

**Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”**

**Redes neurais artificiais auto-organizáveis na classificação
não-supervisionada de imagens multiespectrais de sensoriamento remoto**

Christopher Silva de Pádua

Dissertação apresentada para obtenção do título de
Mestre em Ciências. Área de concentração: Estatística
e Experimentação Agronômica

**Piracicaba
2016**

Christopher Silva de Pádua
Bacharel em Matemática Computacional

**Redes neurais artificiais auto-organizáveis na classificação não-supervisionada de
imagens multiespectrais de sensoriamento remoto**

versão revisada de acordo com a resolução CoPGr 6018 de 2011

Orientador:
Prof. Dr. **VITOR AUGUSTO OZAKI**

Dissertação apresentada para obtenção do título de Mestre
em Ciências. Área de concentração: Estatística e Experi-
mentação Agronômica

Piracicaba
2016

**Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA - DIBD/ESALQ/USP**

Pádua, Christopher Silva de

Redes neurais artificiais auto-organizáveis na classificação não-supervisionada de imagens multiespectrais de sensoriamento remoto / Christopher Silva de Pádua. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - - Piracicaba, 2016. 84 p. : il.

Dissertação (Mestrado) - - Escola Superior de Agricultura "Luiz de Queiroz".

1. Redes neurais artificiais 2. Growing neural gas 3. Sensoriamento remoto 4. Máxima verossimilhança 5. Imagem multiespectral de satélite I. Título

CDD 526.9823
P125r

"Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor"

AGRADECIMENTOS

À minha família, em especial minha mãe Jane Sheila da Silva Batista, pois foram os principais responsáveis pela minha formação me ajudando a trilhar esse caminho e chegar até esse ponto.

Ao Prof. Dr. Vitor Augusto Ozaki pela orientação e apoio fundamental para a realização desse projeto.

À Prof^a. Dr^a. Juliana Garcia Cespedes pelo continuo apoio e ajuda científica.

Aos professores do departamento de Estatística e Experimentação Agronômica da ESALQ/USP pela contribuição e incentivo à minha formação acadêmica.

Aos amigos de mestrado, sobretudo; Viviane Paulenas, Vinícius Menarin e Maíra Fatoretto que compartilharam dessa jornada e a tornaram em algo muito agradável.

À CAPES pelo apoio financeiro.

*“Tudo aquilo que o homem ignora não existe para ele.
Por isso o universo de cada um se resume ao tamanho do seu saber.”
Albert Einstein*

SUMÁRIO

RESUMO	9
ABSTRACT	11
LISTA DE ILUSTRAÇÕES	13
LISTA DE TABELAS	15
1 INTRODUÇÃO	17
2 REVISÃO BIBLIOGRÁFICA	19
2.1 Sensoriamento remoto	19
2.1.1 Aquisição e características de imagens de satélite	19
2.1.2 Landsat 8	22
2.1.2.1 Características das bandas do satélite Landsat 8	24
2.1.3 Espaço de atributos e métodos de classificação	25
2.1.3.1 Métodos supervisionados	25
2.1.3.2 Métodos não supervisionados	27
2.1.3.3 Métodos híbridos	27
2.2 Redes Neurais Artificiais	27
2.2.1 Neurônio artificial	28
2.2.2 Perceptron	31
2.2.3 Multilayer perceptron	32
2.2.3.1 Processo de Aprendizagem	34
2.2.4 Growing Neural Gas	36
2.2.4.1 Algoritmo de aprendizagem	38
2.3 Máxima Verossimilhança	40
2.4 K-Means	42
2.5 Análise de Componentes Principais	43
3 MATERIAL E MÉTODOS	49
4 RESULTADOS E DISCUSSÕES	55
4.1 Biblioteca Espectral	55
4.2 Imagem de Satélite	66
5 CONSIDERAÇÕES FINAIS	75
REFERÊNCIAS	77
ANEXOS	81

