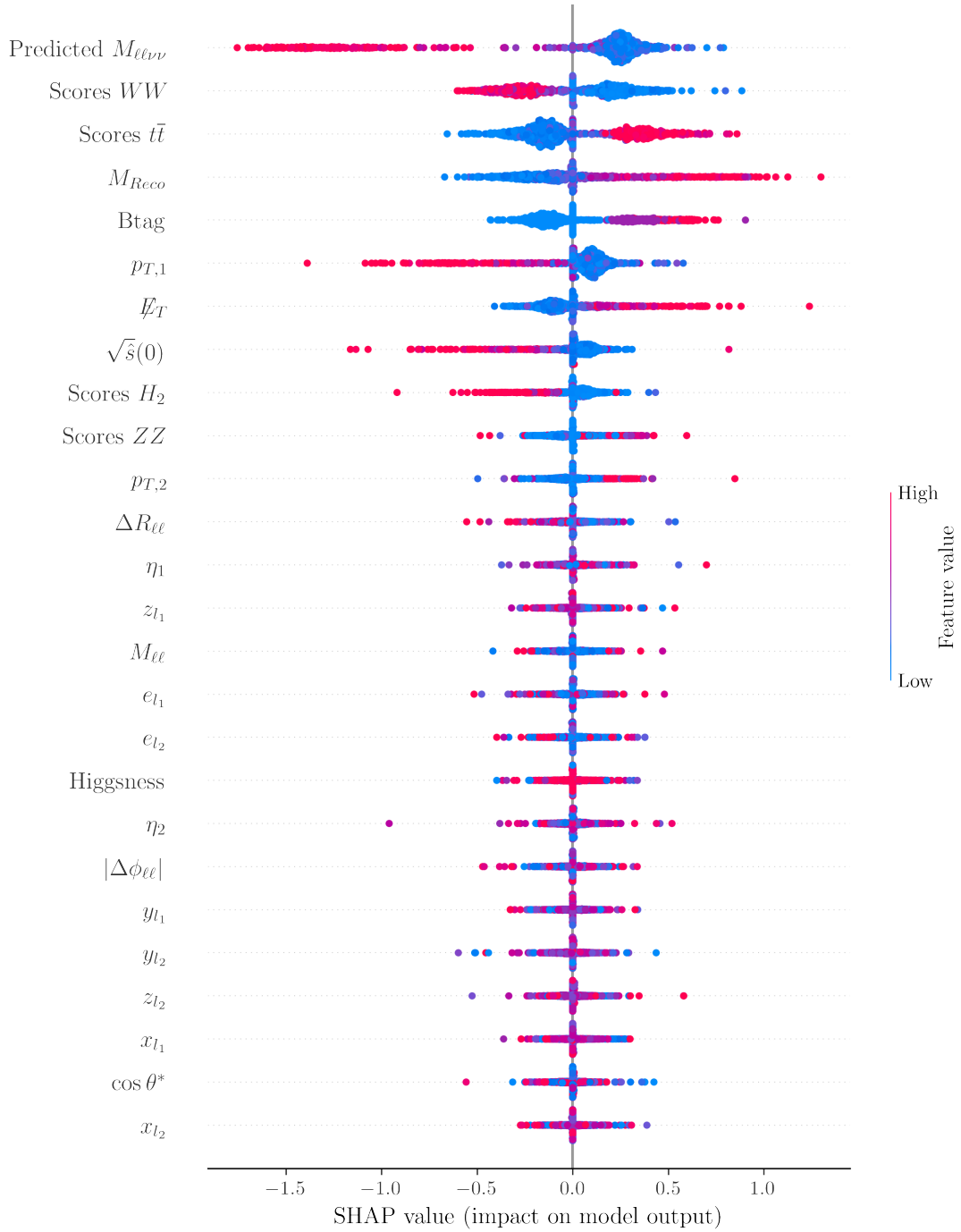


Figura 5.15: Valores de SHAP para H_2 de 1 TeV. O $M_{\ell\nu\nu}$ predito é a variável mais importante para a classificação quando incluída na representação dos dados.

Capítulo 6

Reconstrução de ressonâncias totalmente leptônicas - MSSM

No modelo padrão das interações, as massas das partículas são geradas quando há a quebra espontânea de simetria do potencial escalar $V = \mu^2|\phi|^2 + \lambda|\phi|^4$ para o campo de Higgs, onde assumimos que $\mu^2 < 0$ e $\lambda > 0$ pois o modelo padrão requer um valor esperado do vácuo diferente de zero para o campo ϕ no mínimo do potencial. A descoberta do bóson de Higgs [196, 197, 198, 199], com massa cerca de 125 GeV, implica $\lambda = 0.126$ e $\mu^2 = -(92.6\text{GeV})^2$. Mas há um problema, μ^2 recebe uma enorme quantidade de correções de efeitos virtuais de toda partícula que se acopla com o campo de Higgs, na forma de diagramas com loop. Assumir que o modelo padrão tenha uma grande região de validade desde escala eletrofraca até a escala de Planck $M_P \approx 10^{18}$ GeV requer um cancelamento das correções de uma parte em 10^{36} para obter o valor experimental. Esse é o problema da hierarquia do Higgs [59, 60].

Modelos supersimétricos resolvem esse problema introduzindo uma nova simetria, chamada *supersimetria*, que protege a massa do Higgs de receber correções quadráticas. A supersimetria é uma simetria do espaço-tempo que relaciona bósons e férmions [59, 60]. Escolhemos utilizar como modelo supersimétrico o MSSM, sendo a extensão do modelo padrão usando supersimetria com a menor quantidade de interações e partículas. O candidato à matéria escura do MSSM é a partícula mais leve da teoria que é estável, pela conservação da paridade-R – o neutralino mais leve [122, 11].

Para ilustrar o uso do algoritmo k NNNN na reconstrução de um estado final contendo matéria escura, escolhemos simular o decaimento do parceiro supersimétrico do elétron, o selétron (\tilde{e}), decaindo no neutralino mais leve mais um elétron. O processo de colisão simulado

foi

$$pp \rightarrow \tilde{e}_R^\pm \tilde{e}_R^\mp \rightarrow \ell^\pm \ell^\mp + \cancel{E}_T, \quad (6.1)$$

onde e_R é o selétron de mão direita com massas de 0.5 TeV e 1 TeV e os neutralinos de massa de 100 GeV.

A lagrangiana de interação relevante escrita na base dos autoestados de massa é

$$\begin{aligned} \mathcal{L}_{MSSM} = \mathcal{L}_{MP} &+ \frac{i}{2} \sum_i \bar{\Psi}(\tilde{\chi}_i^0) \not{\partial} \Psi(\tilde{\chi}_i^0) + \sum_{f,i} (\partial^\mu \tilde{f}_i^* \partial_\mu \tilde{f}_i) - \sum_f m_{\tilde{f}}^2 (\tilde{f}_L^* \tilde{f}_L + \tilde{f}_R^* \tilde{f}_R) \\ &- ig_Z \sum_f [\tilde{f}_L^* \overleftrightarrow{\partial}^\mu (T_{3f} - s_W^2 Q_f) \tilde{f}_L - \tilde{f}_R^* \overleftrightarrow{\partial}^\mu (s_W^2 Q_f) \tilde{f}_R] Z_\mu \\ &- ie \sum_f Q_f [\tilde{f}_L^* \overleftrightarrow{\partial}^\mu \tilde{f}_L - \tilde{f}_R^* \overleftrightarrow{\partial}^\mu \tilde{f}_R] A_\mu \\ &- \frac{g}{2} \sum_i \overline{\tilde{\chi}_i^0} (l_i^{NfL} \frac{1 - \gamma_5}{2} + r_i^{NfL} \frac{1 + \gamma_5}{2}) f \tilde{f}_L + \overline{\tilde{\chi}_i^0} (l_i^{NfR} \frac{1 - \gamma_5}{2} + r_i^{NfR} \frac{1 + \gamma_5}{2}) f \tilde{f}_R + h.c., \end{aligned} \quad (6.2)$$

onde $\tilde{\chi}_i^0$ são os neutralinos, $\tilde{f}, \tilde{f} = \tilde{e}_R$ são os Sférmions, $f, f = e$ são os férmions, $l_i^{NfL} = r_i^{NfL} = l_i^{NfR} = 0$, $r_i^{NeR} = 2\eta_i \frac{s_W}{c_W} (\mathcal{O}_N)_{i1}$, $m_{\tilde{f}}$ é a massa do Sférmion, $s_W = \sin \theta_W$, Q é a carga elétrica e T_3 é a terceira componente do isospin fraco.

A lagrangiana completa do MSSM pode ser encontrada na referência [200]. A implementação do MSSM no Madgraph5 segue a convenção SUSY Les Houches Accord [201, 202] que define todas as especificações do modelo, portanto excluindo a massa do neutralino mais leve e de \tilde{e}_R que foram modificados para a simulação todos os parâmetros seguem os valores definidos na convenção.

6.1 Aplicação do k NNNN

Os detalhes da aplicação do k NNNN está na seção anterior, focamos apenas na aplicação para os sinais de \tilde{e}_R com 0.5 e 1 TeV. Os hiperparâmetros ajustados para os sinais estão nas tabelas 6.1 e 6.2, onde observamos uma regularização mais intensa comparado ao sinal de H_2 , consequência dos valores da regularização L_2 e do dropout. O motivo da regularização ser mais intensa pode ser a diferente distribuição da massa $M_{\ell\nu\nu}$ se comparado ao sinal de H_2 . As redes neurais têm

as mesmas condições do processo de treino do sinal de H_2 .

Classe	k	Distância	Peso	Dimensão do espaço latente
\tilde{e}_R 0.5 TeV	1	Chebyshev	Uniform	2
\tilde{e}_R 1 TeV	1	Minkowski	Uniform	3

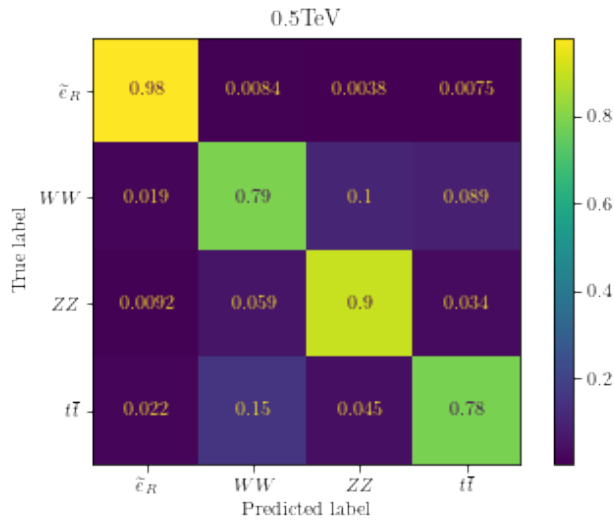
Tabela 6.1: Tabela mostrando os melhores hiperparâmetros encontrados pela busca em rede para a predição da distribuição $M_{\ell\nu\nu}$ para o sinal \tilde{e}_R de 0.5 e 1 TeV.

Hiperparâmetro e arquitetura	0.5 TeV	1 TeV
Regularização L2	1.01×10^{-5}	2.9×10^{-6}
Learning rate inicial	6.5×10^{-4}	5.0×10^{-3}
Tamanho dos lotes	256	64
Taxa de dropout	0.47	0.38
Inicialização dos pesos	uniform	uniform
Função de ativação	ReLU	ReLU
Número de camadas e nós	(256,128,64,32)	(160,80)
Total de parâmetros	48864	16400

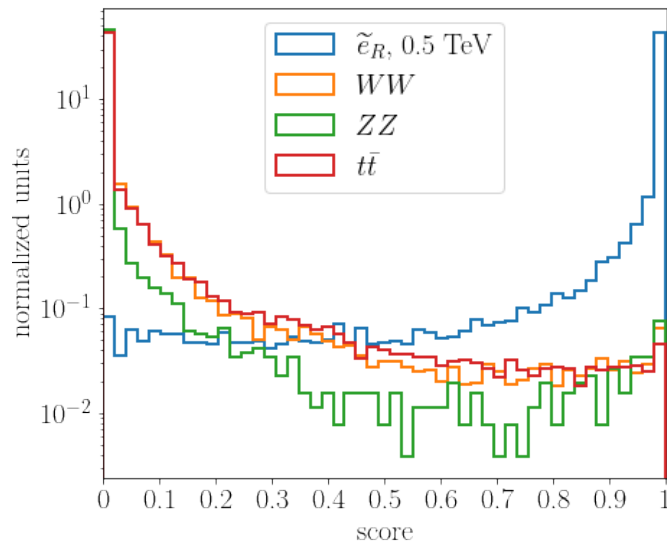
Tabela 6.2: Hiperparâmetros e arquitetura da rede neural para classificação dos sinais de 0.5, 1 TeV dos \tilde{e}_R .

A figura 6.1 mostra a matriz de confusão e os scores de ambos os sinais de \tilde{e}_R do classificador NN_1 . O sinal de \tilde{e}_R de 0.5 TeV tem 15% de eventos de $t\bar{t}$ confundidos como WW , 5.9% de eventos de ZZ confundidos como WW e 10% e 8.9% de eventos de WW confundidos como ZZ e $t\bar{t}$, respectivamente. Enquanto o sinal de \tilde{e}_R de 1 TeV tem 13% de eventos de WW classificados como $t\bar{t}$, 15% de eventos de $t\bar{t}$ confundidos como WW e o ZZ tem 39% de eventos classificados como WW . Em ambos os sinais de \tilde{e}_R o comportamento em relação a WW e $t\bar{t}$ é similar aos sinais de H_2 , o ZZ no sinal de \tilde{e}_R de 0.5 e 1 TeV tem um comportamento similar ao ZZ no sinal de H_2 de 1 e 2 TeV, respectivamente. O comportamento da matriz confusão é refletido no gráfico dos scores, onde o sinal de \tilde{e}_R de 1 TeV mostra a curva dos três backgrounds sobrepondo mutuamente pela abundância de eventos ZZ classificados como WW , enquanto o sinal \tilde{e}_R de 0.5 TeV a curva ZZ tende a ficar embaixo das curvas de WW e $t\bar{t}$ que se sobrepõem. O motivo deve ser similar do sinal de H_2 , onde distribuições com todos os backgrounds similares como $|\Delta\phi_{\ell\ell}|$ e Higgsness afetam os sinais com alta massa em uma maior proporção.

A figura 6.2 mostra que, em relação ao sinal supersimétrico, \tilde{e}_R , há também uma boa separação de sinal e backgrounds, com cerca de 20% dos eventos de ZZ e $t\bar{t}$ sendo considerados como WW após a classificação com NN_2 . Na verdade, a separação dos backgrounds e



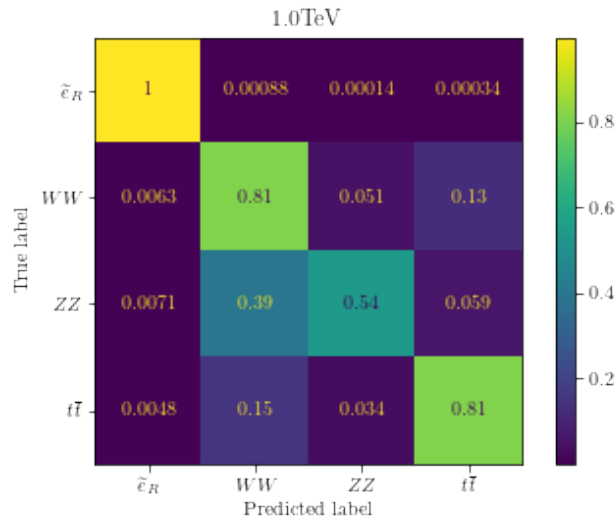
(a) Matriz confusão da classificação NN_1 para o sinal de \tilde{e}_R de 0.5 TeV.



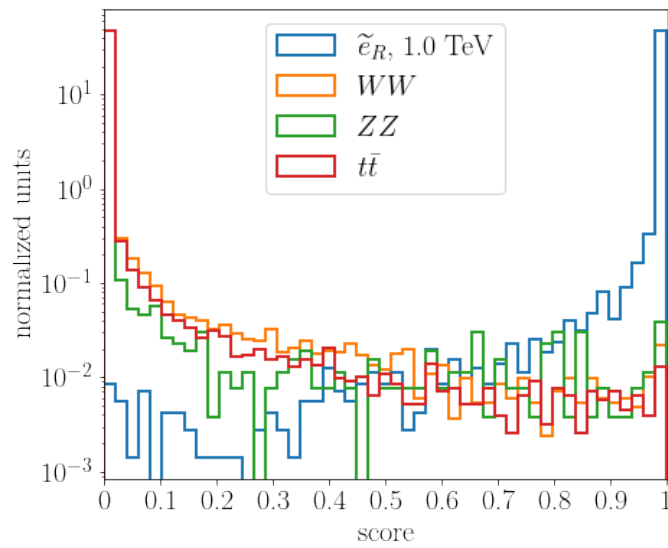
(b) Scores da classificação NN_1 para o sinal de \tilde{e}_R de 0.5 TeV, a linha azul representa o sinal, a linha laranja representa o background WW , a linha verde representa o background ZZ e a linha vermelha representa o background $t\bar{t}$.

o sinal é ainda mais eficiente se comparada ao sinal de H_2 , conforme mostra o comportamento dos gráficos de scores. A quantidade de eventos de sinal confundidos como background chega a ser irrelevante e até nulo no caso de maior massa e a quantidade de backgrounds considerados como sinais tem o mesmo comportamento, mas não chega a ser nulo.

A tabela 6.3 mostra os resultados das métricas utilizadas que comprovam o comportamento observado. A assimetria e AUC mostraram um resultado beirando o ideal mostrando a capacidade do $kNNNN$ em separar os backgrounds do sinal supersimétrico. A $accuracy$, LR_+ e LR_-

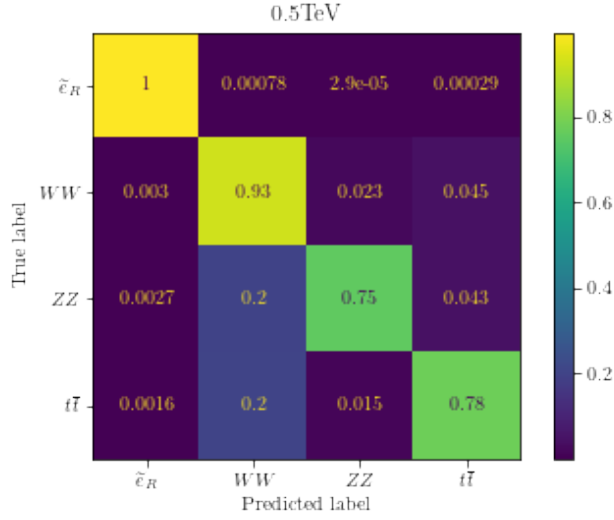


(c) Matriz confusão da classificação NN_1 para o sinal de \tilde{e}_R de 1 TeV.

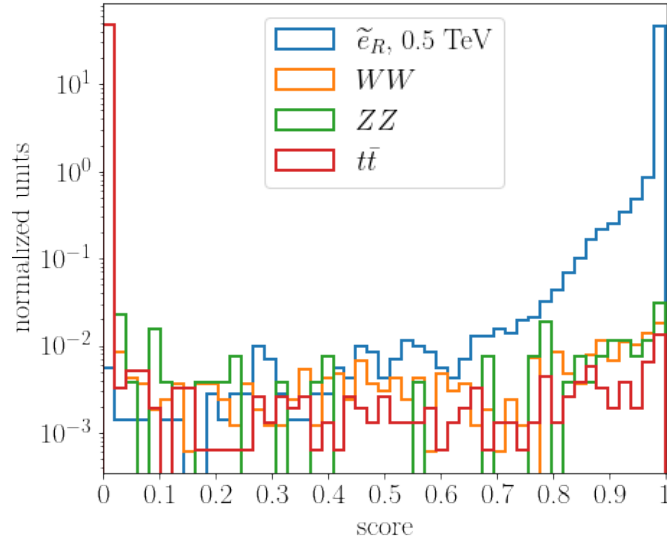


(d) Scores da classificação NN_1 para o sinal de \tilde{e}_R de 1 TeV, a linha azul representa o sinal, a linha laranja representa o background WW , a linha verde representa o background ZZ e a linha vermelha representa o background $t\bar{t}$.

Figura 6.1: Matriz confusão da classificação e saída dos scores da rede neural NN_1 nos painéis da esquerda e direita, respectivamente, para os sinais de \tilde{e}_R de 0.5 TeV(primeira linha) e 1 TeV(segunda linha).



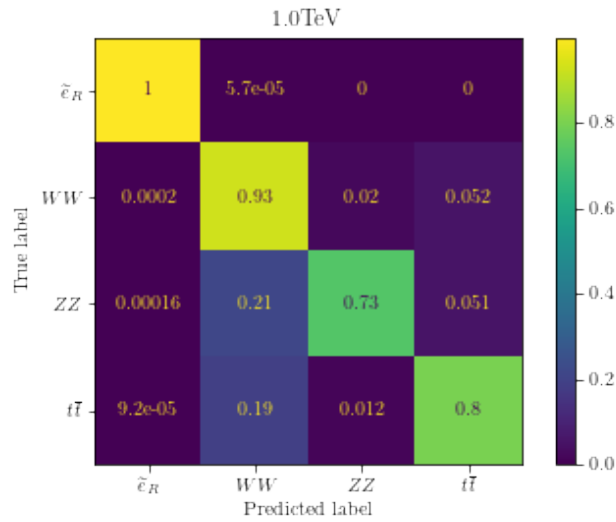
(a) Matriz confusão da classificação NN_2 para o sinal de \tilde{e}_R de 0.5 TeV.



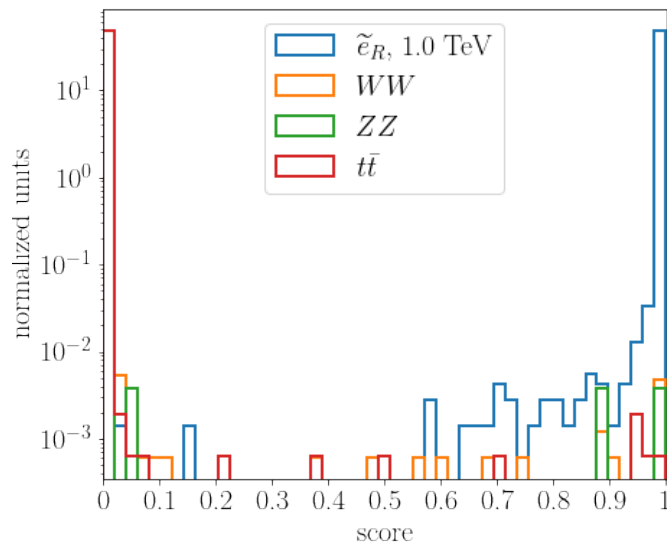
(b) Scores da classificação NN_2 para o sinal de \tilde{e}_R de 0.5 TeV, a linha azul representa o sinal, a linha laranja representa o background WW , a linha verde representa o background ZZ e a linha vermelha representa o background $t\bar{t}$.

mostraram uma melhoria de NN_2 em relação a NN_1 similar ao caso de sinal de H_2 . O $kNNNN$ novamente apresentou um maior acréscimo de performance de NN_1 para NN_2 no caso do sinal de menor massa. Apesar de a assimetria ter um comportamento beirando o ideal, a *accuracy* mostra que a classificação entre os backgrounds ainda pode ser melhorada.

A diferença da matriz de confusão vista na figura 6.3 mostrou que a diagonal principal tem uma grande melhoria de WW para ambos os casos, enquanto ZZ piorou um pouco para o caso de \tilde{e}_R de 0.5 TeV, mas melhorou consideravelmente no caso de \tilde{e}_R de 1 TeV. O sinal e $t\bar{t}$ tiveram



(c) Matriz confusão da classificação NN_2 para o sinal de \tilde{e}_R de 1 TeV.



(d) Scores da classificação NN_2 para o sinal de \tilde{e}_R de 1 TeV, a linha azul representa o sinal, a linha laranja representa o background WW , a linha verde representa o background ZZ e a linha vermelha representa o background $t\bar{t}$.

Figura 6.2: Matriz confusão da classificação e saída dos scores da rede neural NN_2 nos painéis da esquerda e direita, respectivamente, para os sinais de \tilde{e}_R de 0.5 TeV(primeira linha) e 1 TeV(segunda linha).

Métrica	0.5 TeV		1 TeV	
	NN ₁	NN ₂	NN ₁	NN ₂
<i>accuracy</i>	82.8%	87.4%	82.6%	88.0%
assimetria, MSSM	0.947	0.996	0.995	0.999
assimetria, WW	-0.973	-0.994	-0.991	-0.999
assimetria, ZZ	-0.983	-0.994	-0.987	-0.999
assimetria, $t\bar{t}$	-0.970	-0.993	-0.934	-0.999
LR_+	4.688	6.607	4.747	6.924
LR_-	0.111	0.013	0.033	0.001
AUC(MSSM \times WW)	0.9982	0.9999	0.9999	1.0000
AUC(MSSM \times ZZ)	0.9987	0.9998	0.9999	0.9996
AUC(MSSM \times $t\bar{t}$)	0.9981	0.9999	1.0000	1.0000

Tabela 6.3: Comparação do desempenho das métricas obtidas das classificações NN₁ e NN₂ para identificar o \tilde{e}_R das duas massas consideradas.

uma pequena melhoria e piora, respectivamente. Os termos fora da diagonal mostram o aumento de ZZ considerados como $t\bar{t}$ no caso de menor massa e um aumento de pequena quantidade de $t\bar{t}$ considerado como WW em ambos os casos. Os outros termos têm um comportamento ideal apesar de ser em pequena quantidade.

As curvas ROC do sinal de \tilde{e}_R de 0.5 TeV da figura 6.4 mostram um comportamento perto do ideal, para uma aceitação fixa de \tilde{e}_R , a rejeição dos backgrounds em NN₂ é maior que NN₁. O mesmo comportamento foi observado para o caso de sinal de 1 TeV confirmado pelo AUC da tabela 6.3, mas em menor proporção, pois ambas as classificações estão próximas do ideal.

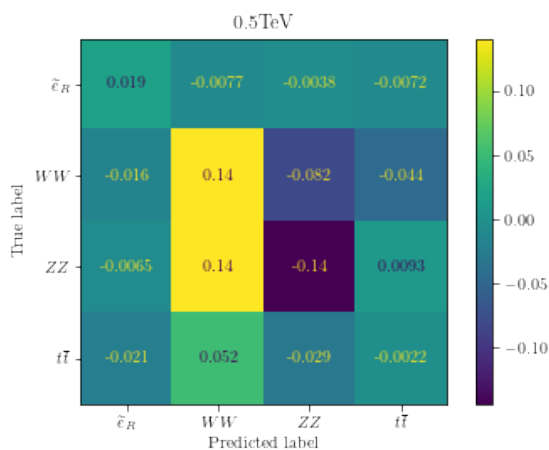
As distribuições da massa $M_{\ell\nu\nu}$ regredida para \tilde{e}_R de 0.5 TeV estão na figura 6.5. As distribuições de NN₁ mostram uma contaminação de backgrounds como os sinais de H_2 e as distribuições de NN₂ mostram distribuições livres de contaminação, um comportamento similar ao caso de H_2 .

A figura 6.6 mostra a significância estatística de descoberta para os sinais de \tilde{e}_R onde destacamos a grande diferença entre a significância estatística de NN₁ e NN₂ sem massa predita com NN₂ com a massa predita. O ganho da significância estatística entre NN₁ e NN₂ sem a massa predita é em torno de 2 enquanto o ganho de NN₂ com massa predita chega a ser maior que 100! nos dois sinais de \tilde{e}_R quando não existe o corte nos scores, mostrando como a significância estatística de descoberta de NN₂ com a massa predita é muito melhor e a diferença cai drasticamente ao aumentar o corte do score mostrando que a melhoria na performance de NN₂ com massa predita é menor que os outros casos. O motivo do ganho enorme de NN₂ com massa predita pode ser a classificação quase ideal entre os backgrounds e sinal conforme as métricas

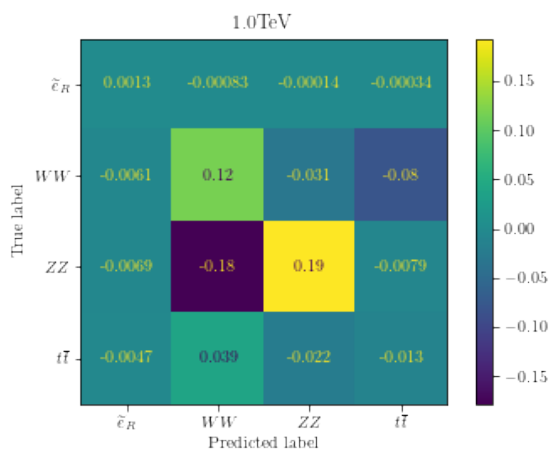
da tabela 6.3 e faz com que o aumento no corte do score seja menos relevante se comparado a NN_1 e NN_2 sem massa predita.

O sinal de \tilde{e}_R de 1 TeV mostra uma grande sobrestimação de NN_2 com massa verdadeira comparado a NN_2 com massa predita, enquanto o sinal de \tilde{e}_R de 0.5 TeV de NN_2 com massa verdadeira é levemente sobrestimado na região de scores entre 0.7 a 0.9 e equivalente no resto se comparado a NN_2 com massa predita mostrando um comportamento contrário aos sinais de H_2 de 1 e 2 TeV que possuem uma sobrestimação de NN_2 com massa predita. Assim obtemos um resultado ainda melhor que os sinais de H_2 mostrando a capacidade do $kNNNN$ em selecionar eventos de sinal.

A importância de $M_{\ell\ell\nu\nu}$ na representação dos dados pode ser observada na figura 6.7, onde

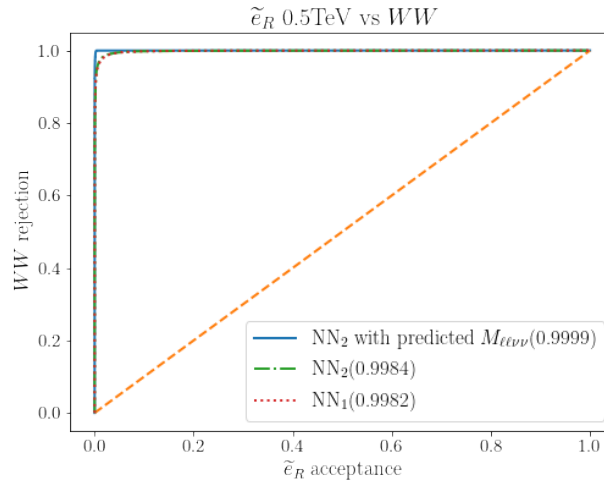


(a) Diferença entre a matriz confusão de NN_1 com NN_2 para o sinal de 0.5 TeV e largura de 10%.

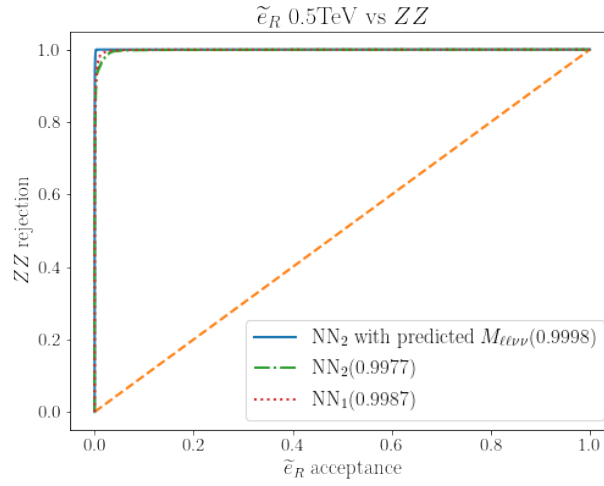


(b) Diferença entre a matriz confusão de NN_1 com NN_2 para o sinal de 1 TeV e largura de 10%.

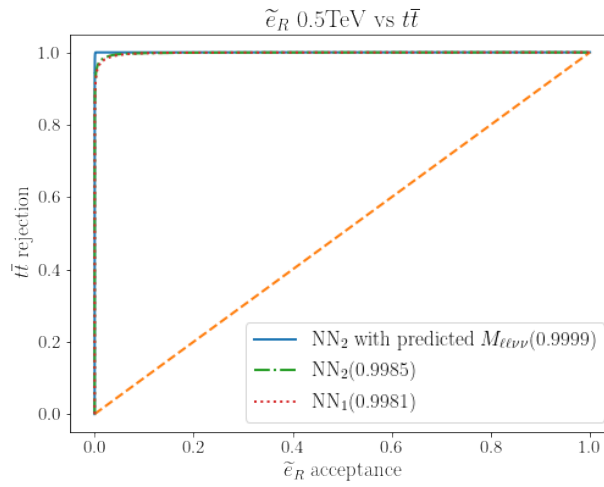
Figura 6.3: Diferença entre a matriz confusão de NN_1 com o NN_2 , ou seja, $NN_2 - NN_1$.



(a)

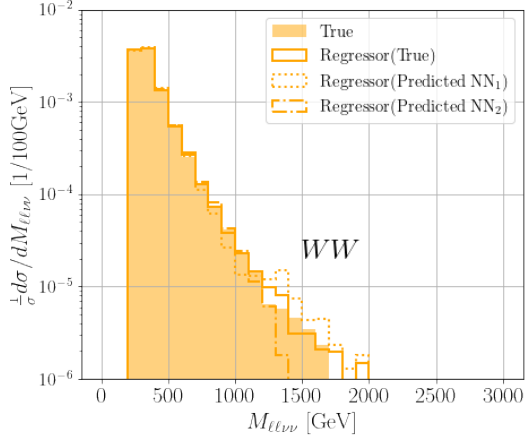


(b)

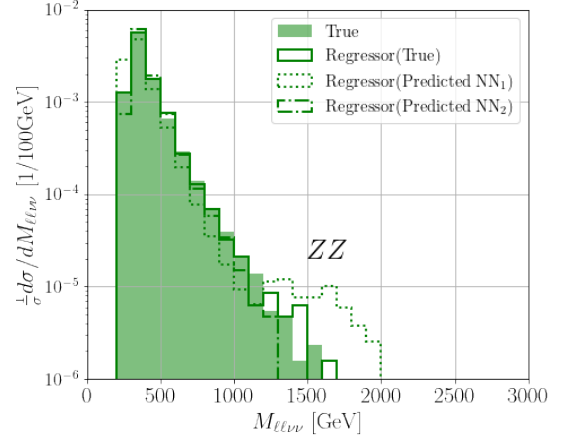


(c)

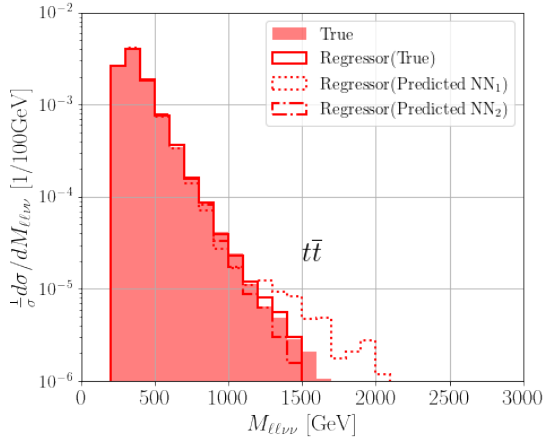
Figura 6.4: As curvas ROC do \tilde{e}_R de 0.5 TeV contra as classes de background WW , ZZ , e $t\bar{t}$. Nas legendas, em parêntesis, mostramos o AUC correspondente da primeira classificação, NN_1 e da segunda classificação, NN_2 , com e sem incluir a massa predita.



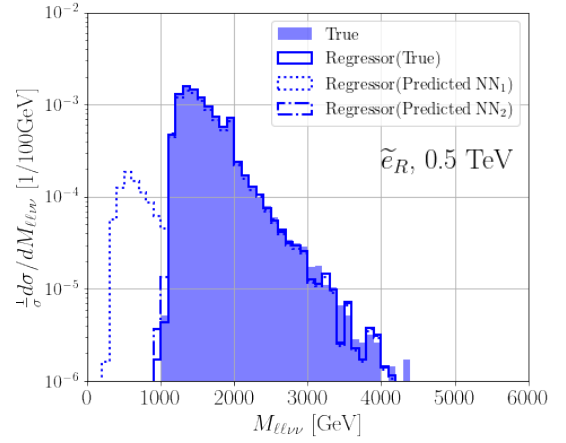
(a) Distribuição $M_{\ell\nu\nu}$ do background WW .



(b) Distribuição $M_{\ell\nu\nu}$ do background ZZ .



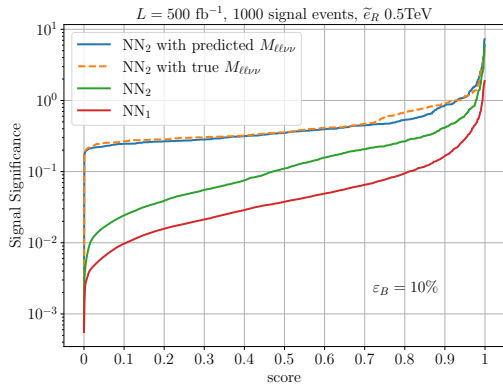
(c) Distribuição $M_{\ell\nu\nu}$ do background $t\bar{t}$.



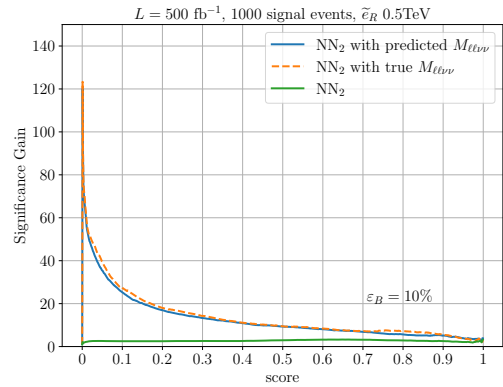
(d) Distribuição $M_{\ell\nu\nu}$ do sinal \tilde{e}_R de 0.5 TeV.

Figura 6.5: Gráficos mostrando a distribuição $M_{\ell\nu\nu}$ verdadeira(fundo), com a regredida a partir dos dados reais(linha sólida), a regredida com os dados identificados com NN_1 (linha pontilhada) para o sinal \tilde{e}_R de 0.5 TeV e a regredida com os dados identificados com NN_2 (linha semi-pontilhada). O azul representa o sinal, laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

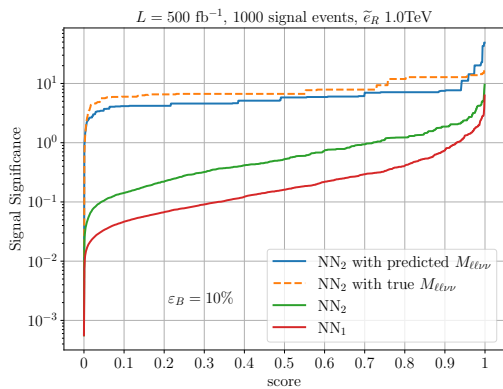
calculamos os valores de SHAP para 1000 eventos de NN_2 do sinal \tilde{e}_R de 0.5 TeV. As variáveis mais importantes são a massa $M_{\ell\nu\nu}$ predita e os scores de $t\bar{t}$ e WW , respectivamente, as mesmas variáveis do sinal de H_2 com 1 TeV. Resultados similares foram obtidos para os outros sinais. Independente do sinal e do formato da distribuição de $M_{\ell\nu\nu}$, o $M_{\ell\nu\nu}$ predito continua sendo a variável mais importante na representação dos dados.



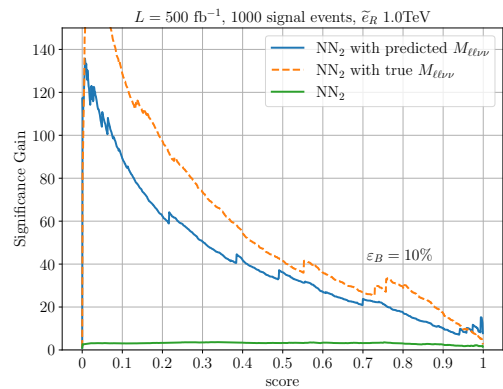
(a) Significância estatística de descoberta para o sinal \tilde{e}_R de massa 0.5 TeV.



(b) Ganho da significância estatística de descoberta para o sinal \tilde{e}_R de massa 0.5 TeV.



(c) Significância estatística de descoberta para o sinal \tilde{e}_R de massa 1 TeV.



(d) Ganho da significância estatística de descoberta para o sinal \tilde{e}_R de massa 1 TeV.

Figura 6.6: A significância estatística do k NNNN algoritmo como uma função do corte na saída de scores (painéis da esquerda) e do ganho da significância, eq.5.13, comparado a classificação NN_1 (painéis da direita) para os sinais de \tilde{e}_R com 0.5 TeV (primeira linha) e 1 TeV(segunda linha). A incerteza sistemática é fixo em 10%.

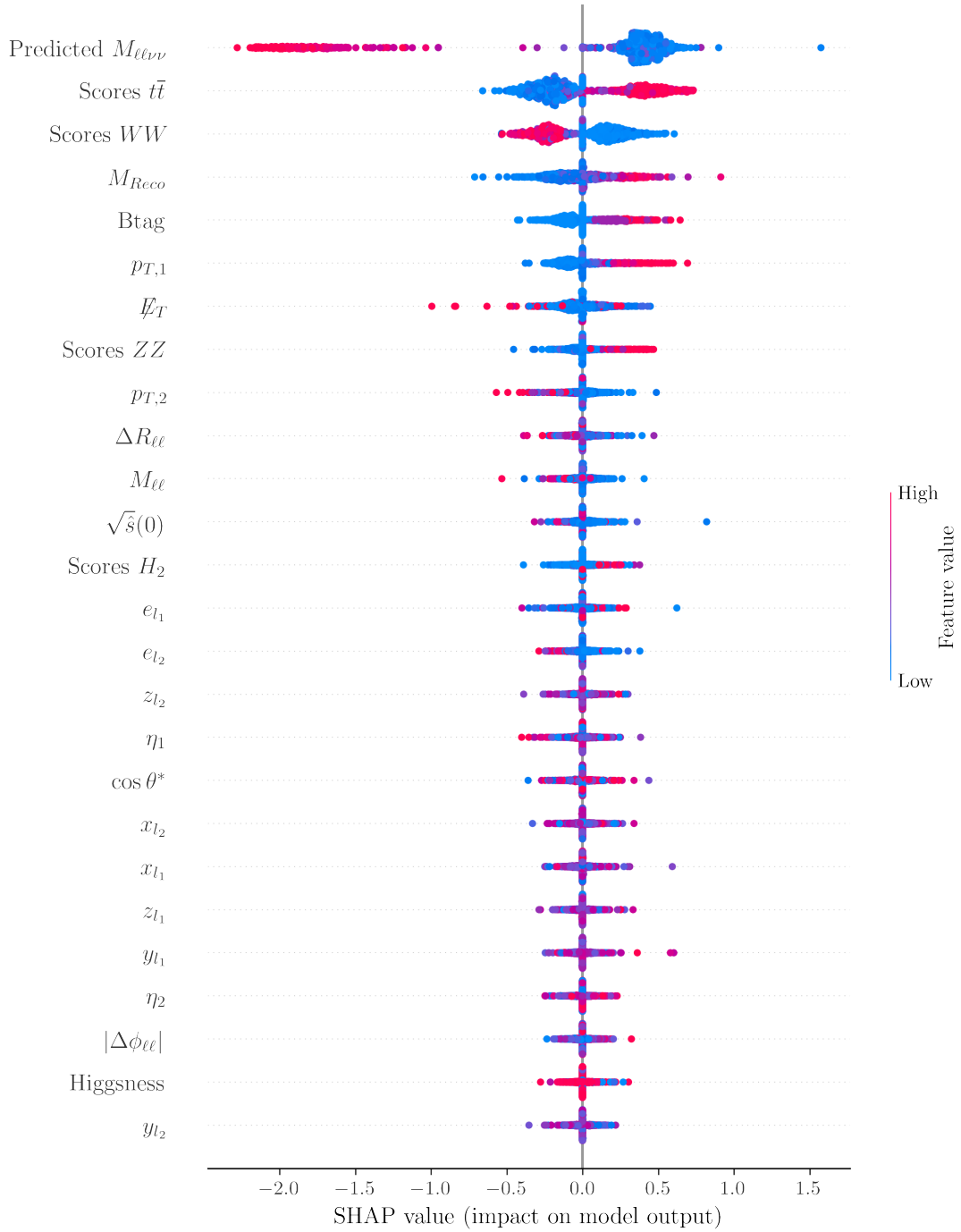


Figura 6.7: Gráfico dos valores de SHAP para \tilde{e}_R de 0.5 TeV. O $M_{\ell\nu\nu}$ predito é a variável mais importante para a classificação quando incluída na representação dos dados.

Capítulo 7

Conclusão

A busca de nova física se intensificará seguindo o cronograma do LHC e novos métodos de identificar partículas que escondem informação através de decaimentos invisíveis são certamente bem-vindos. Nesta dissertação, projetamos um algoritmo capaz de reconstruir a massa de dois sinais; (1) um novo bóson de Higgs pesado com decaimento $W^+W^- \rightarrow \ell^+\ell'^-\nu\bar{\nu}'$ e (2) um slepton do MSSM no decaimento $\tilde{e}_R^+\tilde{e}_R^- \rightarrow \ell^+\ell'^- + \tilde{\chi}^0$; além dos principais backgrounds do modelo padrão para estes processos. O algoritmo assume conhecimento prévio da classe dos eventos e da massa dos sinais, portanto sendo útil em estudos pós-descoberta, por exemplo, uma análise que requer uma seleção de bósons de Higgs na camada de massa.

Mais importante, incluir a massa predita $M_{\ell\nu\nu}$ como atributo para a representação dos dados em redes neurais de classificação melhora a *accuracy*, a taxa de verdadeiro e falso negativo/positivo e a probabilidade de classificações da classe verdadeira quando comparados a uma rede neural cujos dados não possuam informação sobre a massa invariante do estado final contendo o estado escuro, neutrinos ou neutralinos. Computamos as curvas ROC e confirmamos que a segunda classificação aumenta o AUC quando comparado à primeira, especialmente quando a massa predita é incluída na representação dos dados, melhorando a rejeição dos backgrounds a uma taxa fixa de aceitação de sinal. A análise da importância da variável confirma o papel desempenhado pelo $M_{\ell\nu\nu}$ regredido para o meta-classificador NN₂. O ganho da significância estatística é o teste principal do algoritmo proposto. Ambos os sinais mostraram um ganho muito pronunciado na significância estatística do sinal, tipicamente maior do que 10, dependendo do corte selecionado no score da classe do sinal. Verificamos que a massa predita é confiável e robusta como uma nova variável para a classificação comparando os resultados contra um classificador treinado com a massa $M_{\ell\nu\nu}$ verdadeira. Também em relação à distribuição

verdadeira, a significância estatística calculada com as massas regredidas concordam razoavelmente bem.

O algoritmo k NNNN pode ser aplicado a outras variáveis observáveis. Por exemplo, o ângulo de espalhamento dos bósons W pode ser obtido no canal totalmente leptônico ao lado do ângulo dos léptons carregados. A massa das partículas em diferentes topologias também pode ser obtido. Por exemplo, supomos que a distribuição das massas das partículas a partir de cadeias de decaimento de diversos tamanhos podem ser recuperados após a determinação com outros métodos.

O próximo passo nesta categoria de investigação é relaxar o conhecimento prévio dos parâmetros de massa e enfraquecer o nível de supervisionamento quando treinamos os classificadores e regressores. A detecção de valores discrepantes e outras técnicas de aprendizado não supervisionado podem ser usados para dispensar o conhecimento prévio da classe dos sinais, mas utilizar o k NN para a regressão requer o conhecimento prévio do parâmetro de massa. Um algoritmo de regressão semi-supervisionado que assume apenas o conhecimento dos backgrounds é difícil, pois envolve generalização entre classes com perda de informação essencial. Investigações mais profundas podem usar técnicas avançadas de redes neurais como VAEs para a regressão [203] treinadas em uma única classe de backgrounds, mas ainda assumindo o conhecimento prévio das massas do sinal. Utilizar técnicas de redução da dimensionalidade como classificações não supervisionadas sem nenhum conhecimento prévio do sinal também são possibilidades para separar o sinal do background.

No estado atual, a utilização do algoritmo para um estudo pré-descoberta não é possível, mas no futuro existe a possibilidade que seja possível o uso em um estudo pré-descoberta, pois existe um grande potencial de melhorias tornando os métodos propostos ainda mais poderosos.

Apêndice A

Atualização dos pesos dado pelo backpropagation

Consideramos uma rede neural com camada de entrada x , n camadas ocultas e camada de saída, onde cada nó tem k_n entradas, terá a regra de atualização dos pesos para uma função de ativação ϕ dado pelo backpropagation que será proporcional a derivada do erro ($E = L(y, t)$) em função do peso, onde L é a função perda, y é a saída da camada de saída e t é o alvo de interesse. A saída de cada nó é representado por y_j e pela soma $net_j = \sum_{i=0}^m \omega_{ij}y_i$, onde $y_i = x_i$ se j está na primeira camada oculta. Então a derivada do erro em relação ao peso será

$$\frac{\partial E}{\partial \omega_{kj}} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\omega_{kj}} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial net_j} \frac{\partial net_j}{\partial \omega_{kj}}, \quad (\text{A.1})$$

onde usamos a regra da cadeia. Começando pelo termo da direita $\frac{\partial net_j}{\partial \omega_{kj}}$, temos que

$$\frac{\partial net_j}{\partial \omega_{kj}} = \frac{\partial}{\partial \omega_{kj}} \sum_{i=0}^m \omega_{ij}y_i = \frac{\partial}{\partial \omega_{kj}} \omega_{kj}y_i = y_i, \quad (\text{A.2})$$

onde $y_k = x_k$ se o nó está na primeira camada oculta. O termo $\frac{\partial y_j}{\partial net_j}$ será

$$\frac{\partial y_j}{\partial net_j} = \frac{\partial \phi(net_j)}{\partial net_j}, \quad (\text{A.3})$$

ou seja, é simplesmente a derivada da função de ativação $\phi(net_j)$ em relação à soma net_j .