RESUMO

Redes neurais artificiais auto-organizáveis na classificação não-supervisionada de imagens multiespectrais de sensoriamento remoto

O uso de imagens provenientes de sensores remotos, tal como sistemas acoplados em aviões e satélites, é cada vez mais frequente, uma vez que permite o monitoramento contínuo e periódico ao longo do tempo por meio de diversas observações de uma mesma região, por vezes ampla ou de difícil acesso. Essa ferramenta tem se mostrado importante e significativa em aplicações como o mapeamento de solo e fronteiras; acompanhamento de áreas de desmatamento, queimadas e de produção agrícola. Para gerar resultados interpretáveis ao usuário final, essas imagens devem ser processadas. Atualmente, o método de classificação por máxima verossimilhança é o mais empregado para classificação de imagens multiespectrais de sensores remotos, entretanto, por se tratar de uma técnica supervisionada, seus resultados dependem extensivamente da qualidade do conjunto de treinamento, utilizado para definir os parâmetros do método. A seleção de um bom conjunto de treinamento é um processo custoso e inviabiliza a automação da classificação para diversas imagens. O método de classificação por máxima verossimilhança é também paramétrico e portanto existem algumas suposições quanto a distribuição dos dados que devem ser atendidas, caso contrário a aplicação do método pode gerar resultados ruins. Tendo em vista essas desvantagens do método da máxima verossimilhança, este trabalho propõe um novo método para a classificação de imagens multiespectrais provenientes de sensores remotos de forma que o procedimento seja autônomo, veloz e preciso, minimizando dessa forma os possíveis erros humanos inseridos em etapas intermediárias do processo, tal como a definição de conjuntos de treinamento. O método aqui proposto pertence ao conjunto das redes neurais artificiais (RNAs) e é denominado **growing neural gas** (GNG). Este método baseia-se no aprendizado não supervisionado de padrões “naturais” dentro de um conjunto de dados por meio da criação e adaptação de uma rede mínima de neurônios. Os resultados gerados a partir da classificação pela RNA foram comparados com os métodos mais utilizados na literatura atual, sendo eles o método da máxima verossimilhança e o método **k-means**. A partir da biblioteca espectral ASTER, mantida e criada parcialmente pela NASA, foram realizadas várias repetições do experimento, que consiste em classificar os dados de acordo com as diferentes classes existentes, e para cada uma destas repetições calculou-se uma medida de acurácia, denominada índice kappa, além do tempo de execução de cada método, de forma que suas médias foram comparadas via intervalo de confiança gerados por **bootstrap** não paramétrico. Também investigou-se como a análise de componentes principais (ACP), técnica utilizada para reduzir a dimensão dos dados e conseqüentemente o custo computacional, pode influenciar no desempenho dos métodos, tanto em sua qualidade de classificação quanto em relação ao tempo de execução. Os resultados mostram que o método proposto é superior nos dois aspectos estudados, acurácia e tempo de execução, para a maioria dos fatores aplicados. Mostra-se ainda um exemplo de aplicação prática em que uma imagem multiespectral de satélite não satisfaz as pré-suposições estabelecidas para o uso do método da máxima verossimilhança e verifica-se a diferença entre os métodos com relação a qualidade final da imagem classificada.

Palavras-chave: Redes neurais artificiais; **Growing neural gas**; Sensoriamento remoto; Máxima verossimilhança; Imagem multiespectral de satélite

ABSTRACT

Self-organizing artificial neural networks in the unsupervised classification of multispectral remote sensing imagery