Para a camada de saída o termo $\frac{\partial E}{\partial y_j}$ é

$$\frac{\partial E}{\partial y_j} = \frac{\partial E}{\partial y}, \quad (\text{A.4})$$

mas para camadas ocultas encontrar a derivada de E em relação a y_j não é obvio, pois o erro E depende do erro de todos os nós que podem ser conectados $N = u, v, \dots, w$ ao nó de interesse da camada oculta j . Assim $\frac{\partial E}{\partial y_j}$ será

$$\frac{\partial E(y_j)}{\partial y_j} = \frac{\partial E(net_u, net_v, \dots, net_w)}{\partial y_j}, \quad (\text{A.5})$$

efetuando a derivada em relação a y_j , temos

$$\frac{\partial E}{\partial y_j} = \sum_{n \in N} \left(\frac{\partial E}{\partial net_n} \frac{\partial net_n}{\partial y_j} \right) = \sum_{n \in N} \left(\frac{\partial E}{\partial y_n} \frac{\partial y_n}{\partial net_n} \frac{\partial net_n}{\partial y_j} \right) = \sum_{n \in N} \left(\frac{\partial E}{\partial y_n} \frac{\partial y_n}{\partial net_n} \omega_{jn} \right). \quad (\text{A.6})$$

Substituindo as equações A.2,A.3,A.4 e A.6 na equação A.1, temos

$$\frac{\partial E}{\partial \omega_{kj}} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial net_j} \frac{\partial net_j}{\partial \omega_{kj}} = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial net_j} y_k = y_k \delta_j, \quad (\text{A.7})$$

onde

$$\delta_j = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial net_j} = \begin{cases} \frac{\partial L(y_j, t)}{\partial y_j} \frac{d\phi(net_j)}{dnet_j} & \text{se } j \text{ é um nó da camada de saída,} \\ (\sum_{n \in N} \omega_{jn} \delta_n) \frac{d\phi(net_j)}{dnet_j} & \text{se } j \text{ é um nó de uma camada oculta.} \end{cases} \quad (\text{A.8})$$

A regra de atualização dos pesos dado pelo gradiente descendente será proporcional a taxa de aprendizado α , assim a atualização dos pesos será

$$\Delta \omega_{kj} = -\alpha \frac{\partial E}{\partial \omega_{kj}} = -\eta y_k \delta_j, \quad (\text{A.9})$$

onde y_k e δ_j definidos acima e o sinal negativo garante que o erro E sempre diminuir a cada passo realizado pelo gradiente descendente. Por exemplo, para a função de ativação sigmóide δ_j será

$$\delta_j = \frac{\partial E}{\partial y_j} \frac{\partial y_j}{\partial net_j} = \begin{cases} (o_j - t_j) o_j (1 - o_j) & \text{se } j \text{ é um nó da camada de saída,} \\ (\sum_{n \in N} \omega_{jn} \delta_n) o_j (1 - o_j) & \text{se } j \text{ é um nó de uma camada oculta.} \end{cases}, \quad (\text{A.10})$$

onde t_j é o rótulo do evento.

Apêndice B

Análise do sinal de bósons de Higgs pesado com largura diferente

Os métodos propostos usaram sinais com mesma largura. Neste apêndice verificamos que o método proposto funciona também para sinais de larguras diferentes. Usamos o mesmo sinal de H_2 para massa de 1, 1.5 e 2 TeV com largura de 1% da massa. Os hiperparâmetros da regressão e classificação estão nas tabelas B.1 e B.2, respectivamente. Similar ao caso de maior largura apenas o caso de 1 TeV usa o dropout.

Classe	k	Distância	Peso	Dimensão do espaço latente
H_2 1 TeV(1%)	1	Minkowski	Uniform	4
H_2 1.5 TeV(1%)	1	Canberra	Uniform	3
H_2 2 TeV(1%)	1	Canberra	Uniform	4

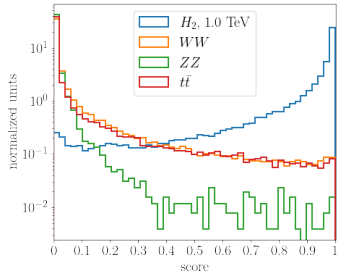
Tabela B.1: Tabela mostrando os melhores hiperparâmetros encontrados pela busca em rede para a predição da distribuição $M_{\ell\nu\nu}$.

As figuras B.1 e B.2 mostram os scores e as matrizes confusões, respectivamente. A principal diferença para os casos com maior largura está no desempenho de NN_1 , onde os casos com massa 1.5 e 2 TeV possuem um comportamento similar com um desempenho levemente superior aos casos com largura de 10%, a *accuracy* mostra 81.6% e 83.8% para largura de 1% contra 81.3% e 80.6% nos casos com largura de 10%. Em NN_2 tanto o comportamento como o desempenho são similares. O caso de 1 TeV mostrou um comportamento diferente, onde o caso com largura de 1% apresentou um comportamento similar a NN_2 , onde o ZZ é bastante confuso com o WW .

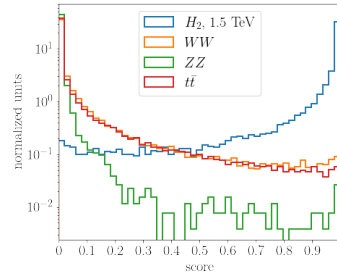
A tabela das métricas B.3 confirma esse comportamento. Os valores de *accuracy*, assimetria

Hiperparâmetro e arquitetura	1 TeV	1.5 TeV	2 TeV
Regularização L2	2.1×10^{-5}	8.7×10^{-6}	1.9×10^{-5}
Learning rate inicial	1.5×10^{-4}	3.8×10^{-4}	2.8×10^{-3}
Tamanho dos lotes	1024	256	1024
Taxa de dropout	0.10	–	–
Inicialização dos pesos	uniform	uniform	uniform
Função de ativação	tanh	Softplus	Softplus
Número de camadas e nós	(96,48,24,12)	(512,256,128)	(384,192,96)
Total de parâmetros	8244	175488	100896

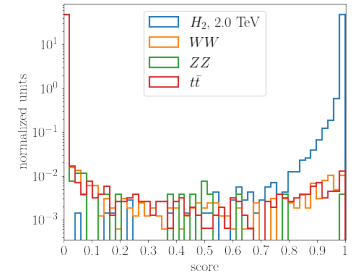
Tabela B.2: Hiperparâmetros e arquitetura da rede neural para classificação dos sinais H_2 de 1, 1.5, 2 TeV com largura de 1%. Os casos de 1.5 e 2.0 TeV não foi necessário camadas de dropout.



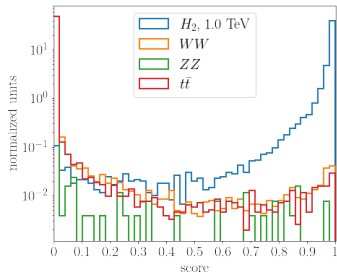
(a) Scores da classificação NN_1 para o sinal de 1 TeV e largura de 1%.



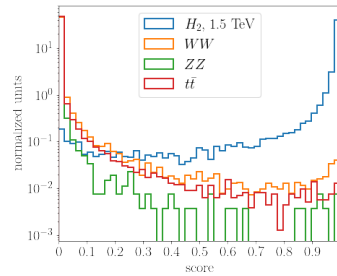
(b) Scores da classificação NN_1 para o sinal de 1.5 TeV e largura de 1%.



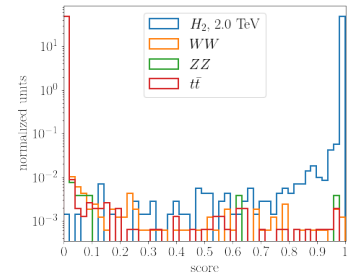
(c) Scores da classificação NN_1 para o sinal de 2 TeV e largura de 1%.



(d) Scores da classificação NN_2 para o sinal de 1 TeV e largura de 1%.

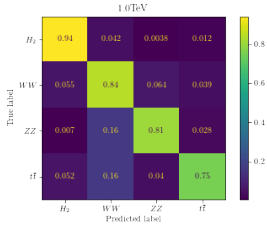


(e) Scores da classificação NN_2 para o sinal de 1.5 TeV e largura de 1%.

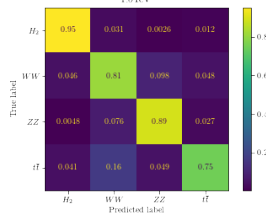


(f) Scores da classificação NN_2 para o sinal de 2 TeV e largura de 1%.

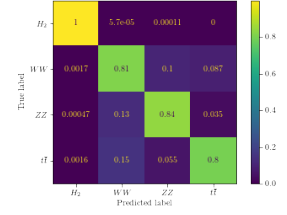
Figura B.1: Probabilidade de cada classe ser sinal utilizando a rede neural como classificador. Valores de score próximos de 1 representam eventos com probabilidade de serem sinais e valores de score próximos de 0 representam eventos com probabilidade de serem background. Em todos os casos, a linha azul representa o sinal, a linha laranja representa o background WW , a linha verde representa o background ZZ e a linha vermelha representa o background $t\bar{t}$.



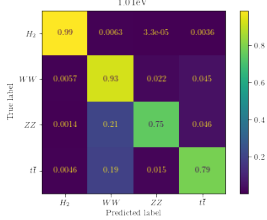
(a) Matriz confusão da classificação NN_2 para o sinal de 1 TeV e largura de 1%.



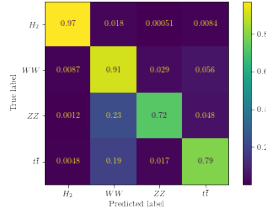
(b) Matriz confusão da classificação NN_2 para o sinal de 1.5 TeV e largura de 1%.



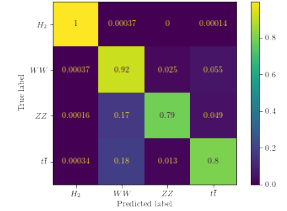
(c) Matriz confusão da classificação NN_2 para o sinal de 2 TeV e largura de 1%.



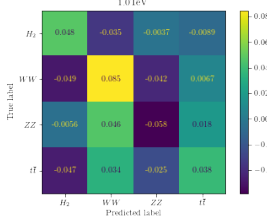
(d) Matriz confusão da classificação NN_2 para o sinal de 1 TeV e largura de 1%.



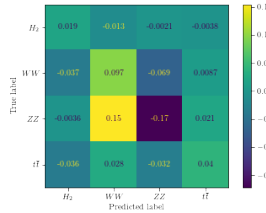
(e) Matriz confusão da classificação NN_2 para o sinal de 1.5 TeV e largura de 1%.



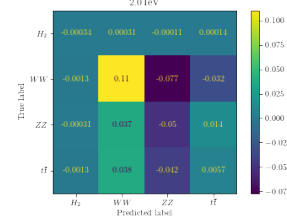
(f) Matriz confusão da classificação NN_2 para o sinal de 2 TeV e largura de 1%.



(g) Diferença entre a matriz confusão de NN_1 com NN_2 para o sinal de 1 TeV e largura de 1%.



(h) Diferença entre a matriz confusão de NN_1 com NN_2 para o sinal de 1.5 TeV e largura de 1%.

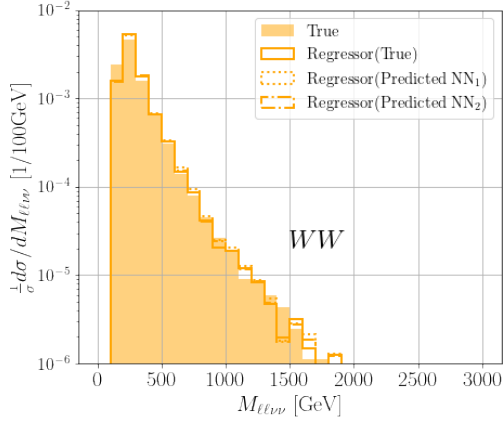


(i) Diferença entre a matriz confusão de NN_1 com NN_2 para o sinal de 2 TeV e largura de 1%.

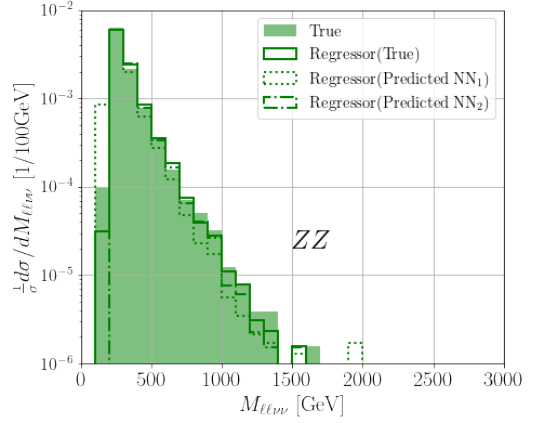
Figura B.2: Matriz confusão do classificador NN_1 (primeira linha,) NN_2 (segunda) e a diferença entre a matriz confusão de NN_1 com o NN_2 , ou seja, $NN_2 - NN_1$, para os sinais de H_2 com 1, 1.5 e 2 TeV com largura de 1%.

são similares aos casos de maior largura para NN_1 , enquanto NN_1 observamos um desempenho levemente superior para os casos com largura menor. A principal diferença entre as larguras são os valores de LR_+ e LR_- que são de uma ordem de grandeza diferente, mas possuem o mesmo comportamento.

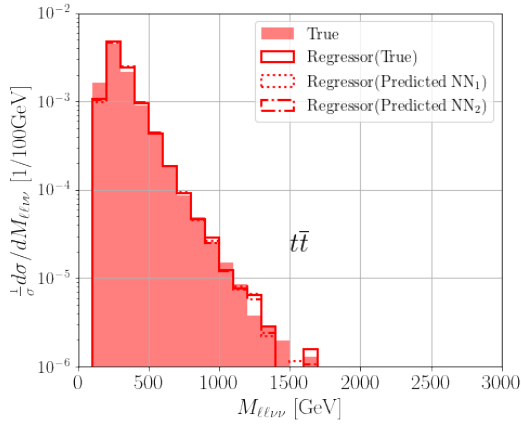
As distribuições de $M_{\ell\nu\nu}$ para o sinal de 2 TeV não contém contaminação do sinal nos backgrounds, figura B.3. A única contaminação que observamos são os backgrounds no sinal. Para massas menores a contaminação do sinal nos backgrounds é visível, mas em menor proporção que os casos com larguras maiores. Assim uma das melhorias que NN_2 tem em relação a NN_1 que é a correção da contaminação acontece em menor proporção nos casos de menores larguras.



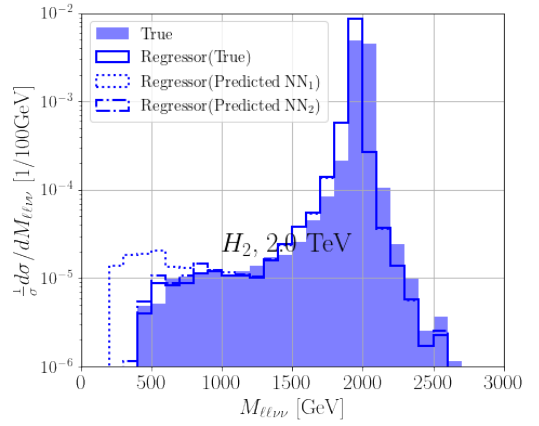
(a) Distribuição $M_{\ell\nu\nu}$ do background WW .



(b) Distribuição $M_{\ell\nu\nu}$ do background ZZ .



(c) Distribuição $M_{\ell\nu\nu}$ do background $t\bar{t}$.



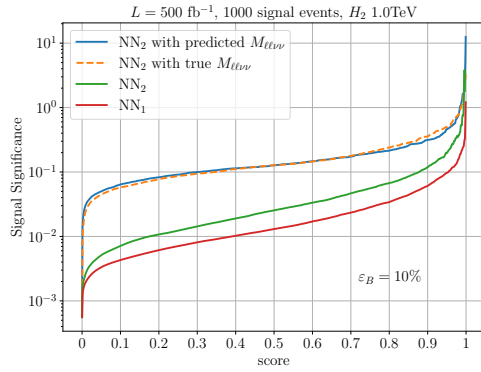
(d) Distribuição $M_{\ell\nu\nu}$ do sinal H_2 de 2 TeV com largura de 1%.

Figura B.3: Gráficos mostrando a distribuição $M_{\ell\nu\nu}$ verdadeira(fundo), com a regredida a partir dos dados reais(linha sólida), a regredida com os dados identificados com NN_1 (linha pontilhada) e a regredida com os dados identificados com NN_2 (linha semi-pontilhada). O azul representa o sinal H_2 de 2 TeV, laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

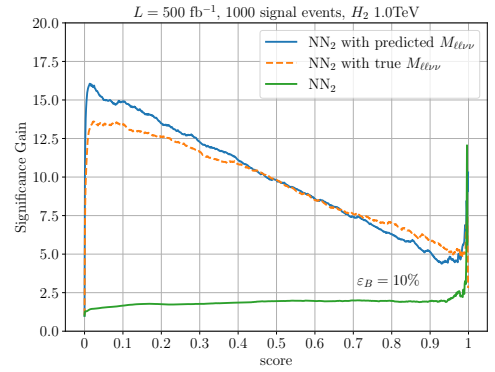
Métrica	1 TeV		1.5 TeV		2 TeV	
	NN ₁	NN ₂	NN ₁	NN ₂	NN ₁	NN ₂
<i>accuracy</i>	81.9%	87.1%	81.6%	86.1%	83.8%	88.1%
assimetria, H_2	0.843	0.977	0.883	0.938	0.999	0.999
assimetria, WW	-0.919	-0.989	-0.930	-0.985	-0.997	-0.999
assimetria, ZZ	-0.990	-0.997	-0.992	-0.998	-0.999	-0.999
assimetria, $t\bar{t}$	-0.919	-0.991	-0.939	-0.992	-0.996	-0.999
LR_+	4.533	6.625	4.434	6.039	5.131	6.974
LR_-	0.279	0.032	0.209	0.036	0.009	0.002
$AUC(H_2 \times WW)$	0.9862	0.9992	0.9910	0.9986	1.0000	1.0000
$AUC(H_2 \times ZZ)$	0.9965	0.9998	0.9981	0.9996	1.0000	1.0000
$AUC(H_2 \times t\bar{t})$	0.9878	0.9995	0.9924	0.9991	1.0000	1.0000

Tabela B.3: Comparação do desempenho das métricas obtidas das classificações NN₁ e NN₂ para identificar o bóson de Higgs com largura de 1% nas três massas consideradas.

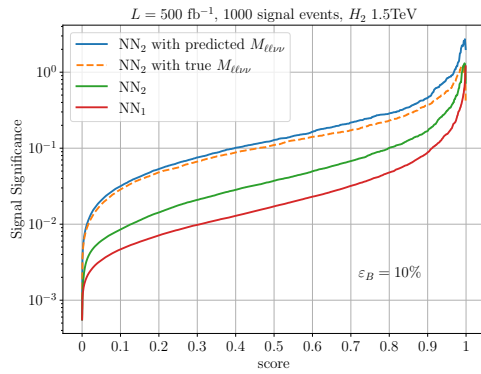
A figura B.4 mostra a significância estatística e o ganho. O comportamento é similar aos outros casos, onde observamos um ganho na significância estatística. A conclusão que podemos tirar é que não existe muita diferença em usar o k NNNN para um sinal com diferentes larguras. As figuras mostram um comportamento idêntico ao caso com largura de 10% com alguns aspectos diferindo em pequena proporção.



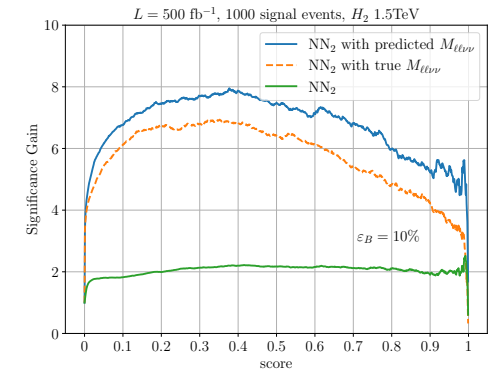
(a) Significância estatística de descoberta para o sinal H_2 de massa 1 TeV.



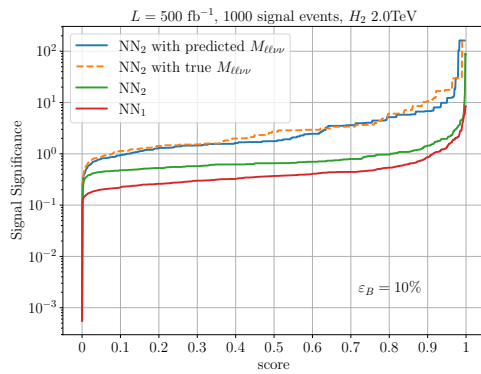
(b) Ganho da significância estatística de descoberta para o sinal H_2 de massa 1 TeV.



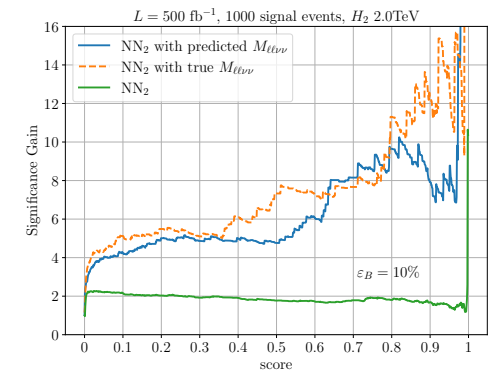
(c) Significância estatística de descoberta para o sinal H_2 de massa 1.5 TeV.



(d) Ganho da significância estatística de descoberta para o sinal H_2 de massa 1.5 TeV.



(e) Significância estatística de descoberta para o sinal H_2 de massa 2 TeV.



(f) Ganho da significância estatística de descoberta para o sinal H_2 de massa 2 TeV.

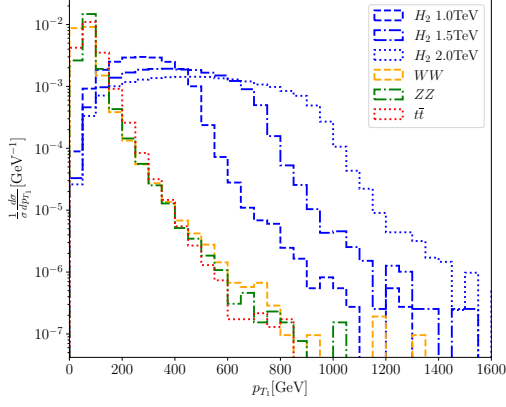
Figura B.4: A significância estatística do k NNNN algoritmo como uma função do corte na saída de scores (gráficos da esquerda) e do ganho da significância, eq.5.13, comparado a classificação NN_1 (gráficos da direita) para os sinais de H_2 com 1 TeV (primeira linha), 1.5 TeV(segunda linha) e 2 TeV(terceira linha). A incerteza sistemática é fixo em 10%.

Apêndice C

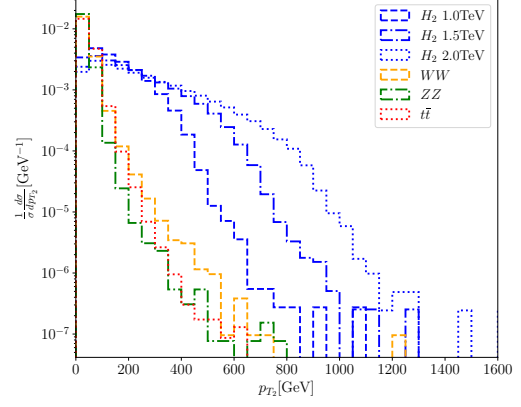
Resultados extras

C.1 Distribuições Cinemáticas

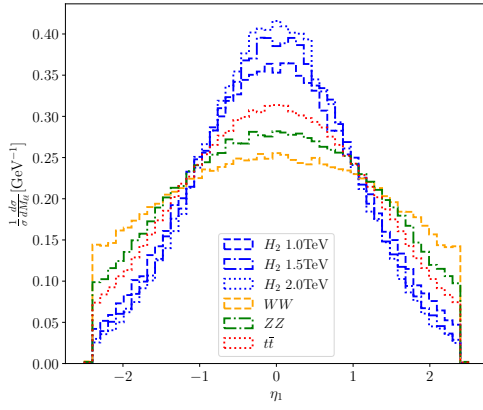
A figura C.1 mostra o restante das distribuições cinemáticas para os sinais de H_2 com largura de 10%, enquanto as figuras C.2, C.3, C.4 e C.5 mostram as distribuições cinemáticas para os sinais de H_2 com largura de 1% (C.2, C.3) e \tilde{e}_R (C.4, C.5), respectivamente. O comportamento das distribuições cinemáticas de H_2 com diferentes larguras são iguais, visto que a única mudança é a largura. Observamos que o comportamento das distribuições cinemáticas de \tilde{e}_R comparado aos sinais de H_2 são diferentes, por exemplo, o comportamento de $\Delta\phi_{\ell\ell}$ de \tilde{e}_R é similar aos backgrounds e, portanto, diferente das distribuições dos sinais de H_2 . O que pode explicar a diferença na regularização entre as redes neurais, principalmente, o dropout.



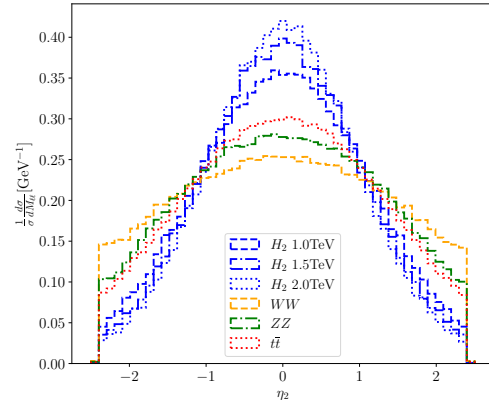
(a) Distribuição de p_{T_1}



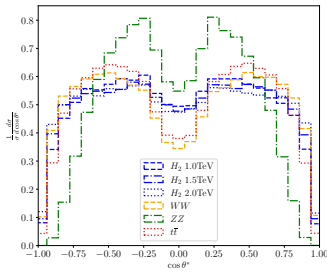
(b) Distribuição de p_{T_2}



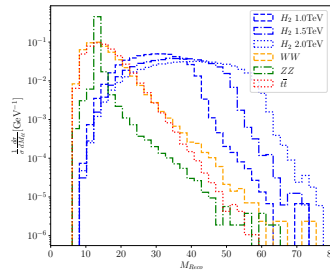
(c) Distribuição de η_1



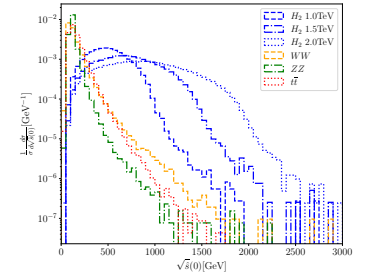
(d) Distribuição de η_2



(e) Distribuição de $\cos \theta^*$



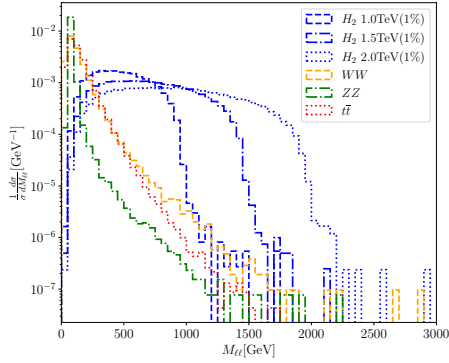
(f) Distribuição de M_{Reco}



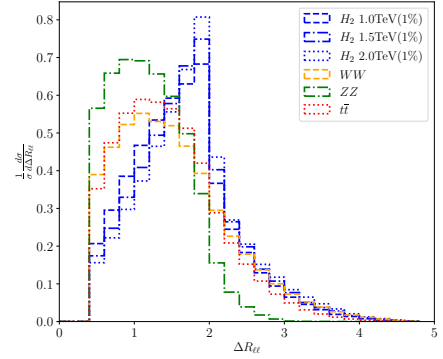
(g) Distribuição de $\sqrt{\hat{s}}(0)$

Figura C.1: Distribuições cinemáticas de algumas informações de alto nível dos backgrounds e sinais para representar os eventos nos algoritmos de ML. O azul representa os sinais de H_2 com largura de 10%, laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

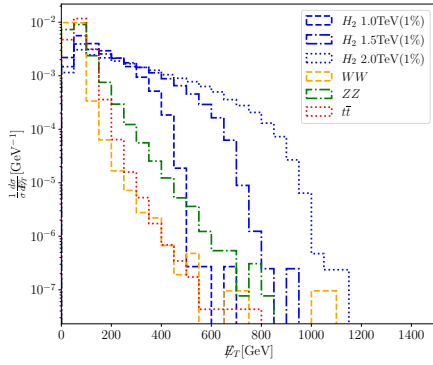
Distribuições cinemáticas do H_2 com largura de 1%.



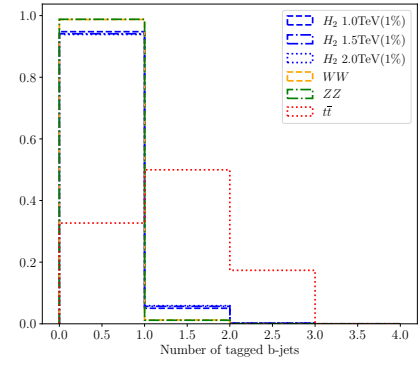
(a) Distribuição da massa invariante dos léptons $M_{\ell\ell}$



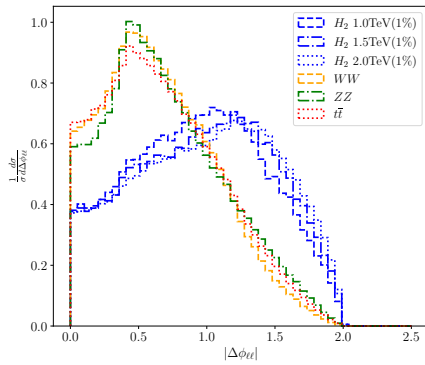
(b) Distribuição de $\Delta R_{\ell\ell}$



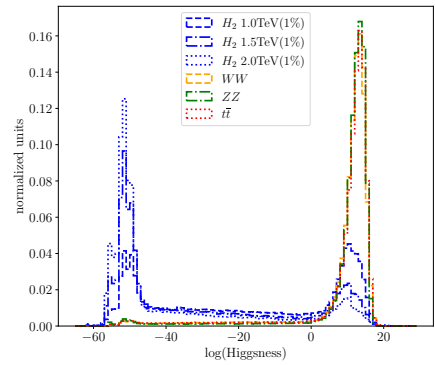
(c) Distribuição do missing energy E_T



(d) Distribuição do B-tag.



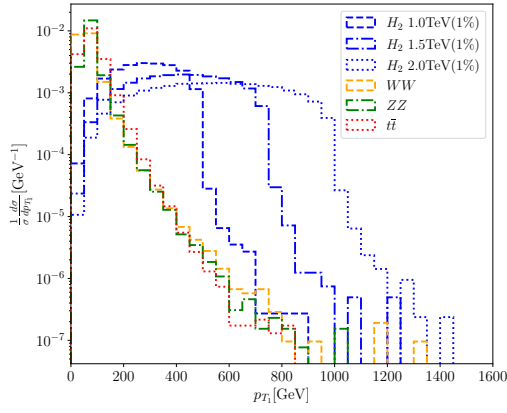
(e) Distribuição de $\Delta\phi_{\ell\ell}$



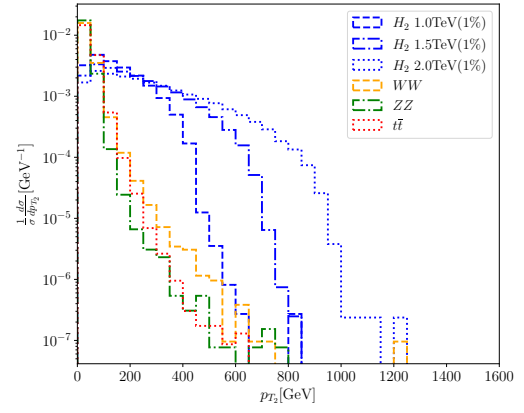
(f) Distribuição de Higgsness

Figura C.2: Algumas distribuições cinemáticas das informações de alto nível dos backgrounds e sinais de H_2 para representar os eventos nos algoritmos de ML. O azul representa os sinais H_2 , laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

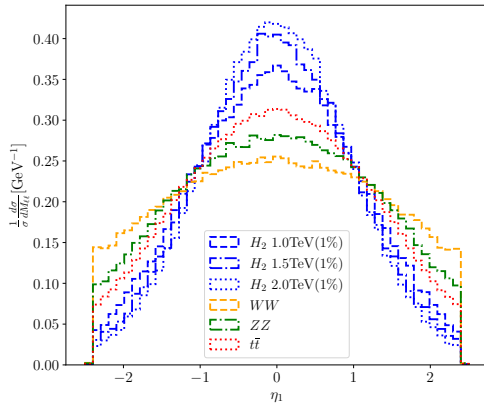
Distribuições cinemáticas do H_2 com largura de 1%.



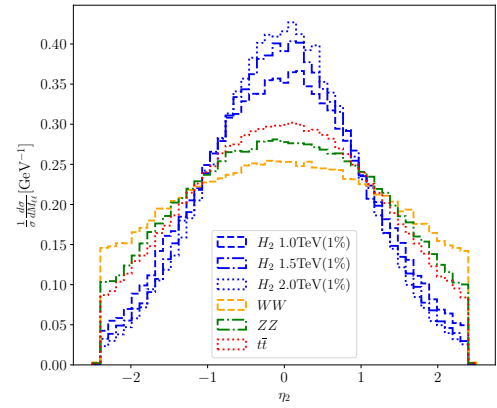
(a) Distribuição de p_{T1}



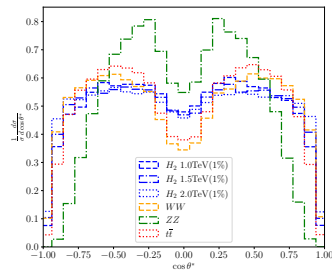
(b) Distribuição de p_{T2}



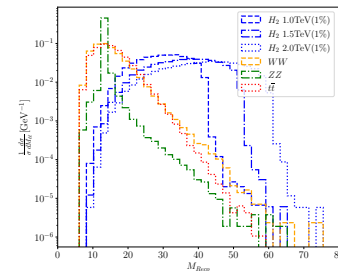
(c) Distribuição de η_1



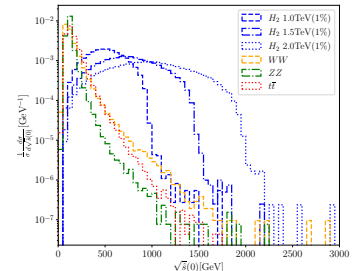
(d) Distribuição de η_2



(e) Distribuição de $\cos \theta^*$



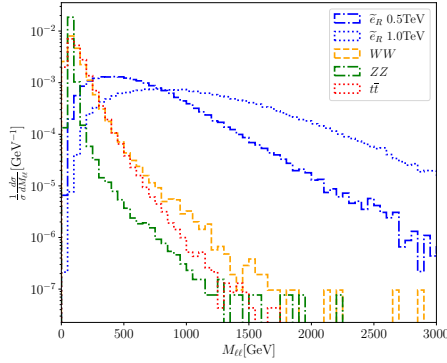
(f) Distribuição de M_{Reco}



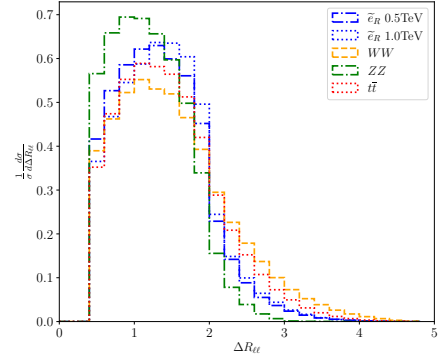
(g) Distribuição de $\sqrt{\hat{s}}(0)$

Figura C.3: Algumas distribuições cinemáticas das informações de alto nível dos backgrounds e sinais de H_2 para representar os eventos nos algoritmos de ML. O azul representa os sinais H_2 , laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

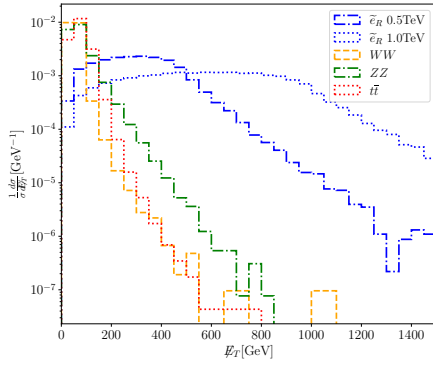
Distribuições cinemáticas do \tilde{e}_R .



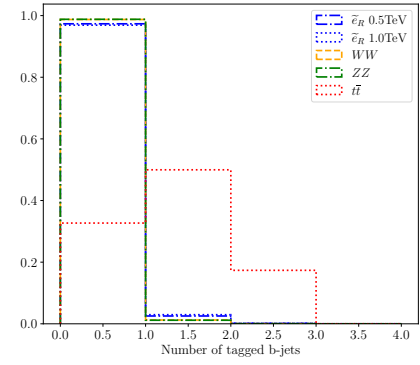
(a) Distribuição da massa invariante dos léptons $M_{\ell\ell}$



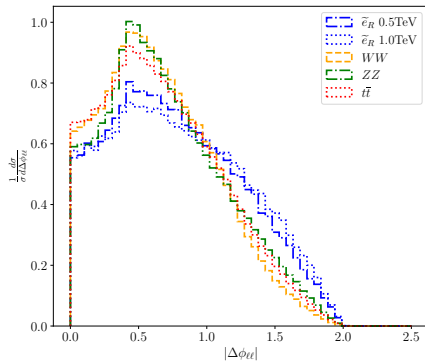
(b) Distribuição de $\Delta R_{\ell\ell}$



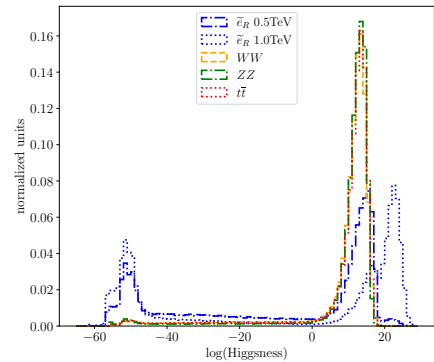
(c) Distribuição do missing energy E_T



(d) Distribuição do B-tag.



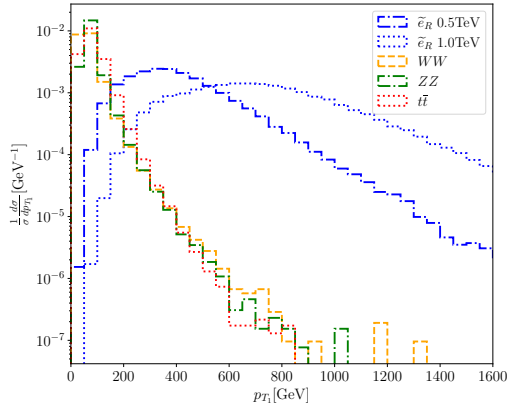
(e) Distribuição de $\Delta\phi_{\ell\ell}$



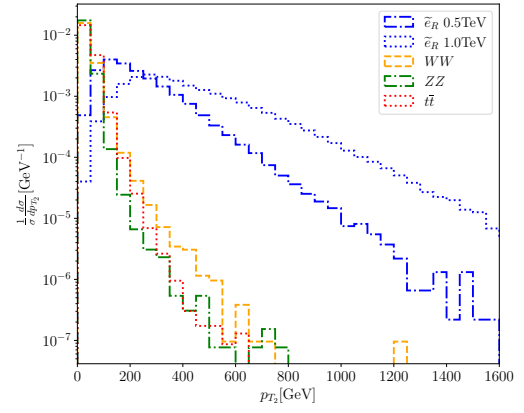
(f) Distribuição de Higgsness

Figura C.4: Algumas distribuições cinemáticas das informações de alto nível dos backgrounds e sinais de \tilde{e}_R para representar os eventos nos algoritmos de ML. O azul representa os sinais \tilde{e}_R , laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

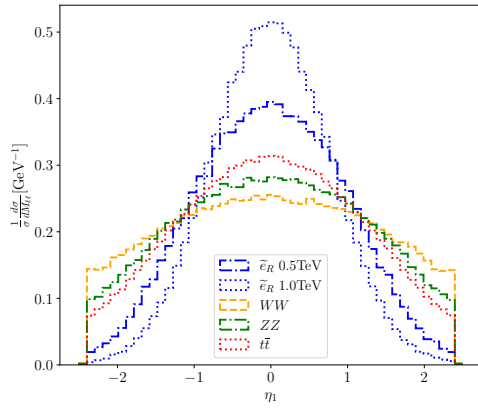
Distribuições cinemáticas do \tilde{e}_R .



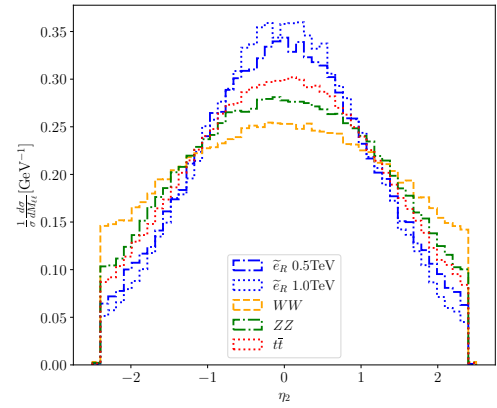
(a) Distribuição de p_{T_1}



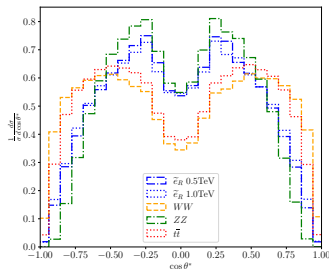
(b) Distribuição de p_{T_2}



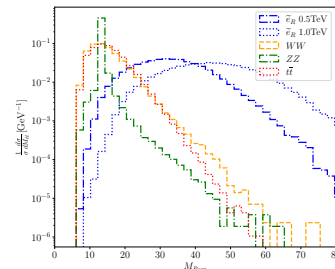
(c) Distribuição de η_1



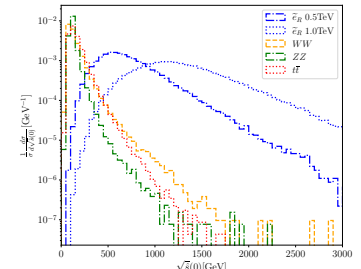
(d) Distribuição de η_2



(e) Distribuição de $\cos \theta^*$



(f) Distribuição de M_{Reco}

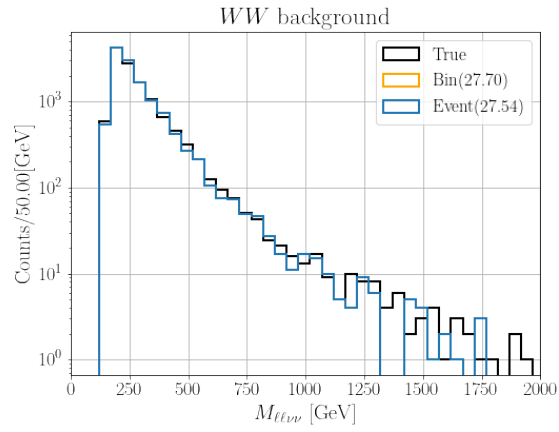


(g) Distribuição de $\sqrt{\hat{s}}(0)$

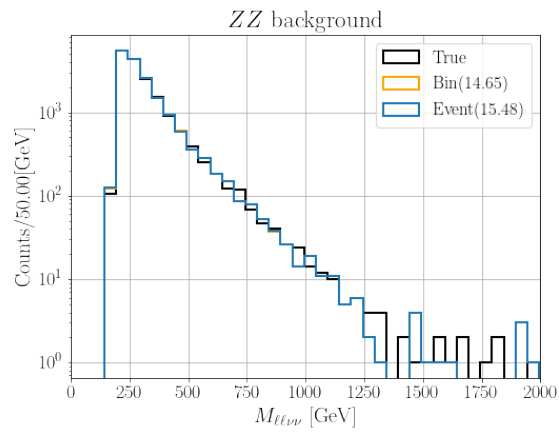
Figura C.5: Algumas distribuições cinemáticas das informações de alto nível dos backgrounds e sinais de \tilde{e}_R para representar os eventos nos algoritmos de ML. O azul representa os sinais \tilde{e}_R , laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

C.2 Comparação entre possíveis aplicações do k NN

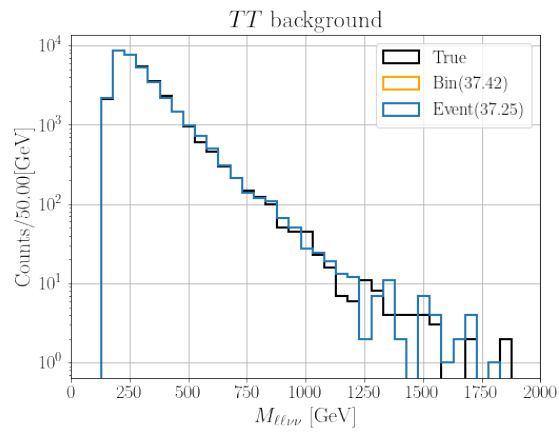
Nesta seção mostramos a comparação entre prever o valor ou o bin do evento, efetuada usando eventos verdadeiros na figura C.6. Cada previsão tem a binagem realizada diferentemente: (i) No primeiro caso, obtemos a previsão do valor dos eventos, fazemos a binagem dos eventos para obter distribuição de $M_{\ell\nu\nu}$. O cálculo do MSE foi realizado após a binagem. (ii) No segundo caso, fazemos a binagem para colocar no k NN e obter a previsão dos bins onde os eventos pertencem, o próximo passo é retornar aos valores originais e obter a distribuição de $M_{\ell\nu\nu}$ usando a binagem pré-previsão. O cálculo do MSE foi realizado após a obtenção da distribuição de $M_{\ell\nu\nu}$ usando a binagem pré-previsão. Na figura C.6 os valores em parêntesis são o MSE calculado, em todos os casos os valores obtidos são próximos um do outro. A diferença entre os valores é pequeno, pois as distribuições obtidas são basicamente idênticas com uma pequena diferença nas caudas do $t\bar{t}$. Assim decidimos que para as classes utilizadas, ambas as previsões são idênticas no resultado e podemos escolher qual método utilizar. Conforme escrito na seção 4.1 escolhemos prever os bins que os eventos pertencem.



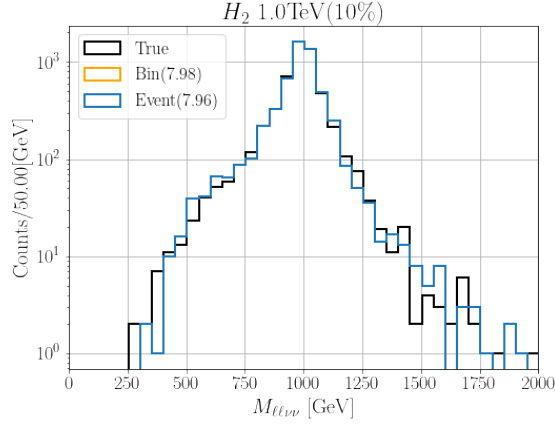
(a) Distribuição $M_{\ell\nu\nu}$ do background WW .



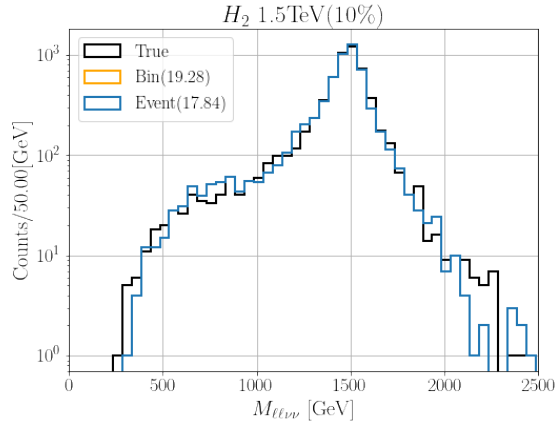
(b) Distribuição $M_{\ell\nu\nu}$ do background WW .



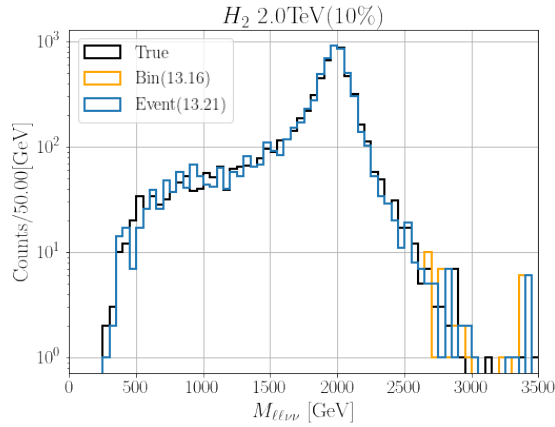
(c) Distribuição $M_{\ell\nu\nu}$ do background WW .



(d) Distribuição $M_{\ell\nu\nu}$ do sinal H_2 de 1 TeV.



(e) Distribuição $M_{\ell\nu\nu}$ do sinal H_2 de 1.5 TeV.



(f) Distribuição $M_{\ell\nu\nu}$ do sinal H_2 de 2 TeV.

Figura C.6: Gráficos mostrando a distribuição $M_{\ell\nu\nu}$ verdadeira (linha preta), com a predita a partir dos valores (linha azul) e a predita a partir dos bins do evento (linha laranja). Todas as distribuições foram obtidas usando eventos verdadeiros. Os painéis da primeira linha mostram os backgrounds e os painéis da segunda linha mostram os sinais de H_2 com 1, 1.5 e 2 TeV com largura de 10%.

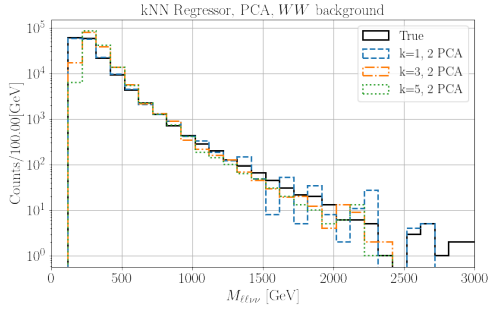
C.3 Hiperparâmetros do k NN

A tabela C.1 mostra todos os hiperparâmetros ajustados do k NN para todas as classes. Para todas as classes o peso é uniforme e o número de vizinhos iguais a $k = 1$. Conforme dito no capítulo 5, a distância é irrelevante. As figuras C.7, C.8, C.9, C.10, C.11, 5.2 mostram a comparação do efeito dos hiperparâmetros ao variar apenas um hiperparâmetro. Os hiperparâmetros padrões são os da tabela C.1. Nas figuras verificamos que variar a distância não tem diferença nos resultados, variar a dimensão de espaço latente tem uma pequena diferença e mudar o número de vizinhos k tem a maior diferença nos resultados.