The use of images from remote sensors, such as coupled systems in airplanes and satellites, are increasingly being used because they allow continuous and periodic surveillance over time through several observations of some particular area, sometimes large or difficult to access. This sort of image has shown an important and meaningful participation in applications such as soil and borders mapping; surveillance of deforestation, forest fires and agricultural production areas. To generate interpretable results to the end user, these images must be processed. Currently, the maximum likelihood classification method is the most used for multispectral image classification of remote sensing, however, because it is a supervised technique, the results depend extensively on the quality of the training set, used to define the parameters of the method. Selecting a good training set is a costly process and prevents the automation of classification for different images. The maximum likelihood classification method is also parametric, and therefore, some assumptions about the distribution of the data must be met, otherwise the application of the method can generate bad results. In view of these disadvantages of the maximum likelihood method, this dissertation proposes a new, autonomous, fast and accurate method for multispectral remote sensing imagery classification thereby minimizing the possible human errors inserted at intermediate stages of the process, such as the definition of training sets. The method proposed here belongs to the set of artificial neural networks (ANN) and is called growing neural gas (GNG). This method is based on unsupervised learning of "natural" patterns in a dataset through the creation and adaptation of a minimum network of neurons. The results generated from the classification by ANN were compared with the most commonly used methods in the literature: the maximum likelihood method and the k-means method. From the spectral library Aster, maintained and made in part by NASA, several replications of the experiment were made, which is to classify the data according to different preestablished classes, and a measure of accuracy called kappa index was calculated for each of the replicates, in addition to the execution time of each method, so that their means were compared via confidence interval generated by nonparametric bootstrap. It was additionally investigated how principal component analysis (PCA), technique which reduces dimension of data and consequently the computational cost, can influence the performance of methods, both in its quality rating and runtime. The results show that the proposed method is superior in both aspects studied, accuracy and runtime, for the majority of applied factors. Furthermore, it is shown an example of a practical application in which a multispectral satellite image does not necessarily meet the established assumptions for using the maximum likelihood method, and there is a difference between the methods, regarding to its final classified image quality.

Keywords: Artificial neural networks; Growing neural gas; Remote sensing; Maximum likelihood; Multispectral satellite image

LISTA DE ILUSTRAÇÕES

Figura 2.1 – Relação entre fonte de radiação (sol), objeto (árvore) e sensor (satélite)	20
Figura 2.2 – Representação de uma imagem multiespectral	20
Figura 2.3 – Curvas espectrais para vegetação seca e saudável	21
Figura 2.4 – Comparação de imagens a diferentes resolução radiométricas	22
Figura 2.5 – Comparação de resoluções entre Landsat 7 e Landsat 8.	23
Figura 2.6 – Dimensão de um pixel a diferentes resoluções	23
Figura 2.7 – Representação de classes no espaço de atributos	25
Figura 2.8 – Exemplo de redes neurais artificiais	28
Figura 2.9 – Neurônio Artificial	29
Figura 2.10–Função de transferência linear	30
Figura 2.11–Função de transferência rampa	30
Figura 2.12–Função de transferência degrau	31
Figura 2.13–Função de transferência sigmoideal	31
Figura 2.14–Classificação de espaço linearmente separável via perceptron	32
Figura 2.15–Exemplo de rede MLP com duas camadas intermediárias	32
Figura 2.16–Classificação de espaço não linearmente separável via MLP	33
Figura 2.17–Fluxo de propagação de erros	34
Figura 2.18–Exemplo de caminho percorrido na superfície de erro pelo algoritmo back-propagation	36
Figura 2.19–Forma de definir similaridade entre um conjunto de pontos	37
Figura 2.20–Exemplo de adaptação e crescimento da rede GNG aplicada a um conjunto de dados genérico	40
Figura 2.21–Exemplo de superfície de distribuição normal bivariada para classes distintas no espaço formado por duas bandas espectrais	41
Figura 2.22–Exemplo de aplicação ACP	44
Figura 3.1 – Visualização dos dados utilizados para comparação dos métodos	50
Figura 3.2 – Diagrama representando o delineamento do experimento	51
Figura 4.1 – Gráfico de dispersão dos pontos representados no espaço bidimensional formado pelos pares de bandas	57
Figura 4.2 – Densidade dos pontos para todas as bandas e classes	58
Figura 4.3 – Boxplot dos pontos para todas as bandas e classes	59
Figura 4.4 – Biplot dos pontos transformados pela análise de componentes principais	60
Figura 4.5 – Densidade dos pontos para os componentes principais e classes	61
Figura 4.6 – Boxplot dos pontos para os componentes principais e classes	62
Figura 4.7 – Índice kappa obtido pela análise	63
Figura 4.8 – Média do índice kappa com intervalo de confiança para os tratamentos	63