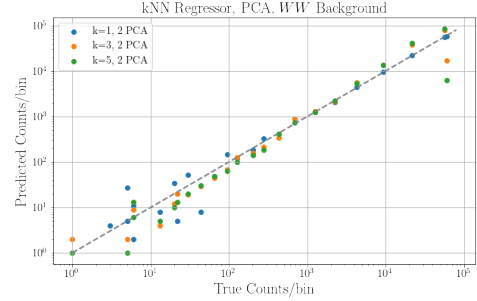
Classe	k	Distância	Peso	Dimensão do espaço latente
WW	1	Chebyshev	Uniform	2
ZZ	1	Minkowski	Uniform	1
$t\bar{t}$	1	Canberra	Uniform	8
H_2 1 TeV(10%)	1	Manhattan	Uniform	1
H_2 1.5TeV(10%)	1	Canberra	Uniform	3
H_2 2 TeV(10%)	1	Chebyshev	Uniform	6
H_2 1 TeV(1%)	1	Minkowski	Uniform	4
H_2 1.5 TeV(1%)	1	Canberra	Uniform	3
H_2 2 TeV(1%)	1	Canberra	Uniform	4
\tilde{e}_R 0.5 TeV	1	Chebyshev	Uniform	2
\tilde{e}_R 1 TeV	1	Minkowski	Uniform	3

Tabela C.1: Tabela mostrando os melhores hiperparâmetros encontrados pela busca em rede para a predição da distribuição $M_{\ell\nu\nu}$ de todas as classes.

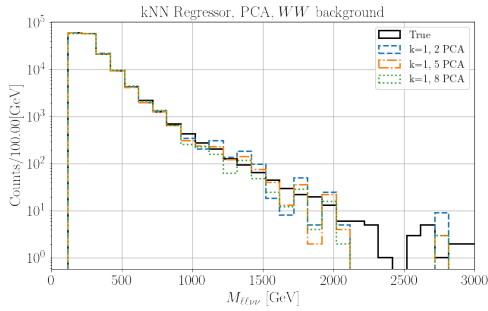
Background WW



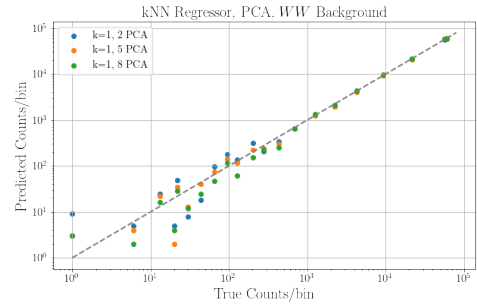
(a) Comparação entre as distribuições da massa invariante do sinal variando o número de vizinhos k . Os valores descritos na legenda são os hiperparâmetros: o k e o número dimensões projetadas pelo PCA



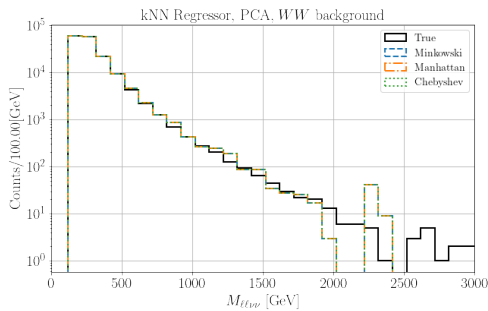
(b) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de vizinhos k .



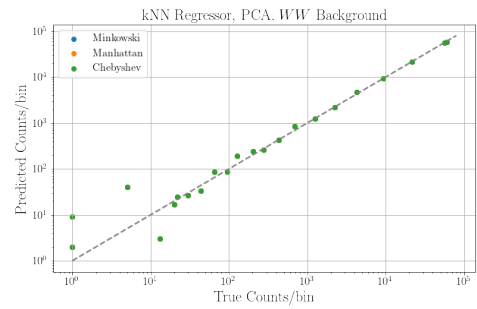
(c) Comparação entre as distribuições da massa invariante do sinal variando o número de dimensão projetado pelo PCA.



(d) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de dimensão projetado pelo PCA.



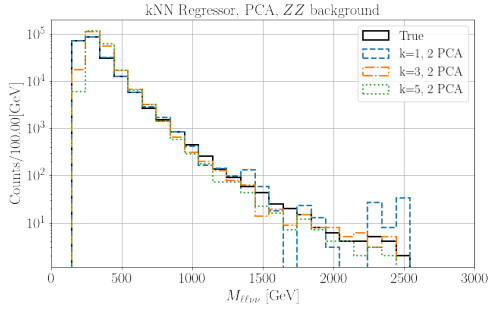
(e) Comparação entre as distribuições da massa invariante do sinal variando a métrica utilizada. A legenda mostra as métricas utilizadas.



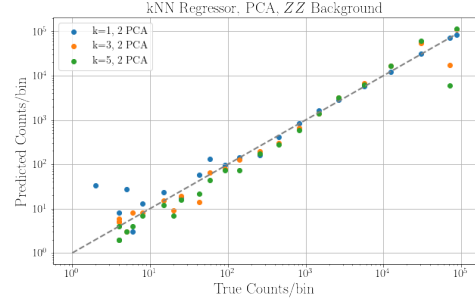
(f) Comparação do número de eventos dentro de um certo bin para os resultados obtidos variando a métrica utilizada.

Figura C.7: Os gráficos C.7a, C.7c, C.7e mostram a comparação da distribuição da massa invariante do sinal variando os hiperparâmetros com a real (linha preta). Os gráficos C.7b, C.7d, C.7f mostram a comparação do número de eventos para dado bin entre o evento verdadeiro (eixo x) e o predito pela regressão (eixo y), idealmente todos os pontos estariam na reta $y = x$ (linha pontilhada e cinza).

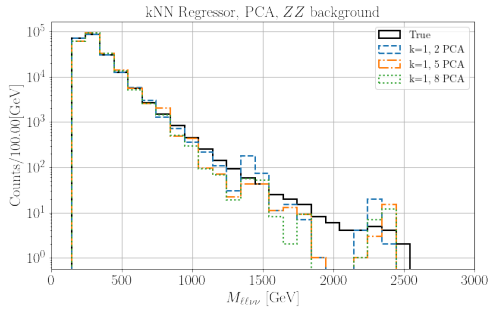
Background ZZ



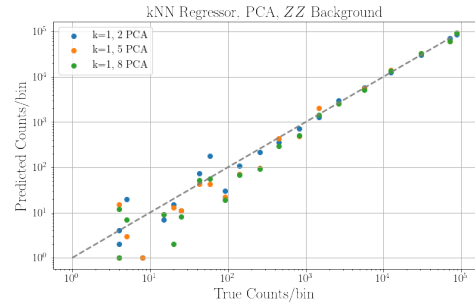
(a) Comparação entre as distribuições da massa invariante do sinal variando o número de vizinhos k . Os valores descritos na legenda são os hiperparâmetros: o k e o número dimensões projetadas pelo PCA



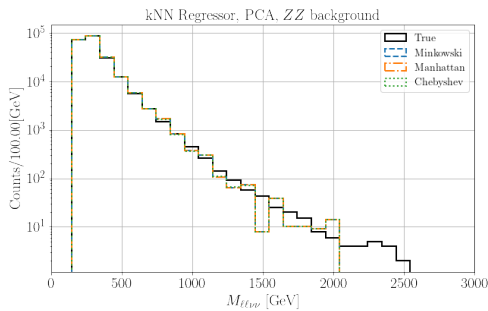
(b) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de vizinhos k .



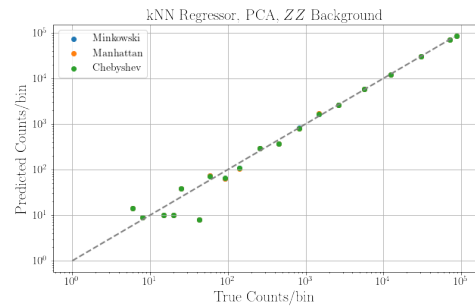
(c) Costra a comparação entre as distribuições da massa invariante do sinal variando o número de dimensão projetado pelo PCA.



(d) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de dimensão projetado pelo PCA.



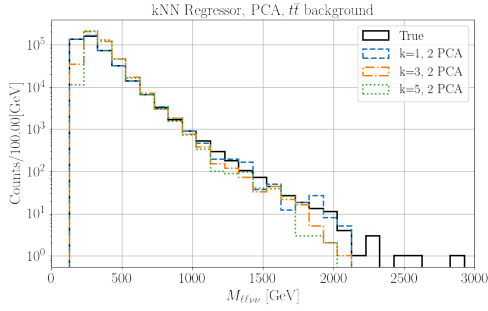
(e) Comparação entre as distribuições da massa invariante do sinal variando a métrica utilizada. A legenda mostra as métricas utilizadas.



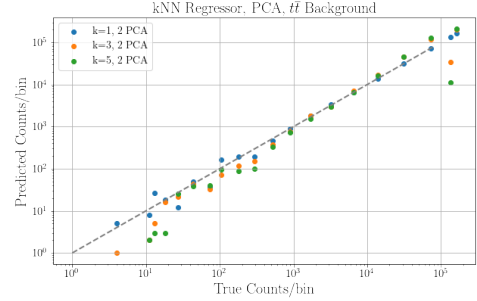
(f) Comparação do número de eventos dentro de um certo bin para os resultados obtidos variando a métrica utilizada.

Figura C.8: Os gráficos C.8a, C.8c, C.8e mostram a comparação da distribuição da massa invariante do sinal variando os hiperparâmetros com a real (linha preta). Os gráficos C.8b, C.8d, C.8f mostram a comparação do número de eventos para dado bin entre o evento verdadeiro (eixo x) e o predito pela regressão (eixo y), idealmente todos os pontos estariam na reta $y = x$ (linha pontilhada e cinza).

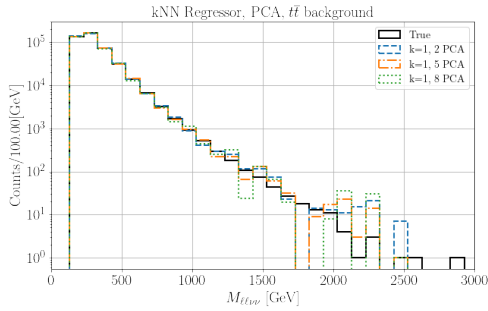
Background $t\bar{t}$



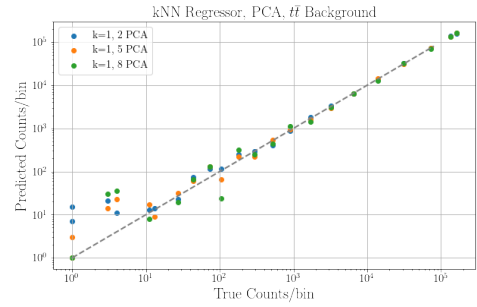
(a) Comparação entre as distribuições da massa invariante do sinal variando o número de vizinhos k . Os valores descritos na legenda são os hiperparâmetros: o k e o número dimensões projetadas pelo PCA



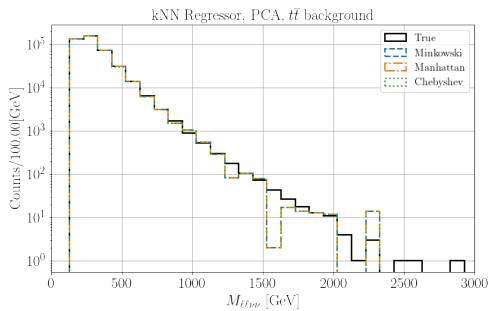
(b) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de vizinhos k .



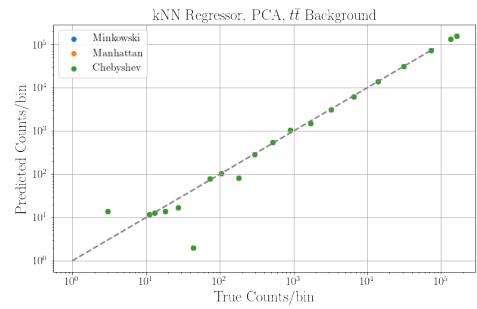
(c) Comparação entre as distribuições da massa invariante do sinal variando o número de dimensão projetado pelo PCA.



(d) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de dimensão projetado pelo PCA.



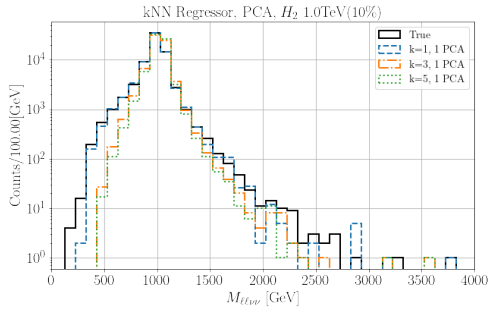
(e) Comparação entre as distribuições da massa invariante do sinal variando a métrica utilizada. A legenda mostra as métricas utilizadas.



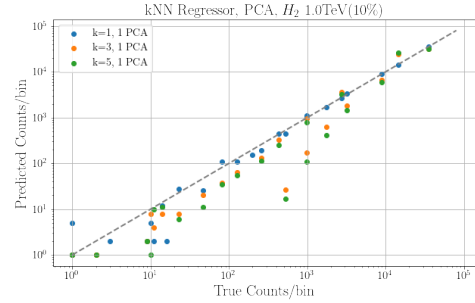
(f) Comparação do número de eventos dentro de um certo bin para os resultados obtidos variando a métrica utilizada.

Figura C.9: Os gráficos C.9a, C.9c, C.9e mostram a comparação da distribuição da massa invariante do sinal variando os hiperparâmetros com a real (linha preta). Os gráficos C.9b, C.9d, C.9f mostram a comparação do número de eventos para dado bin entre o evento verdadeiro (eixo x) e o predito pela regressão (eixo y), idealmente todos os pontos estariam na reta $y = x$ (linha pontilhada e cinza).

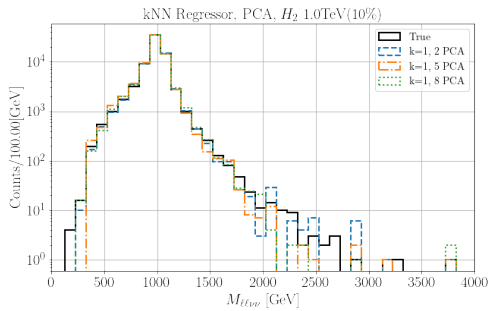
Sinal de H_2 com 1 TeV e largura de 10%.



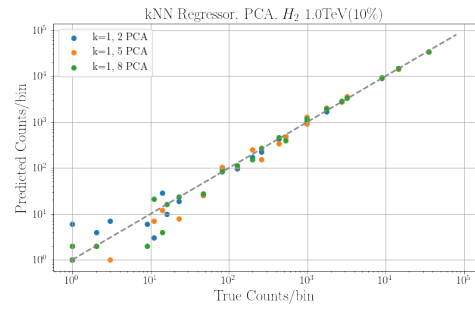
(a) Comparação entre as distribuições da massa invariante do sinal variando o número de vizinhos k . Os valores descritos na legenda são os hiperparâmetros: o k e o número dimensões projetadas pelo PCA



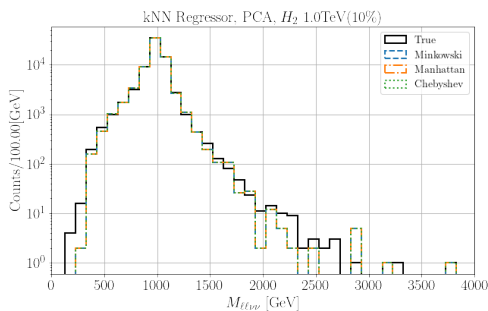
(b) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de vizinhos k .



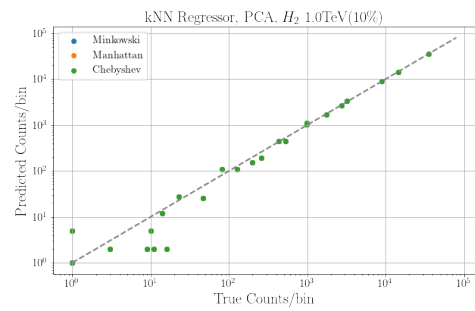
(c) Comparação entre as distribuições da massa invariante do sinal variando o número de dimensão projetado pelo PCA.



(d) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de dimensão projetado pelo PCA.



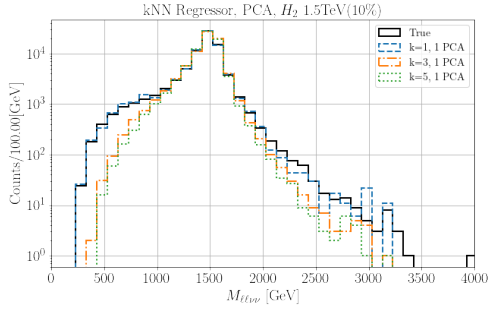
(e) Comparação entre as distribuições da massa invariante do sinal variando a métrica utilizada. A legenda mostra as métricas utilizadas.



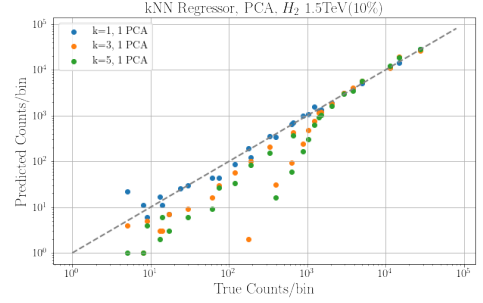
(f) Comparação do número de eventos dentro de um certo bin para os resultados obtidos variando a métrica utilizada.

Figura C.10: Os gráficos C.10a, C.10c, C.10e mostram a comparação da distribuição da massa invariante do sinal variando os hiperparâmetros com a real (linha preta). Os gráficos C.10b, C.10d, C.10f mostram a comparação do número de eventos para dado bin entre o evento verdadeiro (eixo x) e o predito pela regressão (eixo y), idealmente todos os pontos estariam na reta $y = x$ (linha pontilhada e cinza).

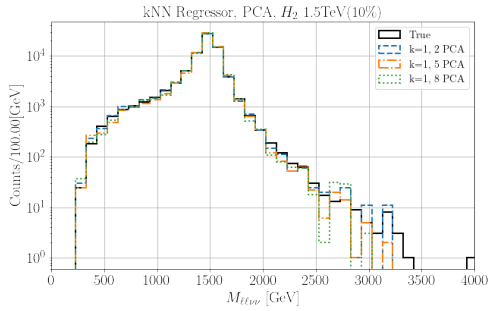
Sinal de H_2 com 1.5 TeV e largura de 10%.



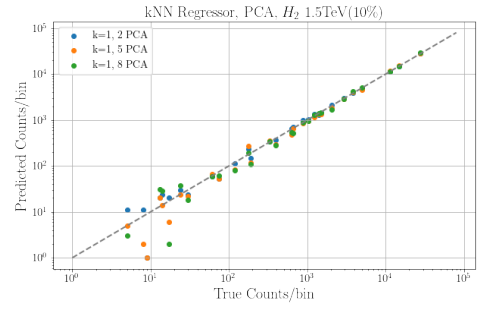
(a) Comparação entre as distribuições da massa invariante do sinal variando o número de vizinhos k . Os valores descritos na legenda são os hiperparâmetros: o k e o número dimensões projetadas pelo PCA



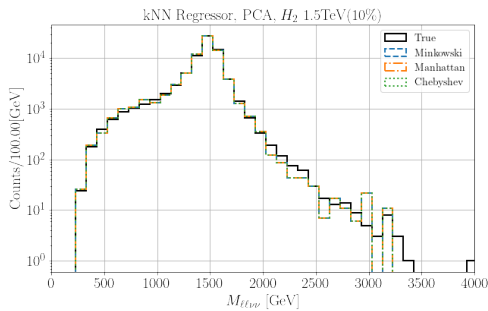
(b) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de vizinhos k .



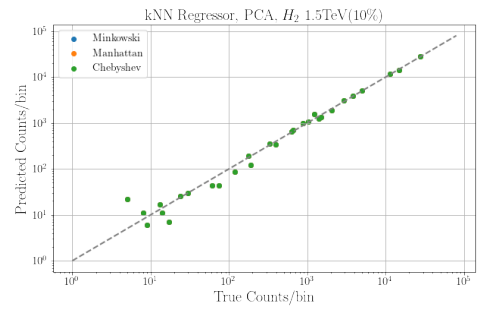
(c) Comparação entre as distribuições da massa invariante do sinal variando o número de dimensão projetado pelo PCA.



(d) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de dimensão projetado pelo PCA.



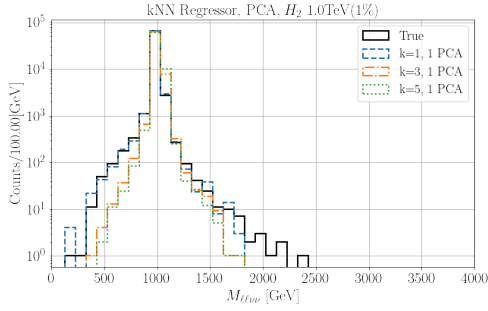
(e) Comparação entre as distribuições da massa invariante do sinal variando a métrica utilizada. A legenda mostra as métricas utilizadas.



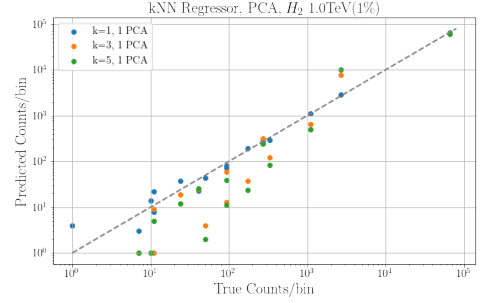
(f) Comparação do número de eventos dentro de um certo bin para os resultados obtidos variando a métrica utilizada.

Figura C.11: Os gráficos C.11a, C.11c, C.11e mostram a comparação da distribuição da massa invariante do sinal variando os hiperparâmetros com a real (linha preta). Os gráficos C.11b, C.11d, C.11f mostram a comparação do número de eventos para dado bin entre o evento verdadeiro (eixo x) e o predito pela regressão (eixo y), idealmente todos os pontos estariam na reta $y = x$ (linha pontilhada e cinza).

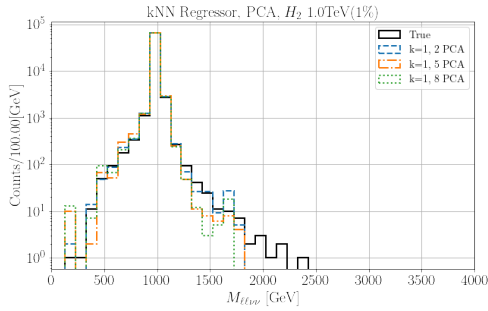
Sinal de H_2 com 1 TeV e largura de 1%.



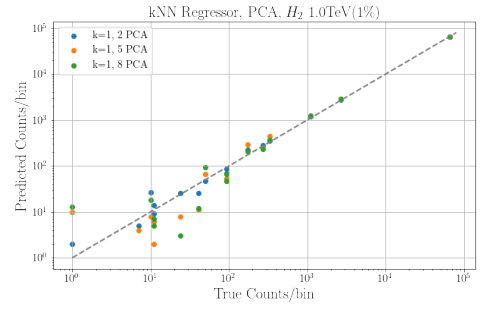
(a) Comparação entre as distribuições da massa invariante do sinal variando o número de vizinhos k . Os valores descritos na legenda são os hiperparâmetros: o k e o número dimensões projetadas pelo PCA



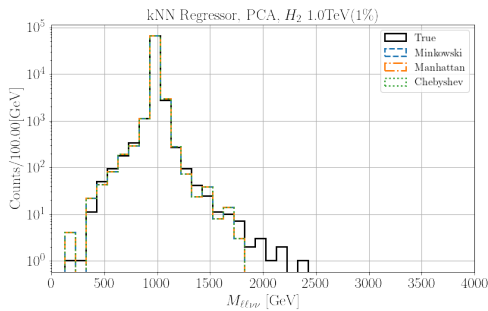
(b) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de vizinhos k .



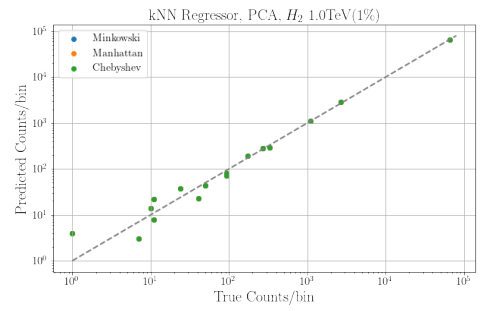
(c) Comparação entre as distribuições da massa invariante do sinal variando o número de dimensão projetado pelo PCA.



(d) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de dimensão projetado pelo PCA.



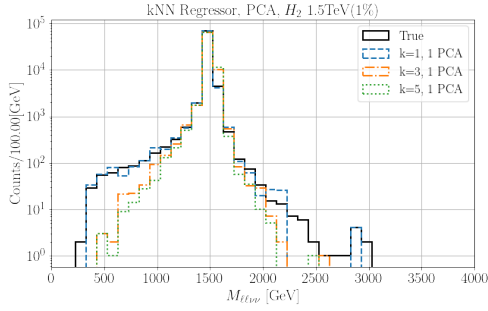
(e) Comparação entre as distribuições da massa invariante do sinal variando a métrica utilizada. A legenda mostra as métricas utilizadas.



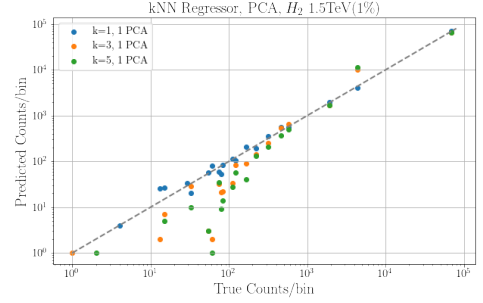
(f) Comparação do número de eventos dentro de um certo bin para os resultados obtidos variando a métrica utilizada.

Figura C.12: Os gráficos C.12a, C.12c, C.12e mostram a comparação da distribuição da massa invariante do sinal variando os hiperparâmetros com a real (linha preta). Os gráficos C.12b, C.12d, C.12f mostram a comparação do número de eventos para dado bin entre o evento verdadeiro (eixo x) e o predito pela regressão (eixo y), idealmente todos os pontos estariam na reta $y = x$ (linha pontilhada e cinza).

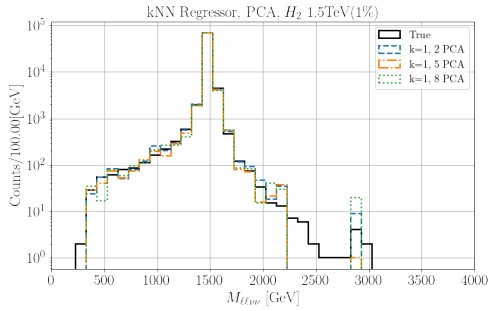
Sinal de H_2 com 1.5 TeV e largura de 1%.



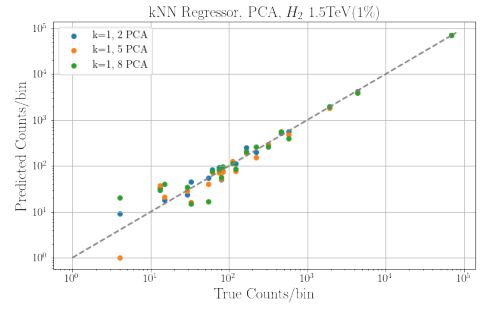
(a) Comparação entre as distribuições da massa invariante do sinal variando o número de vizinhos k . Os valores descritos na legenda são os hiperparâmetros: o k e o número dimensões projetadas pelo PCA



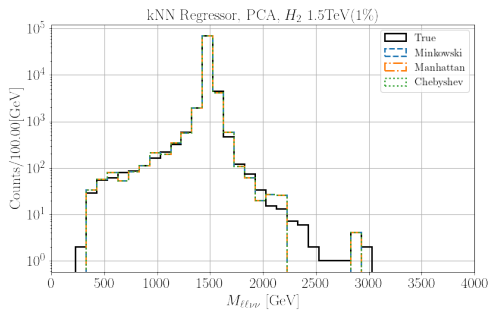
(b) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de vizinhos k .



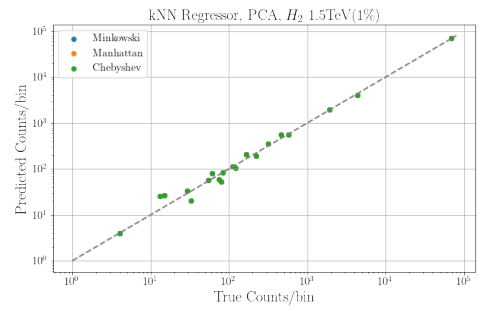
(c) Comparação entre as distribuições da massa invariante do sinal variando o número de dimensão projetado pelo PCA.



(d) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de dimensão projetado pelo PCA.



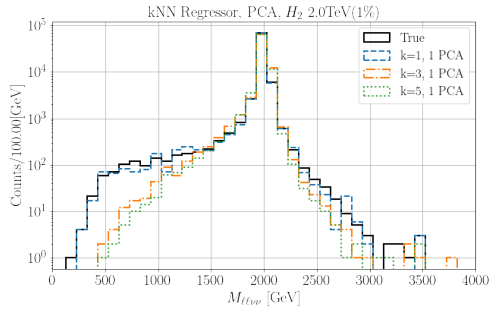
(e) Comparação entre as distribuições da massa invariante do sinal variando a métrica utilizada. A legenda mostra as métricas utilizadas.



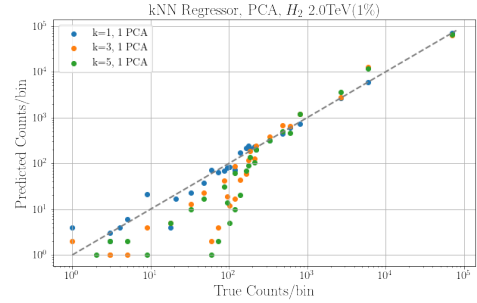
(f) Comparação do número de eventos dentro de um certo bin para os resultados obtidos variando a métrica utilizada.

Figura C.13: Os gráficos C.13a, C.13c, C.11e mostram a comparação da distribuição da massa invariante do sinal variando os hiperparâmetros com a real (linha preta). Os gráficos C.13b, C.13d, C.13f mostram a comparação do número de eventos para dado bin entre o evento verdadeiro (eixo x) e o predito pela regressão (eixo y), idealmente todos os pontos estariam na reta $y = x$ (linha pontilhada e cinza).

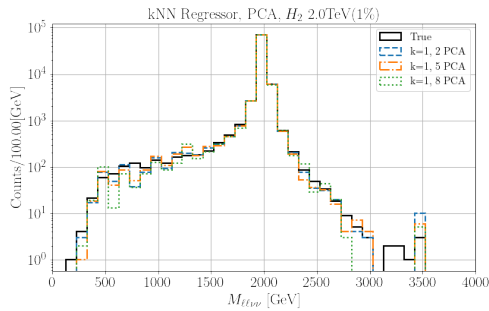
Sinal de H_2 com 2 TeV e largura de 1%.



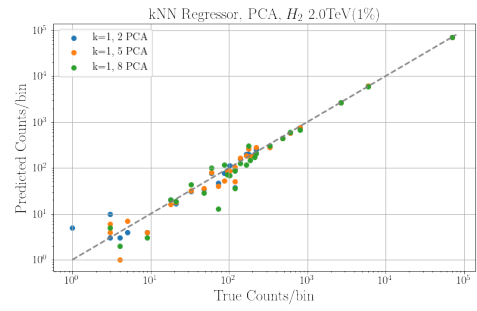
(a) Comparação entre as distribuições da massa invariante do sinal variando o número de vizinhos k . Os valores descritos na legenda são os hiperparâmetros: o k e o número dimensões projetadas pelo PCA



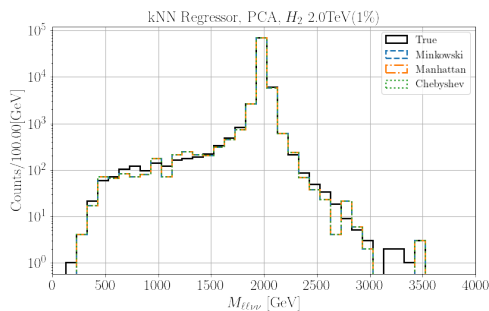
(b) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de vizinhos k .



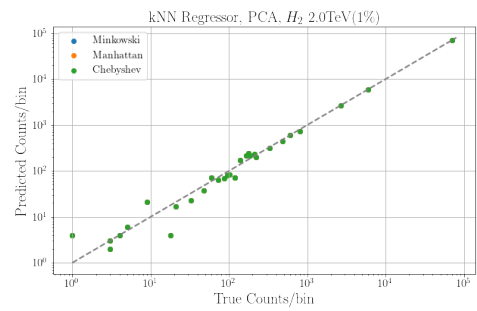
(c) Comparação entre as distribuições da massa invariante do sinal variando o número de dimensão projetado pelo PCA.



(d) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de dimensão projetado pelo PCA.



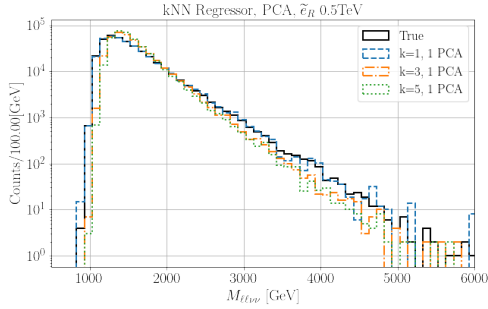
(e) Comparação entre as distribuições da massa invariante do sinal variando a métrica utilizada. A legenda mostra as métricas utilizadas.



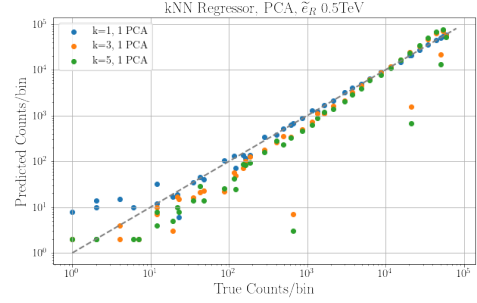
(f) Comparação do número de eventos dentro de um certo bin para os resultados obtidos variando a métrica utilizada.

Figura C.14: Os gráficos C.14a, C.14c, 5.2e mostram a comparação da distribuição da massa invariante do sinal variando os hiperparâmetros com a real(linha preta). Os gráficos C.14b, C.14d, C.14f mostram a comparação do número de eventos para dado bin entre o evento verdadeiro(eixo x) e o predito pela regressão(eixo y), idealmente todos os pontos estariam na reta $y = x$ (linha pontilhada e cinza).

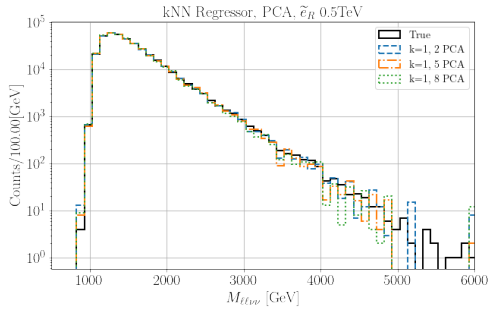
Sinal de \tilde{e}_R com 0.5 TeV.



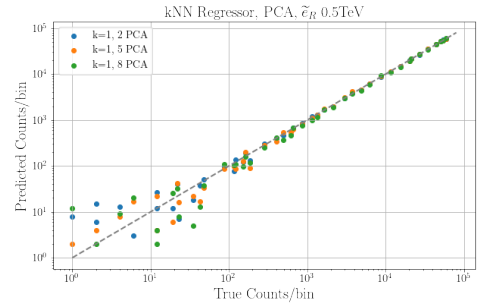
(a) Comparação entre as distribuições da massa invariante do sinal variando o número de vizinhos k . Os valores descritos na legenda são os hiperparâmetros: o k e o número dimensões projetadas pelo PCA



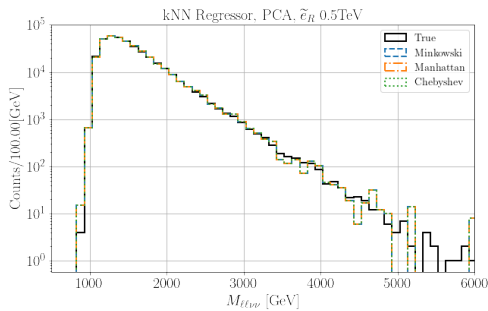
(b) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de vizinhos k .



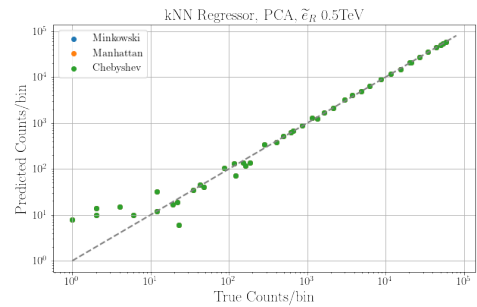
(c) Comparação entre as distribuições da massa invariante do sinal variando o número de dimensão projetado pelo PCA.



(d) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de dimensão projetado pelo PCA.



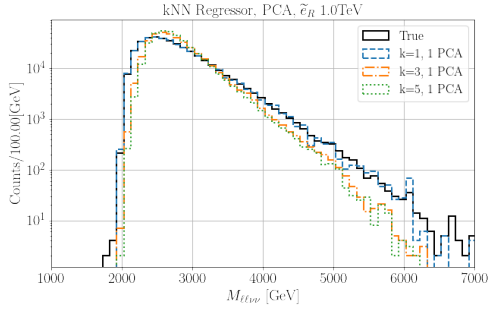
(e) Comparação entre as distribuições da massa invariante do sinal variando a métrica utilizada. A legenda mostra as métricas utilizadas.



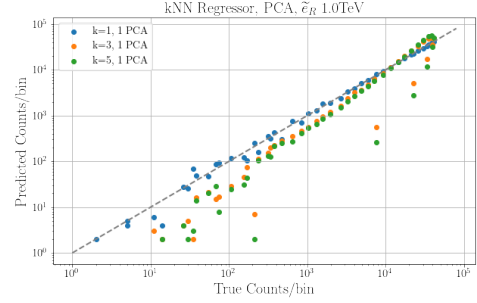
(f) Comparação do número de eventos dentro de um certo bin para os resultados obtidos variando a métrica utilizada.

Figura C.15: Os gráficos C.15a, C.15c, C.15e mostram a comparação da distribuição da massa invariante do sinal variando os hiperparâmetros com a real (linha preta). Os gráficos C.15b, C.15d, C.15f mostram a comparação do número de eventos para dado bin entre o evento verdadeiro (eixo x) e o predito pela regressão (eixo y), idealmente todos os pontos estariam na reta $y = x$ (linha pontilhada e cinza).

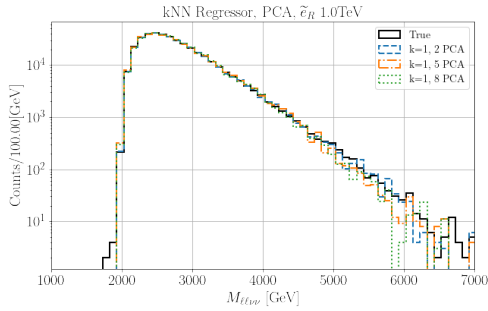
Sinal de \tilde{e}_R com 1 TeV.



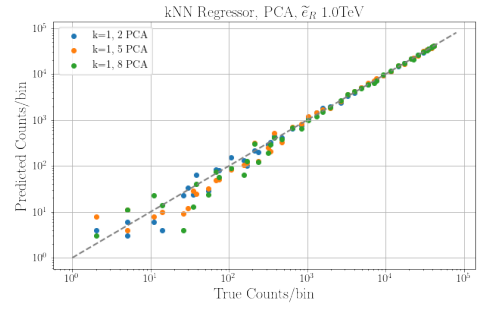
(a) Comparação entre as distribuições da massa invariante do sinal variando o número de vizinhos k . Os valores descritos na legenda são os hiperparâmetros: o k e o número dimensões projetadas pelo PCA



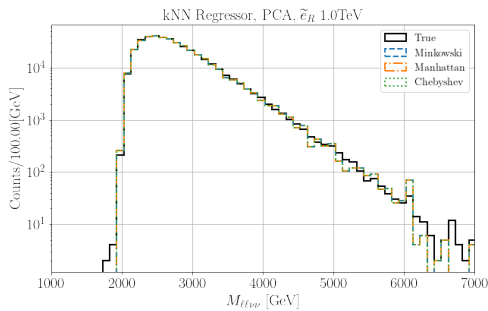
(b) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de vizinhos k .



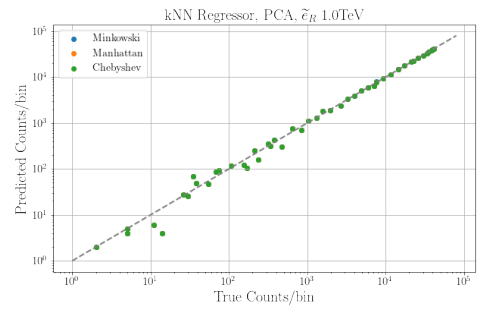
(c) Comparação entre as distribuições da massa invariante do sinal variando o número de dimensão projetado pelo PCA.



(d) Comparação do número de eventos dentro do bin para os resultados obtidos variando o número de dimensão projetado pelo PCA.



(e) Comparação entre as distribuições da massa invariante do sinal variando a métrica utilizada. A legenda mostra as métricas utilizadas.



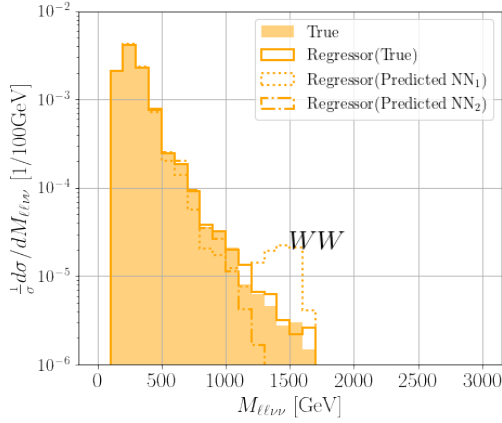
(f) Comparação do número de eventos dentro de um certo bin para os resultados obtidos variando a métrica utilizada.

Figura C.16: Os gráficos C.16a, C.16c, C.16e mostram a comparação da distribuição da massa invariante do sinal variando os hiperparâmetros com a real (linha preta). Os gráficos C.16b, C.16d, C.16f mostram a comparação do número de eventos para dado bin entre o evento verdadeiro (eixo x) e o predito pela regressão (eixo y), idealmente todos os pontos estariam na reta $y = x$ (linha pontilhada e cinza).

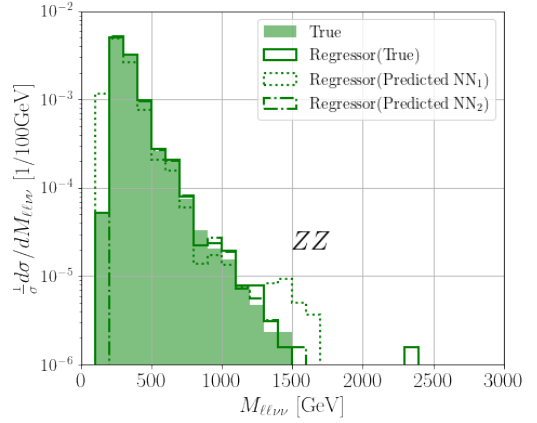
C.4 Distribuições da massa invariante para outros casos.

As figuras C.17, C.18, C.19, C.20, C.21 mostram a distribuição da massa invariante $M_{\ell\nu\nu}$ verdadeira para os outros sinais, regredida a partir de dados verdadeiros, após NN_1 e NN_2 .

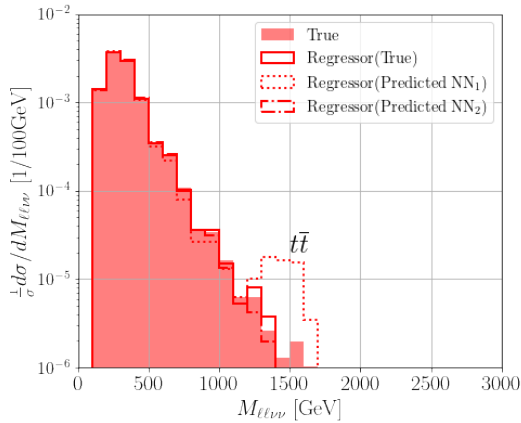
Sinal de H_2 com 1.5 TeV e largura de 10%.



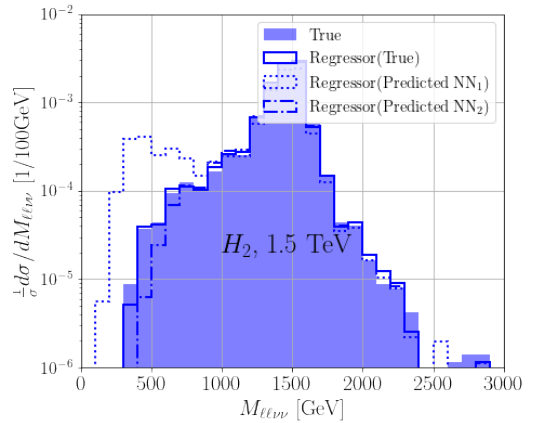
(a) Distribuição $M_{\ell\nu\nu}$ do background WW .



(b) Distribuição $M_{\ell\nu\nu}$ do background ZZ .



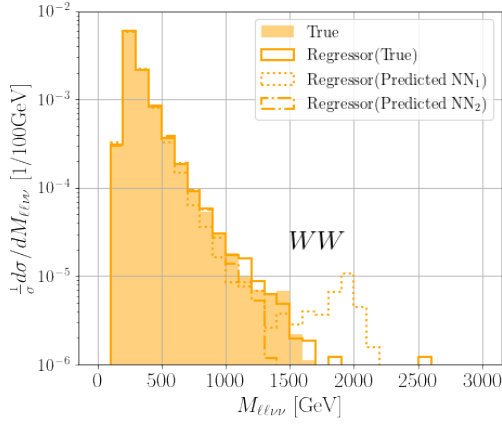
(c) Distribuição $M_{\ell\nu\nu}$ do background $t\bar{t}$.



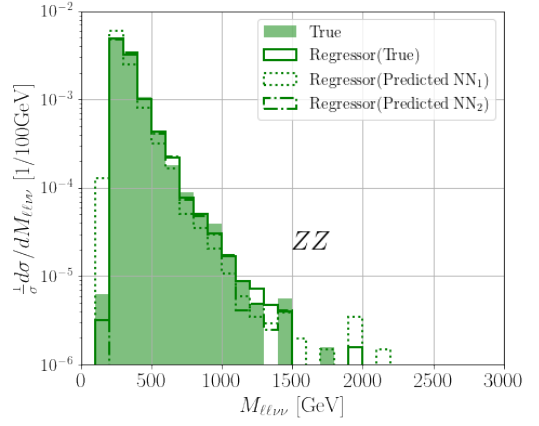
(d) Distribuição $M_{\ell\nu\nu}$ do sinal H_2 de 1.5 TeV.

Figura C.17: Gráficos mostrando a distribuição $M_{\ell\nu\nu}$ verdadeira(fundo), com a regredida a partir dos dados reais(linha sólida), a regredida com os dados identificados com NN_1 (linha pontilhada) para o sinal H_2 de 1.5 TeV e a regredida com os dados identificados com NN_2 (linha semi-pontilhada). O azul representa o sinal, laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

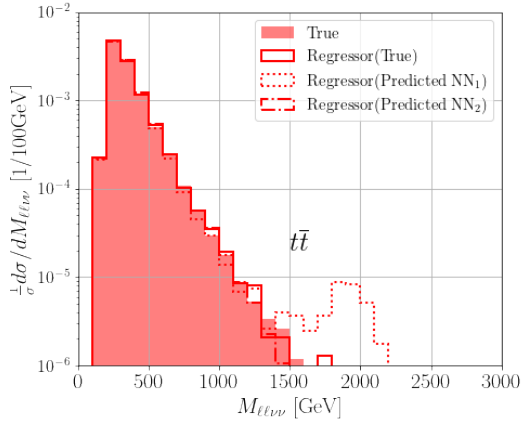
Sinal de H_2 com 2 TeV e largura de 10%.



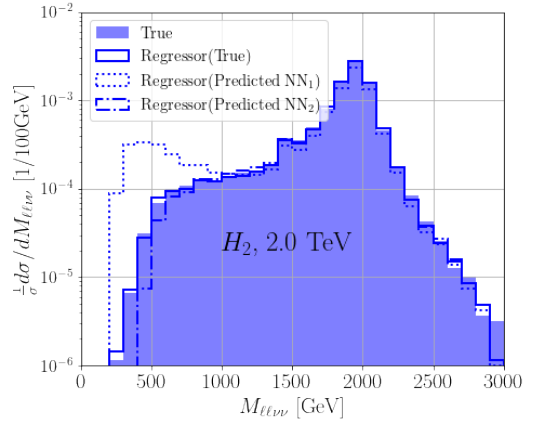
(a) Distribuição $M_{\ell\nu\nu}$ do background WW .



(b) Distribuição $M_{\ell\nu\nu}$ do background ZZ .



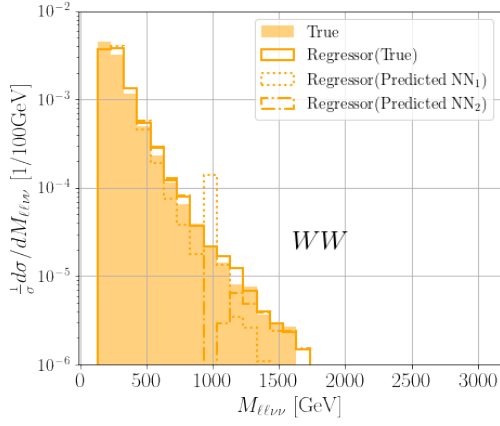
(c) Distribuição $M_{\ell\nu\nu}$ do background $t\bar{t}$.