Figura 4.9 – Tempo de execução dos métodos	64
Figura 4.10–Média do tempo de execução com intervalo de confiança para os tratamentos	65
Figura 4.11–Separação por método para melhor visualização dos tempos de execução . .	66
Figura 4.12–Área de estudo representada em falsa cor pelas bandas 7, 5 e 4	66
Figura 4.13–Análise exploratória dos pontos da imagem e do conjunto de treinamento (conclusão)	70
Figura 4.14–Análise exploratória dos componentes principais da imagem e conjunto de treinamento	71
Figura 4.15–Imagem classificada utilizando os métodos estudados com e sem análise de componentes principais	72
Figura 4.16–Ampliação de parte da imagem	73

LISTA DE TABELAS

Tabela 2.1 – Resolução espacial e espectral do satélite Landsat 8	23
Tabela 3.1 – Classes dos dados a diferentes níveis	51
Tabela 3.2 – Coordenadas da região de estudo	53
Tabela 4.1 – Variância explicada pelos componentes principais	55

1 INTRODUÇÃO

A forma mais convencional de coleta de informações sobre objetos de uma região, em que é necessário deslocar recursos humanos ao local de interesse, apesar de fundamental em algumas situações, está se tornando ultrapassada pois é custosa e por vezes não consegue entregar resultados em tempo hábil, tornando difícil o acompanhamento de grandes áreas e períodos frequentes. Portanto o uso de informações provenientes de sensores remotos, tal como satélites, se torna uma ferramenta imprescindível.

De acordo com Joseph (2005) o ano 1972 marca uma nova era para o sensoriamento remoto com o lançamento do satélite ERTS-1, primeiro satélite que permitia observações repetitivas e sistemáticas dedicado primariamente a observação de terras e que a partir do segundo lançamento passou a ser chamado de Landsat. Esse avanço, aliado as novas tecnologias emergentes da década de 80 promoveu o incentivo e o progresso de muitos países, inclusive o Brasil, no que diz respeito a lançamentos, operação e monitoramento por sensores aeroportados.

Entretanto esse avanço não faria sentido se os procedimentos para tratar as informações contidas nas imagens de forma a gerar resultados interpretáveis não evoluísse conjuntamente. A necessidade de processamento advém da quantidade massiva de informação presente nas imagens coletas, afim de facilitar seu entendimento e interpretação, agindo como um interprete extraindo apenas a parte essencial para uma determinada aplicação (CRÓSTA, 1992).

Neste sentido, o estudo de modelagem matemática e estatística de dados deve acompanhar o progresso tecnológico permitindo formas cada vez mais poderosas de extrair informação significativa das fontes de dados. Um dos métodos mais consagrados e utilizados nesse contexto é a classificação por máxima verossimilhança, que assume que cada uma das classes de informação presentes na imagem possui distribuição normal multivariada com média μ_i e matriz de covariância Σ_i .

A medida que o conceito de inteligência artificial ganhava força, as redes neurais artificiais se destacaram pela sua capacidade de aprendizado e ampla aplicação em diversas áreas da ciência, sobretudo a rede neural **multy-layer perceptron** (MLP). Com o tempo esse estudo chegou a área de sensoriamento remoto, onde foi utilizado para verificar como esta rede se comporta frente ao problema de classificar imagens multiespectrais de satélite e consecutivamente comparar o seu desempenho com o método estatístico mais difundido na área.

É vasta a literatura a esse respeito e de forma geral os estudos chegam a mesma conclusão, a rede MLP atinge resultados superiores ao método de classificação por máxima verossimilhança, porém o esforço computacional é excessivo e o tempo de execução se torna um problema grave sendo necessário o uso de computadores de alto desempenho e um grande nível de paralelismo.