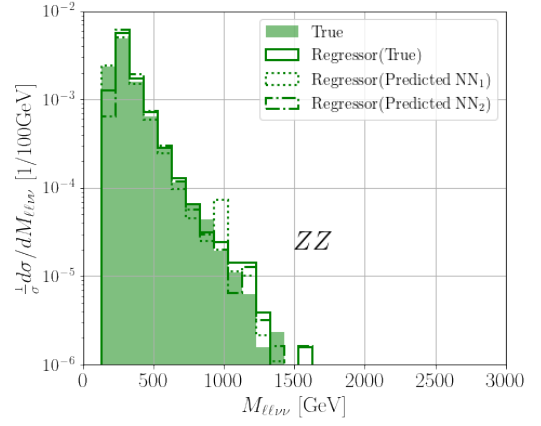
(d) Distribuição $M_{\ell\nu\nu}$ do sinal H_2 de 2 TeV.

Figura C.18: Gráficos mostrando a distribuição $M_{\ell\nu\nu}$ verdadeira(fundo), com a regredida a partir dos dados reais(linha sólida), a regredida com os dados identificados com NN_1 (linha pontilhada) para o sinal H_2 de 2 TeV e a regredida com os dados identificados com NN_2 (linha semi-pontilhada). O azul representa o sinal, laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

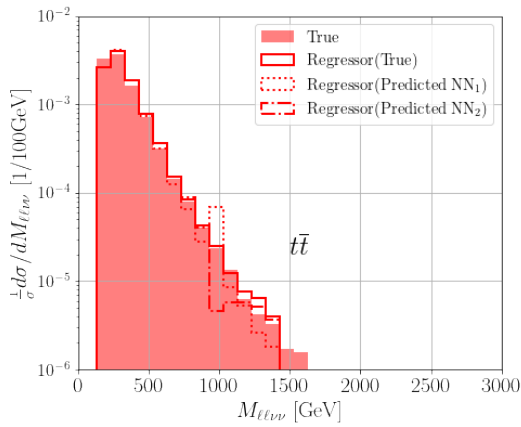
Sinal de H_2 com 1 TeV e largura de 1%.



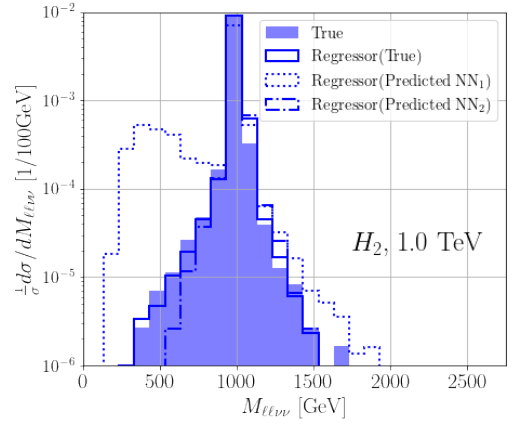
(a) Distribuição $M_{\ell\nu\nu}$ do background WW .



(b) Distribuição $M_{\ell\nu\nu}$ do background ZZ .



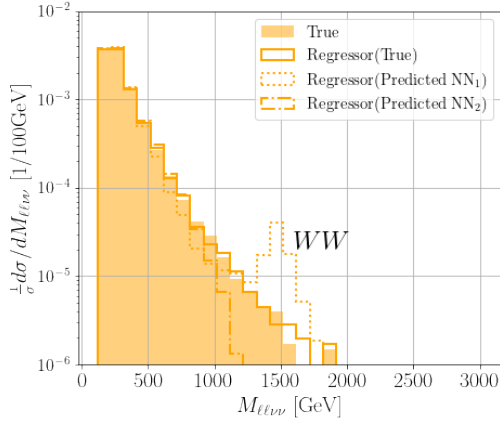
(c) Distribuição $M_{\ell\nu\nu}$ do background $t\bar{t}$.



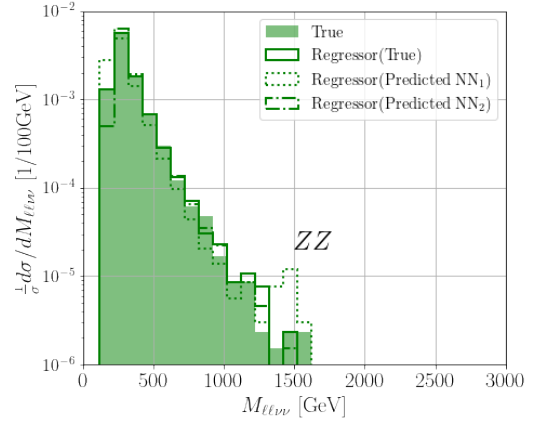
(d) Distribuição $M_{\ell\nu\nu}$ do sinal H_2 de 1 TeV com largura de 1%.

Figura C.19: Gráficos mostrando a distribuição $M_{\ell\nu\nu}$ verdadeira(fundo), com a regredida a partir dos dados reais(linha sólida), a regredida com os dados identificados com NN_1 (linha pontilhada) e a regredida com os dados identificados com NN_2 (linha semi-pontilhada). O azul representa o sinal H_2 de 1 TeV, laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

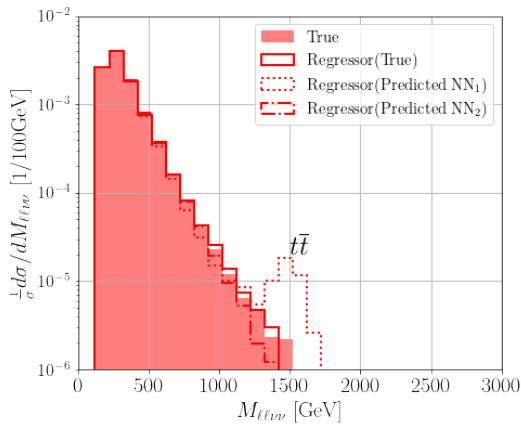
Sinal de H_2 com 1.5 TeV e largura de 1%.



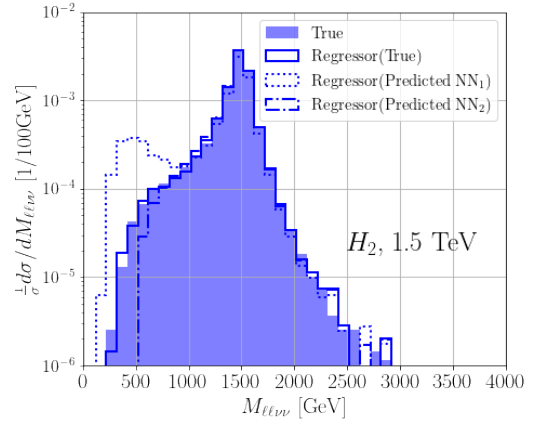
(a) Distribuição $M_{\ell\ell\nu\nu}$ do background WW .



(b) Distribuição $M_{\ell\ell\nu\nu}$ do background ZZ .



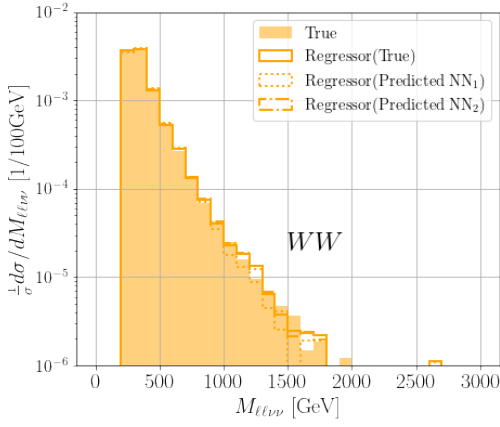
(c) Distribuição $M_{\ell\ell\nu\nu}$ do background $t\bar{t}$.



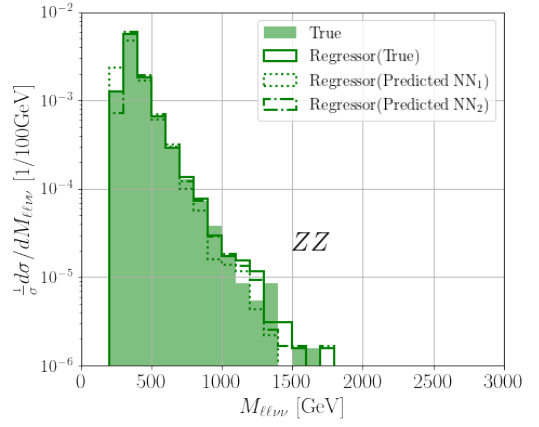
(d) Distribuição $M_{\ell\ell\nu\nu}$ do sinal H_2 de 1.5 TeV com largura de 1%.

Figura C.20: Gráficos mostrando a distribuição $M_{\ell\ell\nu\nu}$ verdadeira(fundo), com a regredida a partir dos dados reais(linha sólida), a regredida com os dados identificados com NN_1 (linha pontilhada) e a regredida com os dados identificados com NN_2 (linha semi-pontilhada). O azul representa o sinal H_2 de 1.5 TeV, laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

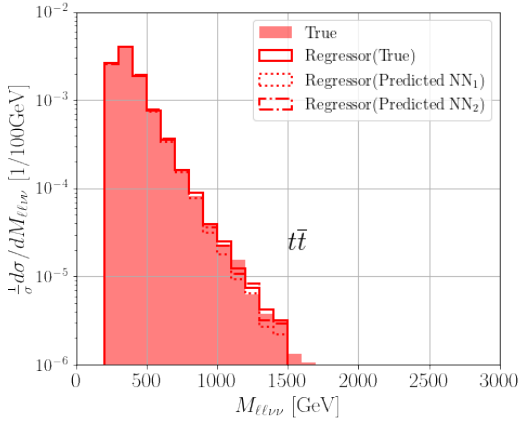
Sinal de \tilde{e}_R com 1 TeV e neutralino de 0.1 TeV.



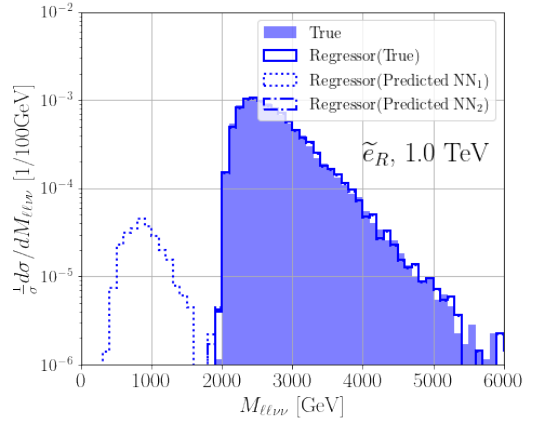
(a) Distribuição $M_{\ell\nu\nu}$ do background WW .



(b) Distribuição $M_{\ell\nu\nu}$ do background ZZ .



(c) Distribuição $M_{\ell\nu\nu}$ do background $t\bar{t}$.



(d) Distribuição $M_{\ell\nu\nu}$ do sinal \tilde{e}_R de 1 TeV.

Figura C.21: Gráficos mostrando a distribuição $M_{\ell\nu\nu}$ verdadeira(fundo), com a regredida a partir dos dados reais(linha sólida), a regredida com os dados identificados com NN_1 (linha pontilhada) para o sinal \tilde{e}_R de 1 TeV e a regredida com os dados identificados com NN_2 (linha semi-pontilhada). O azul representa o sinal, laranja representa o background WW , verde representa o background ZZ e vermelho representa o background $t\bar{t}$.

Referências Bibliográficas

- [1] Edvige Corbelli and Paolo Salucci. The Extended Rotation Curve and the Dark Matter Halo of M33. *Mon. Not. Roy. Astron. Soc.*, 311:441–447, 2000.
- [2] Steven W. Allen, August E. Evrard, and Adam B. Mantz. Cosmological parameters from observations of galaxy clusters. *Annual Review of Astronomy and Astrophysics*, 49(1):409–470, Sep 2011.
- [3] Douglas Clowe, Marusa Bradac, Anthony H. Gonzalez, Maxim Markevitch, Scott W. Randall, Christine Jones, and Dennis Zaritsky. A direct empirical proof of the existence of dark matter. *Astrophys. J. Lett.*, 648:L109–L113, 2006.
- [4] Alexandre Refregier. Weak gravitational lensing by large scale structure. *Ann. Rev. Astron. Astrophys.*, 41:645–668, 2003.
- [5] Priyamvada Natarajan et al. Mapping substructure in the HST Frontier Fields cluster lenses and in cosmological simulations. *Mon. Not. Roy. Astron. Soc.*, 468(2):1962–1980, 2017.
- [6] P. A. R. Ade et al. Planck 2015 results. XIII. Cosmological parameters. *Astron. Astrophys.*, 594:A13, 2016.
- [7] Dan Hooper. Particle Dark Matter. In *Theoretical Advanced Study Institute in Elementary Particle Physics: The Dawn of the LHC Era*, pages 709–764, 2010.
- [8] Martin Bauer and Tilman Plehn. *Yet Another Introduction to Dark Matter.*, volume 959 of *Lecture Notes in Physics*. Springer, 2019.
- [9] Stefano Profumo. Astrophysical Probes of Dark Matter. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Searching for New Physics at Small and Large Scales*, pages 143–189, 2013.

- [10] Stefano Profumo, Leonardo Giani, and Oliver F. Piattella. An introduction to particle dark matter. *Universe*, 5(10):213, 2019.
- [11] Leszek Roszkowski, Enrico Maria Sessolo, and Sebastian Trojanowski. WIMP dark matter candidates and searches—current status and future prospects. *Rept. Prog. Phys.*, 81(6):066201, 2018.
- [12] Marta Felcini. Searches for Dark Matter Particles at the LHC. In *53rd Rencontres de Moriond on Cosmology*, pages 327–336, 2018.
- [13] Burkhard Schmidt. The high-luminosity upgrade of the LHC: Physics and technology challenges for the accelerator and the experiments. 706:022002, apr 2016.
- [14] The hl-lhc project. <https://hilumilhc.web.cern.ch/content/hl-lhc-project>.
- [15] Michel Della Negra, Peter Jenni, and Tejinder S. Virdee. The construction of atlas and cms. *Annual Review of Nuclear and Particle Science*, 68(1):183–209, 2018.
- [16] G. Aad et al. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008.
- [17] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.
- [18] Albert M Sirunyan et al. Search for a heavy Higgs boson decaying to a pair of W bosons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 03:034, 2020.
- [19] Shankha Banerjee, Brian Batell, and Michael Spannowsky. Invisible decays in higgs boson pair production. *Phys. Rev. D*, 95:035009, Feb 2017.
- [20] Alexandre Alves, Tathagata Ghosh, and Farinaldo S. Queiroz. Dark and bright signatures of di-Higgs boson production. *Phys. Rev. D*, 100(3):036012, 2019.
- [21] Albert M. Sirunyan et al. Search for supersymmetric partners of electrons and muons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 790:140–166, 2019.
- [22] Georges Aad et al. Search for electroweak production of charginos and sleptons decaying into final states with two leptons and missing transverse momentum in $\sqrt{s} = 13$ TeV pp collisions using the ATLAS detector. *Eur. Phys. J. C*, 80(2):123, 2020.

- [23] Morad Aaboud et al. Search for a right-handed gauge boson decaying into a high-momentum heavy neutrino and a charged lepton in pp collisions with the ATLAS detector at $\sqrt{s} = 13$ TeV. *Phys. Lett. B*, 798:134942, 2019.
- [24] Vardan Khachatryan et al. Search for a Heavy Gauge Boson W' in the Final State with an Electron and Large Missing Transverse Energy in pp Collisions at $\sqrt{s} = 7$ TeV. *Phys. Lett. B*, 698:21–39, 2011.
- [25] Georges Aad et al. Combination of the W boson polarization measurements in top quark decays using ATLAS and CMS data at $\sqrt{s} = 8$ TeV. *JHEP*, 08(08):051, 2020.
- [26] Lars Sonnenschein. Analytical solution of $t\bar{t}$ dilepton equations. *Phys. Rev. D*, 73:054015, 2006. [Erratum: *Phys.Rev.D* 78, 079902 (2008)].
- [27] A. M. Sirunyan et al. Measurement of normalized differential $t\bar{t}$ cross sections in the dilepton channel from pp collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 04:060, 2018.
- [28] C. G. Lester and D. J. Summers. Measuring masses of semiinvisibly decaying particles pair produced at hadron colliders. *Phys. Lett. B*, 463:99–103, 1999.
- [29] Vernon Barger and Peisi Huang. Higgs boson finder and mass estimator: The Higgs boson to WW to leptons decay channel at the LHC. *Phys. Rev. D*, 84:093001, 2011.
- [30] Alan J. Barr and Christopher G. Lester. A Review of the Mass Measurement Techniques proposed for the Large Hadron Collider. *J. Phys. G*, 37:123001, 2010.
- [31] Partha Konar, Kyoungchul Kong, Konstantin T. Matchev, and Myeonghun Park. RECO level \sqrt{s}_{min} and subsystem \sqrt{s}_{min} : Improved global inclusive variables for measuring the new physics mass scale in E_T events at hadron colliders. *JHEP*, 06:041, 2011.
- [32] S. Kawabata, Y. Shimizu, Y. Sumino, and H. Yokoya. Measurement of physical parameters with a weight function method and its application to the Higgs boson mass reconstruction. *JHEP*, 08:129, 2013.
- [33] Won Sang Cho, Kiwoon Choi, Yeong Gyun Kim, and Chan Beom Park. M(T2)-assisted on-shell reconstruction of missing momenta and its application to spin measurement at the LHC. *Phys. Rev. D*, 79:031701, 2009.

- [34] Alan J. Barr, Ben Gripaios, and Christopher G. Lester. Weighing Wimps with Kinks at Colliders: Invisible Particle Mass Measurements from Endpoints. *JHEP*, 02:014, 2008.
- [35] Partha Konar and Abhaya Kumar Swain. Mass reconstruction with M_2 under constraint in semi-invisible production at a hadron collider. *Phys. Rev. D*, 93(1):015021, 2016.
- [36] Chan Beom Park. A singular way to search for heavy resonances in missing energy events. *JHEP*, 07(07):089, 2020.
- [37] Kiwoon Choi, Suyong Choi, Jae Sik Lee, and Chan Beom Park. Reconstructing the Higgs boson in dileptonic W decays at hadron collider. *Phys. Rev. D*, 80:073010, 2009.
- [38] Konstantin T. Matchev and Prasanth Shyamsundar. Singularity Variables for Missing Energy Event Kinematics. *JHEP*, 04:027, 2020.
- [39] Ji-Chong Yang, Jin-Hua Chen, and Yu-Chen Guo. Extract the energy scale of anomalous $\gamma\gamma \rightarrow W^+W^-$ scattering in the vector boson scattering process using artificial neural networks. *JHEP*, 21:085, 2020.
- [40] Jinmian Li, Cong Zhang, and Rao Zhang. Polarization measurement for the dileptonic channel of W^+W^- scattering using generative adversarial network. *Phys. Rev. D*, 105(1):016005, 2022.
- [41] Miguel Arratia, Daniel Britzger, Owen Long, and Benjamin Nachman. Reconstructing the kinematics of deep inelastic scattering with deep learning. *Nucl. Instrum. Meth. A*, 1025:166164, 2022.
- [42] David F. Rentería-Estrada, Roger J. Hernández-Pinto, German F. R. Sborlini, and Pia Zurita. Reconstructing partonic kinematics at colliders with Machine Learning. 12 2021.
- [43] Stephen B. Menary and Darren D. Price. Learning to discover: expressive Gaussian mixture models for multi-dimensional simulation and parameter inference in the physical sciences. *Mach. Learn. Sci. Tech.*, 3(1):015021, 2022.
- [44] Peter Flach. *Machine Learning: The Art and Science of Algorithms That Make Sense of Data*. Cambridge University Press, USA, 2012.

- [45] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer Series in Statistics. Springer New York Inc., New York, NY, USA, 2001.
- [46] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G.R. Day, Clint Richardson, Charles K. Fisher, and David J. Schwab. A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports*, 810:1–124, May 2019.
- [47] M. Grossi, J. Novak, B. Kersevan, and D. Rebuffi. Comparing traditional and deep-learning techniques of kinematic reconstruction for polarization discrimination in vector boson scattering. *Eur. Phys. J. C*, 80(12):1144, 2020.
- [48] Pinar Donmez, Guy Lebanon, and Krishnakumar Balasubramanian. Unsupervised supervised learning i: Estimating classification and regression errors without labels. *Journal of Machine Learning Research*, 11(44):1323–1351, 2010.
- [49] Raymond Davis, Don S. Harmer, and Kenneth C. Hoffman. Search for neutrinos from the sun. *Phys. Rev. Lett.*, 20:1205–1209, May 1968.
- [50] Q. R. Ahmad et al. Measurement of the rate of $\nu_e + d \rightarrow p + p + e^-$ interactions produced by ^8B solar neutrinos at the Sudbury Neutrino Observatory. *Phys. Rev. Lett.*, 87:071301, 2001.
- [51] Y. Fukuda et al. Evidence for oscillation of atmospheric neutrinos. *Phys. Rev. Lett.*, 81:1562–1567, 1998.
- [52] Nicola Cabibbo. Unitary symmetry and leptonic decays. *Phys. Rev. Lett.*, 10:531–533, Jun 1963.
- [53] M. Kobayashi and T. Maskawa. CP-Violation in the Renormalizable Theory of Weak Interaction. *Progress of Theoretical Physics*, 49(2):652–657, February 1973.
- [54] Ziro Maki, Masami Nakagawa, and Shoichi Sakata. Remarks on the Unified Model of Elementary Particles. *Progress of Theoretical Physics*, 28(5):870–880, 11 1962.
- [55] Marvin Gerlach, Florian Herren, and Matthias Steinhauser. Wilson coefficients for Higgs boson production and decoupling relations to $\mathcal{O}(\alpha_s^4)$. *JHEP*, 11:141, 2018.

- [56] Long-Bin Chen, Hai Tao Li, Hua-Sheng Shao, and Jian Wang. The gluon-fusion production of Higgs boson pair: N³LO QCD corrections and top-quark mass effects. *JHEP*, 03:072, 2020.
- [57] S. F. King. Neutrino mass models. *Rept. Prog. Phys.*, 67:107–158, 2004.
- [58] Stephen F. King. Models of Neutrino Mass, Mixing and CP Violation. *J. Phys. G*, 42:123001, 2015.
- [59] Stephen P. Martin. A Supersymmetry primer. *Adv. Ser. Direct. High Energy Phys.*, 18:1–98, 1998.
- [60] Michael E. Peskin. Supersymmetry in Elementary Particle Physics. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Exploring New Frontiers Using Colliders and Neutrinos*, pages 609–704, 1 2008.
- [61] Jihn E. Kim. Light pseudoscalars, particle physics and cosmology. *Physics Reports*, 150(1):1–177, 1987.
- [62] C. S. Frenk and Simon D. M. White. Dark matter and cosmic structure. *Annalen Phys.*, 524:507–534, 2012.
- [63] Kevork Abazajian. Linear cosmological structure limits on warm dark matter. *Phys. Rev. D*, 73:063513, 2006.
- [64] Roland de Putter, Olga Mena, Elena Giusarma, Shirley Ho, Antonio Cuesta, Hee-Jong Seo, Ashley J. Ross, Martin White, Dmitry Bizyaev, Howard Brewington, David Kirkby, Elena Malanushenko, Viktor Malanushenko, Daniel Oravetz, Kaike Pan, Will J. Percival, Nicholas P. Ross, Donald P. Schneider, Alaina Shelden, Audrey Simmons, and Stephanie Snedden. New Neutrino Mass Bounds from SDSS-III Data Release 8 Photometric Luminous Galaxies. , 761(1):12, December 2012.
- [65] Vladimir N. Lukash, E.V. Mikheeva, and A.M. Malinovsky. Formation of the large-scale structure of the universe. *Uspekhi Fizicheskikh Nauk*, 181(10):1017, 2011.
- [66] Andreas Birkedal-Hansen and Jay G. Wacker. Scalar dark matter from theory space. *Phys. Rev. D*, 69:065022, 2004.

- [67] Jay Hubisz and Patrick Meade. Phenomenology of the lightest Higgs with T-parity. *Phys. Rev. D*, 71:035016, 2005.
- [68] Hsin-Chia Cheng, Jonathan L. Feng, and Konstantin T. Matchev. Kaluza-Klein dark matter. *Phys. Rev. Lett.*, 89:211301, 2002.
- [69] Geraldine Servant and Timothy M. P. Tait. Is the lightest Kaluza-Klein particle a viable dark matter candidate? *Nucl. Phys. B*, 650:391–419, 2003.
- [70] Jonathan M. Cornell, Stefano Profumo, and William Shepherd. Dark matter in minimal universal extra dimensions with a stable vacuum and the “right” Higgs boson. *Phys. Rev. D*, 89(5):056005, 2014.
- [71] Peter W. Graham, Igor G. Irastorza, Steven K. Lamoreaux, Axel Lindner, and Karl A. van Bibber. Experimental Searches for the Axion and Axion-Like Particles. *Ann. Rev. Nucl. Part. Sci.*, 65:485–514, 2015.
- [72] Stephen F. King and Alexander Merle. Warm Dark Matter from keVins. *JCAP*, 08:016, 2012.
- [73] Miguel D. Campos and Werner Rodejohann. Testing keV sterile neutrino dark matter in future direct detection experiments. *Phys. Rev. D*, 94(9):095010, 2016.
- [74] A. Boyarsky, M. Drewes, T. Lasserre, S. Mertens, and O. Ruchayskiy. Sterile neutrino Dark Matter. *Prog. Part. Nucl. Phys.*, 104:1–45, 2019.
- [75] B. J. Carr. The primordial black hole mass spectrum. , 201:1–19, October 1975.
- [76] B. J. Carr, Kazunori Kohri, Yuuiti Sendouda, and Jun’ichi Yokoyama. New cosmological constraints on primordial black holes. *Phys. Rev. D*, 81:104019, 2010.
- [77] Bernard Carr, Florian Kuhnel, and Marit Sandstad. Primordial Black Holes as Dark Matter. *Phys. Rev. D*, 94(8):083504, 2016.
- [78] Latham A. Boyle, Robert R. Caldwell, and Marc Kamionkowski. Spintessence! New models for dark matter and dark energy. *Phys. Lett. B*, 545:17–22, 2002.
- [79] Neven Bilic, Gary B. Tupper, and Raoul D. Viollier. Unification of dark matter and dark energy: The Inhomogeneous Chaplygin gas. *Phys. Lett. B*, 535:17–21, 2002.

- [80] Alexandre Arbey. Dark fluid: A Complex scalar field to unify dark energy and dark matter. *Phys. Rev. D*, 74:043516, 2006.
- [81] Zong-Kuan Guo and Yuan-Zhong Zhang. Cosmology with a variable Chaplygin gas. *Phys. Lett. B*, 645:326–329, 2007.
- [82] Riccardo Catena and Piero Ullio. A novel determination of the local dark matter density. *JCAP*, 08:004, 2010.
- [83] Fabrizio Nesti and Paolo Salucci. The Dark Matter halo of the Milky Way, AD 2013. *JCAP*, 07:016, 2013.
- [84] S. Sivertsson, H. Silverwood, J. I. Read, G. Bertone, and P. Steger. The local dark matter density from SDSS-SEGUE G-dwarfs. *Mon. Not. Roy. Astron. Soc.*, 478(2):1677–1693, 2018.
- [85] E. Aprile et al. Light Dark Matter Search with Ionization Signals in XENON1T. *Phys. Rev. Lett.*, 123(25):251801, 2019.
- [86] E. Aprile et al. Search for Light Dark Matter Interactions Enhanced by the Migdal Effect or Bremsstrahlung in XENON1T. *Phys. Rev. Lett.*, 123(24):241803, 2019.
- [87] Xiangyi Cui et al. Dark Matter Results From 54-Ton-Day Exposure of PandaX-II Experiment. *Phys. Rev. Lett.*, 119(18):181302, 2017.
- [88] R. Agnese et al. Search for Low-Mass Dark Matter with CDMSlite Using a Profile Likelihood Fit. *Phys. Rev. D*, 99(6):062001, 2019.
- [89] F. Ruppin, J. Billard, E. Figueroa-Feliciano, and L. Strigari. Complementarity of dark matter detectors in light of the neutrino background. *Phys. Rev. D*, 90(8):083510, 2014.
- [90] A. Arbey and F. Mahmoudi. Dark matter and the early Universe: a review. *Prog. Part. Nucl. Phys.*, 119:103865, 2021.
- [91] A. Arbey, M. Boudaud, F. Mahmoudi, and G. Robbins. Robustness of dark matter constraints and interplay with collider searches for New Physics. *JHEP*, 11:132, 2017.
- [92] Julio F. Navarro, Carlos S. Frenk, and Simon D. M. White. The Structure of cold dark matter halos. *Astrophys. J.*, 462:563–575, 1996.

- [93] Paul J. McMillan. Mass models of the Milky Way. , 414(3):2446–2457, July 2011.
- [94] A. Burkert. The Structure of dark matter halos in dwarf galaxies. *Astrophys. J. Lett.*, 447:L25, 1995.
- [95] Alister W. Graham, David Merritt, Ben Moore, Juerg Diemand, and Balsa Terzic. Empirical models for Dark Matter Halos. I. Nonparametric Construction of Density Profiles and Comparison with Parametric Models. *Astron. J.*, 132:2685–2700, 2006.
- [96] Julio F. Navarro, Aaron Ludlow, Volker Springel, Jie Wang, Mark Vogelsberger, Simon D. M. White, Adrian Jenkins, Carlos S. Frenk, and Amina Helmi. The Diversity and Similarity of Cold Dark Matter Halos. *Mon. Not. Roy. Astron. Soc.*, 402:21, 2010.
- [97] H. Abdallah et al. Search for γ -Ray Line Signals from Dark Matter Annihilations in the Inner Galactic Halo from 10 Years of Observations with H.E.S.S. *Phys. Rev. Lett.*, 120(20):201101, 2018.
- [98] Benjamin Zitzer. The VERITAS Dark Matter Program. *PoS, ICRC2017*:904, 2018.
- [99] V. Barger, Wai Yee Keung, D. Marfatia, and G. Shaughnessy. PAMELA and dark matter. *Phys. Lett. B*, 672:141–146, 2009.
- [100] O. Adriani et al. PAMELA Measurements of Cosmic-ray Proton and Helium Spectra. *Science*, 332:69–72, 2011.
- [101] M. Aguilar, L. Ali Cavasonza, B. Alpat, G. Ambrosi, L. Arruda, N. Attig, S. Aupetit, P. Azzarello, A. Bachlechner, F. Barao, A. Barrau, L. Barrin, A. Bartoloni, L. Basara, S. Başğmez-du Pree, M. Battarbee, R. Battiston, J. Bazo, U. Becker, M. Behlmann, B. Beischer, J. Berdugo, B. Bertucci, V. Bindi, G. Boella, W. de Boer, K. Bollweg, V. Bonnivard, B. Borgia, M. J. Boschini, M. Bourquin, E. F. Bueno, J. Burger, F. Cadoux, X. D. Cai, M. Capell, S. Caroff, J. Casaus, G. Castellini, I. Cernuda, F. Cervelli, M. J. Chae, Y. H. Chang, A. I. Chen, G. M. Chen, H. S. Chen, L. Cheng, H. Y. Chou, E. Choumilov, V. Choutko, C. H. Chung, C. Clark, R. Clavero, G. Coignet, C. Consonlandi, A. Contin, C. Corti, B. Coste, W. Creus, M. Crispoltoni, Z. Cui, Y. M. Dai, C. Delgado, S. Della Torre, M. B. Demirköz, L. Derome, S. Di Falco, F. Dimiccoli, C. Díaz, P. von Doetinchem, F. Dong, F. Donnini, M. Duranti, D. D’Urso, A. Egorov, A. Eline, T. Eronen, J. Feng, E. Fiandrini, E. Finch, P. Fisher, V. Formato, Y. Galaktionov,

G. Gallucci, B. García, R. J. García-López, C. Gargiulo, H. Gast, I. Gebauer, M. Gervasi, A. Ghelfi, F. Giovacchini, P. Goglov, D. M. Gómez-Coral, J. Gong, C. Goy, V. Grabski, D. Grandi, M. Graziani, I. Guerri, K. H. Guo, M. Habiby, S. Haino, K. C. Han, Z. H. He, M. Heil, J. Hoffman, T. H. Hsieh, H. Huang, Z. C. Huang, C. Huh, M. Incagli, M. Ionica, W. Y. Jang, H. Jinchi, S. C. Kang, K. Kanishev, G. N. Kim, K. S. Kim, Th. Kirn, C. Konak, O. Kounina, A. Kounine, V. Koutsenko, M. S. Krafczyk, G. La Vacca, E. Laudi, G. Laurenti, I. Lazzizzera, A. Lebedev, H. T. Lee, S. C. Lee, C. Leluc, H. S. Li, J. Q. Li, J. Q. Li, Q. Li, T. X. Li, W. Li, Z. H. Li, Z. Y. Li, S. Lim, C. H. Lin, P. Lipari, T. Lippert, D. Liu, Hu Liu, S. Q. Lu, Y. S. Lu, K. Luebelsmeyer, F. Luo, J. Z. Luo, S. S. Lv, R. Majka, C. Mañá, J. Marín, T. Martin, G. Martínez, N. Masi, D. Maurin, A. Menchaca-Rocha, Q. Meng, D. C. Mo, L. Morescalchi, P. Mott, T. Nelson, J. Q. Ni, N. Nikonov, F. Nozzoli, P. Nunes, A. Oliva, M. Orcinha, F. Palmonari, C. Palomares, M. Paniccia, M. Pauluzzi, S. Pensotti, R. Pereira, N. Picot-Clemente, F. Pilo, C. Pizzolotto, V. Plyaskin, M. Pohl, V. Poireau, A. Putze, L. Quadrani, X. M. Qi, X. Qin, Z. Y. Qu, T. Rähkä, P. G. Rancoita, D. Rapin, J. S. Ricol, I. Rodríguez, S. Rosier-Lees, A. Rozhkov, D. Rozza, R. Sagdeev, J. Sandweiss, P. Saouter, S. Schael, S. M. Schmidt, A. Schulz von Dratzig, G. Schwing, E. S. Seo, B. S. Shan, J. Y. Shi, T. Siedenbueg, D. Son, J. W. Song, W. H. Sun, M. Tacconi, X. W. Tang, Z. C. Tang, L. Tao, D. Tesaro, Samuel C. C. Ting, S. M. Ting, N. Tomassetti, J. Torsti, C. Türkoğlu, T. Urban, V. Vagelli, E. Valente, C. Vannini, E. Valtonen, M. Vázquez Acosta, M. Vecchi, M. Velasco, J. P. Vialle, V. Vitale, S. Vitillo, L. Q. Wang, N. H. Wang, Q. L. Wang, X. Wang, X. Q. Wang, Z. X. Wang, C. C. Wei, Z. L. Weng, K. Whitman, J. Wienkenhöver, M. Willenbrock, H. Wu, X. Wu, X. Xia, R. Q. Xiong, W. Xu, Q. Yan, J. Yang, M. Yang, Y. Yang, H. Yi, Y. J. Yu, Z. Q. Yu, S. Zeissler, C. Zhang, J. Zhang, J. H. Zhang, S. D. Zhang, S. W. Zhang, Z. Zhang, Z. M. Zheng, Z. Q. Zhu, H. L. Zhuang, V. Zhukov, A. Zichichi, N. Zimmermann, and P. Zuccon. Antiproton flux, antiproton-to-proton flux ratio, and properties of elementary particle fluxes in primary cosmic rays measured with the alpha magnetic spectrometer on the international space station. *Phys. Rev. Lett.*, 117:091103, Aug 2016.

[102] M. Aguilar et al. Towards Understanding the Origin of Cosmic-Ray Positrons. *Phys. Rev. Lett.*, 122(4):041102, 2019.

[103] Christoph Tönnis. Search for secluded dark matter with 6 years of IceCube data. In *37th International Cosmic Ray Conference*, 7 2021.

- [104] Rasha Abbasi et al. Searching for Dark Matter from the Sun with the IceCube Detector. *PoS*, ICRC2021:020, 2022.
- [105] D. Maurin, F. Donato, R. Taillet, and P. Salati. Cosmic rays below $z=30$ in a diffusion model: new constraints on propagation parameters. *Astrophys. J.*, 555:585–596, 2001.
- [106] F. Donato, Nicolao Fornengo, D. Maurin, and P. Salati. Antiprotons in cosmic rays from neutralino annihilation. *Phys. Rev. D*, 69:063501, 2004.
- [107] M. Boudaud, E. F. Bueno, S. Caroff, Y. Genolini, V. Poulin, V. Poireau, A. Putze, S. Rosier, P. Salati, and M. Vecchi. The pinching method for Galactic cosmic ray positrons: implications in the light of precision measurements. *Astron. Astrophys.*, 605:A17, 2017.
- [108] Yang Bai, James Bourbeau, and Tongyan Lin. Dark matter searches with a mono- Z' jet. *JHEP*, 06:205, 2015.
- [109] Christoph Englert, Matthew McCullough, and Michael Spannowsky. S-channel dark matter simplified models and unitarity, 2016.
- [110] Seng Pei Liew, Michele Papucci, Alessandro Vichi, and Kathryn M. Zurek. Mono-X Versus Direct Searches: Simplified Models for Dark Matter at the LHC. *JHEP*, 06:082, 2017.
- [111] Glen Cowan, Kyle Cranmer, Eilam Gross, and Ofer Vitells. Asymptotic formulae for likelihood-based tests of new physics. *Eur. Phys. J. C*, 71:1554, 2011. [Erratum: *Eur.Phys.J.C* 73, 2501 (2013)].
- [112] Alexandre Alves and Kuver Sinha. Searches for Dark Matter at the LHC: A Multivariate Analysis in the Mono- Z Channel. *Phys. Rev. D*, 92(11):115013, 2015.
- [113] Identification of b quark jets at the CMS Experiment in the LHC Run 2. 3 2016.
- [114] Optimisation and performance studies of the ATLAS b -tagging algorithms for the 2017-18 LHC run. 7 2017.
- [115] Steven Schramm. Machine learning at CERN: ATLAS, LHCb, and more. *PoS*, ICHEP2018:158, 2019.

- [116] Rüdiger Haake. Machine and deep learning techniques in heavy-ion collisions with ALICE. *PoS, EPS-HEP2017*, 2017.
- [117] Nicholas Choma, Federico Monti, Lisa Gerhardt, Tomasz Palczewski, Zahra Ronaghi, Prabhat, Wahid Bhimji, Michael M. Bronstein, Spencer R. Klein, and Joan Bruna. Graph neural networks for icecube signal classification, 2018.
- [118] Vernon Barger, Paul Langacker, Mathew McCaskey, Michael J. Ramsey-Musolf, and Gabe Shaughnessy. LHC Phenomenology of an Extended Standard Model with a Real Scalar Singlet. *Phys. Rev. D*, 77:035005, 2008.
- [119] Stefano Profumo, Michael J. Ramsey-Musolf, Carroll L. Wainwright, and Peter Winslow. Singlet-catalyzed electroweak phase transitions and precision Higgs boson studies. *Phys. Rev. D*, 91(3):035018, 2015.
- [120] T. Huang, J.M. No, L. Pernié, M. Ramsey-Musolf, A. Safonov, M. Spannowsky, and P. Winslow. Resonant di-Higgs boson production in the $b\bar{b}W\bar{W}$ channel: Probing the electroweak phase transition at the LHC. *Phys. Rev. D*, 96(3):035007, 2017.
- [121] Stefano Profumo, Michael J. Ramsey-Musolf, and Gabe Shaughnessy. Singlet Higgs phenomenology and the electroweak phase transition. *JHEP*, 08:010, 2007.
- [122] Gerard Jungman, Marc Kamionkowski, and Kim Griest. Supersymmetric dark matter. *Phys. Rept.*, 267:195–373, 1996.
- [123] Albert M Sirunyan et al. Search for supersymmetry in final states with two oppositely charged same-flavor leptons and missing transverse momentum in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JHEP*, 04:123, 2021.
- [124] Charu C. Aggarwal. *Neural Networks and Deep Learning A Textbook*. Springer, 2018.
- [125] François Chollet. *Deep Learning with Python*. Manning Publications, 2018.
- [126] Rakshit Jain, Asmita Chotani, and G. Anuradha. 9 - disease diagnosis using machine learning: A comparative study. In Kun Chang Lee, Sanjiban Sekhar Roy, Pijush Samui, and Vijay Kumar, editors, *Data Analytics in Biomedical Engineering and Healthcare*, pages 145–161. Academic Press, 2021.
- [127] Miroslav Kubat. *An Introduction to Machine Learning*. Springer, 2017.

- [128] S. Lloyd. Least squares quantization in pcm. *IEEE Transactions on Information Theory*, 28(2):129–137, 1982.
- [129] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, 1996.
- [130] David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. A learning algorithm for boltzmann machines. *Cognitive Science*, 9(1):147–169, 1985.
- [131] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.
- [132] Diederik P Kingma and Max Welling. Auto-Encoding Variational Bayes. *arXiv e-prints*, page arXiv:1312.6114, December 2013.
- [133] Slawomir Wierzchoń and Mieczysław Kłopotek. *Modern Algorithms of Cluster Analysis*. Springer International Publishing, 2018.
- [134] Fionn Murtagh Christian Hennig, Marina Meila and Roberto Rocci. *Handbook of cluster analysis*. CRC Press, 2016.
- [135] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [136] V. V. Gligorov and Mike Williams. Efficient, reliable and fast high-level triggering using a bonsai boosted decision tree. *JINST*, 8:P02013, 2013.
- [137] Morad Aaboud et al. Measurements of b-jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13$ TeV. *JHEP*, 08:089, 2018.
- [138] Albert M Sirunyan et al. Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13$ TeV. *JINST*, 15(10):P10017, 2020.
- [139] Leo Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001.
- [140] Marc Peter Deisenroth, A. Aldo Faisal, and Cheng Soon Ong. *Mathematics for Machine Learning*. Cambridge University Press, 2020.

- [141] Jacob Goldberger, Geoffrey E Hinton, Sam Roweis, and Russ R Salakhutdinov. Neighbourhood components analysis. In L. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 17, pages 513–520. MIT Press, 2005.
- [142] Tom Fawcett. An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874, 2006. ROC Analysis in Pattern Recognition.
- [143] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1):79 – 86, 1951.
- [144] et al Charles R. Harris. Array programming with NumPy. *Nature*, 585(7825):357–362, September 2020.
- [145] et al. Virtanen, Pauli. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.
- [146] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007.
- [147] Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010.
- [148] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [149] et al. Martín Abadi. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.
- [150] Francois Chollet et al. Keras, 2015.
- [151] Eric Postma Laurens van der Maaten and Jaap van den Herik. Dimensionality reduction: A comparative review, 2009.
- [152] Ian T. Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philos Trans A Math Phys Eng Sci*, 374(2065), 2015.

- [153] Laurens van der Maaten. Accelerating t-sne using tree-based algorithms. *Journal of Machine Learning Research*, 15(93):3221–3245, 2014.
- [154] Leland McInnes, John Healy, Nathaniel Saul, and Lukas Großberger. Umap: Uniform manifold approximation and projection. *Journal of Open Source Software*, 3(29):861, 2018.
- [155] The brain: Understanding neurobiology through the study of addiction. <https://archives.drugabuse.gov/publications/brain-understanding-neurobiology-through-study-addiction>, March 2010.
- [156] Charu C. Aggarwal. *Neural Networks and Deep Learning*. Springer Cham, 1 edition, 2018.
- [157] Nikhil Ketkar. *Deep learning with python: A hands-on introduction*. Apress, 2017.
- [158] Kurt Hornik. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, 4(2):251–257, 1991.
- [159] Dmitry Yarotsky. Universal approximations of invariant maps by neural networks. *Constructive Approximation*, 55:407–474, 2018.
- [160] Zhou Lu, Hongming Pu, Feicheng Wang, Zhiqiang Hu, and Liwei Wang. The Expressive Power of Neural Networks: A View from the Width. *arXiv e-prints*, page arXiv:1709.02540, September 2017.
- [161] Boris Hanin and Mark Sellke. Approximating Continuous Functions by ReLU Nets of Minimal Width. *arXiv e-prints*, page arXiv:1710.11278, October 2017.
- [162] Hong Hui Tan and King Hann Lim. Review of second-order optimization techniques in artificial neural networks backpropagation. *IOP Conference Series: Materials Science and Engineering*, 495:012003, jun 2019.
- [163] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.
- [164] Pranjal Awasthi, Abhimanyu Das, Rajat Sen, and Ananda Theertha Suresh. On the benefits of maximum likelihood estimation for regression and forecasting, 2021.

- [165] Sebastian Ruder. An overview of gradient descent optimization algorithms, 2016.
- [166] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12:2121–2159, 07 2011.
- [167] Geoffrey Hinton. Lecture notes 6a in introduction to neural networks and machine learning, February 2014.
- [168] Geoffrey E. Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *ArXiv*, abs/1207.0580, 2012.
- [169] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and Joao P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:1–102, 2012.
- [170] James M. Cline, Kimmo Kainulainen, and Axel P. Vischer. Dynamics of two-higgs-doublet cp violation and baryogenesis at the electroweak phase transition. *Physical Review D*, 54(4):2451–2472, aug 1996.
- [171] Lars Fromme, Stephan J Huber, and Michael Seniuch. Baryogenesis in the two-higgs doublet model. *Journal of High Energy Physics*, 2006(11):038–038, nov 2006.
- [172] Mark Trodden. Electroweak baryogenesis: A brief review, 1998.
- [173] Frank F. Deppisch, Chandan Hati, Sudhanwa Patra, Utpal Sarkar, and Josè W. F. Valle. 331 Models and Grand Unification: From Minimal SU(5) to Minimal SU(6). *Phys. Lett. B*, 762:432–440, 2016.
- [174] Qing-Hong Cao and Dong-Ming Zhang. Collider Phenomenology of the 3-3-1 Model. 11 2016.
- [175] S. Chatrchyan, V. Khachatryan, A. M. Sirunyan, A. Tumasyan, W. Adam, T. Bergauer, M. Dragicevic, J. Erö, C. Fabjan, and et al. Measurement of higgs boson production and properties in the ww decay channel with leptonic final states. *Journal of High Energy Physics*, 2014(1), Jan 2014.

- [176] R. Frederix, S. Frixione, V. Hirschi, D. Pagani, H.-S. Shao, and M. Zaro. The automation of next-to-leading order electroweak calculations. *Journal of High Energy Physics*, 2018(7), Jul 2018.
- [177] Torbjörn Sjöstrand, Stefan Ask, Jesper R. Christiansen, Richard Corke, Nishita Desai, Philip Ilten, Stephen Mrenna, Stefan Prestel, Christine O. Rasmussen, and Peter Z. Skands. An introduction to pythia 8.2. *Computer Physics Communications*, 191:159–177, Jun 2015.
- [178] Tao Han. Collider phenomenology: Basic knowledge and techniques. In *Theoretical Advanced Study Institute in Elementary Particle Physics: Physics in $D \geq 4$* , pages 407–454, 8 2005.
- [179] J. de Favereau, C. Delaere, P. Demin, A. Giammanco, V. Lemaître, A. Mertens, and M. Selvaggi. Delphes 3: a modular framework for fast simulation of a generic collider experiment. *Journal of High Energy Physics*, 2014(2), Feb 2014.
- [180] Marco Bellagente, Anja Butter, Gregor Kasieczka, Tilman Plehn, and Ramon Winterhalder. How to get away detector effects. *SciPost Physics*, 8(4), Apr 2020.
- [181] Marco Bellagente, Anja Butter, Gregor Kasieczka, Tilman Plehn, Armand Rousselot, Ramon Winterhalder, Lynton Ardizzone, and Ullrich Köthe. Invertible networks or partons to detector and back again. *SciPost Physics*, 9(5), Nov 2020.
- [182] Luca Scodellaro. b tagging in ATLAS and CMS. In *5th Large Hadron Collider Physics Conference*, 9 2017.
- [183] A. J. Barr. Measuring slepton spin at the LHC. *JHEP*, 02:042, 2006.
- [184] Jeong Han Kim, Kyoungchul Kong, Konstantin T. Matchev, and Myeonghun Park. Probing the Triple Higgs Self-Interaction at the Large Hadron Collider. *Phys. Rev. Lett.*, 122(9):091801, 2019.
- [185] Jerome Friedman. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29, 11 2000.
- [186] J. Bergstra, D. Yamins, and D. D. Cox. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *Proceedings*

of the 30th International Conference on International Conference on Machine Learning
- Volume 28, ICML'13, page I-115-I-123. JMLR.org, 2013.

- [187] James Bergstra. Hyperopt: Distributed asynchronous hyper-parameter optimization.
- [188] James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyper-parameter optimization. NIPS'11, page 2546–2554, Red Hook, NY, USA, 2011. Curran Associates Inc.
- [189] Alexandre Alves. Stacking machine learning classifiers to identify Higgs bosons at the LHC. *JINST*, 12(05):T05005, 2017.
- [190] Pierre Baldi, Peter Sadowski, and Daniel Whiteson. Searching for Exotic Particles in High-Energy Physics with Deep Learning. *Nature Commun.*, 5:4308, 2014.
- [191] Ye Ren, Le Zhang, and P.N. Suganthan. Ensemble classification and regression-recent developments, applications and future directions [review article]. *IEEE Computational Intelligence Magazine*, 11(1):41–53, 2016.
- [192] Nathalie Japkowicz and Mohak Shah. *Evaluating Learning Algorithms: A Classification Perspective*. Cambridge University Press, 2011.
- [193] David Hand and Robert Till. A simple generalisation of the area under the roc curve for multiple class classification problems. *Hand, The*, 45:171–186, 11 2001.
- [194] Ti-Pei Li and YuQian Ma. Analysis methods for results in gamma-ray astronomy. *The Astrophysical Journal*, 272:317–324, 08 1983.
- [195] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017.
- [196] Serguei Chatrchyan et al. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Phys. Lett. B*, 716:30–61, 2012.
- [197] Georges Aad et al. Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29, 2012.

- [198] Georges Aad et al. Combined Measurement of the Higgs Boson Mass in pp Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments. *Phys. Rev. Lett.*, 114:191803, 2015.
- [199] Georges Aad et al. Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV. *JHEP*, 08:045, 2016.
- [200] Masaaki Kuroda. Complete Lagrangian of MSSM. 2 1999.
- [201] Peter Z. Skands et al. SUSY Les Houches accord: Interfacing SUSY spectrum calculators, decay packages, and event generators. *JHEP*, 07:036, 2004.
- [202] B. C. Allanach et al. SUSY Les Houches Accord 2. *Comput. Phys. Commun.*, 180:8–25, 2009.
- [203] Qingyu Zhao, Ehsan Adeli, Nicolas Honnorat, Tuo Leng, and Kilian M Pohl. Variational autoencoder for regression: Application to brain aging analysis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 823–831. Springer, 2019.