Nos trabalhos de Benediktsson, Swain e Ersoy (1990) e de Zhuang *et al.* (1991), por exemplo, os autores concluem que um dos principais benefícios de se trabalhar com redes neurais é que estas, além de obterem resultados superiores aos de máxima verossimilhança, não precisam de nenhuma suposição quanto a distribuição dos dados, mas que o treinamento da rede demora a convergir.

Mesmo para um número elevado de quantidade de classes as redes neurais apresentam bons resultados, como é apresentado em (KANELLOPOULOS *et al.*, 1992), onde o autor estuda a classificação de uma imagem com 20 tipos de terrenos e obtém um valor de acurácia igual a 81% para a rede neural contra 52.6% para a máxima verossimilhança, uma diferença de 28.4%.

Apesar do amplo estudo de classificadores neurais aplicados as imagens de sensoriamento remoto, a maioria destes foca a atenção na rede MLP e em formas de otimizar seu algoritmo de treinamento afim de agilizá-lo. Apesar do amplo uso de redes neurais não supervisionadas como método de pré-classificação de imagens, em geral usando o mapa auto-organizável de Kohonen (**Kohonen self organization maps** - KSOM), apenas uma pequena parte dos estudos tratam este como um método para a geração de um produto final. Enquanto isso outros métodos não supervisionados, como a rede GNG, são raramente levados em consideração ou apenas utilizados como ferramentas auxiliares para métodos híbridos (LALONDE, 2005).

Neste contexto, objetiva-se com este trabalho buscar uma metodologia autônoma para classificação de imagens provenientes de sensoriamento remoto e que proporcionem uma classificação adequada com uma performance eficiente. Para tanto utilizar-se-á a rede neural artificial GNG e seus resultados serão comparados com o classificador de máxima verossimilhança e o método **k-means** por meio do índice **kappa** e tempo de execução.

2 REVISÃO BIBLIOGRÁFICA

2.1 Sensoriamento remoto

Colwell (1997) define sensoriamento remoto como a arte, ciência e tecnologia de obtenção de informação confiável sobre objetos físicos e ambientais pelo processo de gravação, mensuração e interpretação de imagens e representação digital de padrões de energia derivados por meio de sistemas de sensores remotos. Este processo é alcançado por meio da medição de radiação eletromagnética emitida ou refletida por um objeto, tal como o sol e a superfície terrestre.

Esta área é importante, sobretudo em países com uma dimensão territorial extensa como a do Brasil, pois permite o monitoramento contínuo e periódico ao longo do tempo por meio de diversas observações de uma mesma área, por vezes ampla ou de difícil acesso, de forma automatizada, isto é, sem que seja necessária a alocação de uma força massiva a fim de cobrir e mensurar grandes áreas.

Atualmente existem diversas aplicações do uso de sensores remotos, tais como: previsões climáticas; acompanhamento e transmissão de eventos ao redor do globo pela mídia; sistema de navegação embutidos em carros, celulares e computadores; mapeamento de solo e fronteiras; acompanhamento de áreas de desmatamento, queimadas, desabamento e de produção agrícola. (AKINTUNDE; ADZANDEH; FABIYI, 2016; FOX *et al.*, 2016; MITCHELL *et al.*, 2016)

2.1.1 Aquisição e características de imagens de satélite

A teoria de sensoriamento remoto é pautada sobre a interação entre fonte de radiação, objeto e sensor como mostra a Figura 2.1.

Os feixes de luz solar ao entrarem pela atmosfera são separados em diversos comprimentos de onda λ que atingem a superfície terrestre e seus objetos. Parte dessa energia incidente é refletida, parte é absorvida e parte é refratada, de forma que:

$$E_I(\lambda) = E_R(\lambda) + E_A(\lambda) + E_T(\lambda) \quad (2.1)$$

sendo:

E_I a energia incidente.

E_R a energia refletida.

E_A a energia absorvida.

E_T a energia refratada.

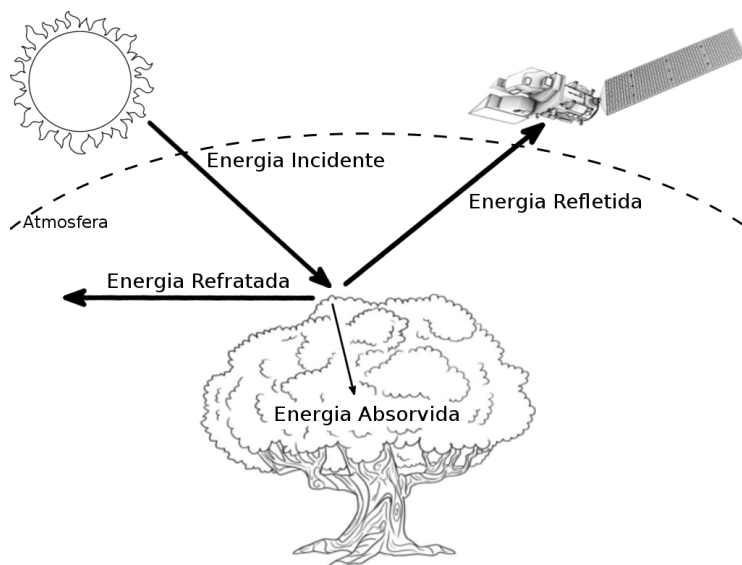
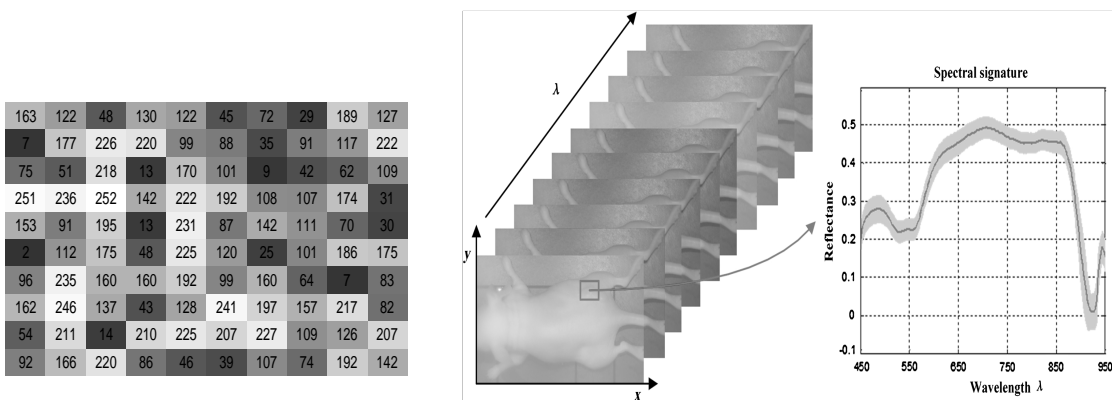


Figura 2.1 – Relação entre fonte de radiação (sol), objeto (árvore) e sensor (satélite)

As imagens multiespectrais de satélite são compostas por várias camadas denominadas de bandas. Cada banda é o resultado da radiação refletida por um comprimento de onda diferente e juntas compõem um espaço multidimensional de atributos, ao qual os objetos pertencem e são classificados, como mostra a Figura 2.2.



(a) Banda espectral registrada em número digital (**Digital Number - DN**)

(b) Composição de bandas espectrais em função do comprimento de onda
Fonte: (LU *et al.*, 2014)

Figura 2.2 – Representação de uma imagem multiespectral

Objetos absorvem e refletem radiação de forma distinta desta forma pode-se, com certa precisão, determinar o objeto em si medindo-se a radiação refletida por este. Um exemplo claro disso é a comparação de vegetação seca e saudável como mostra a Figura 2.3. Percebe-se que na região espectral visível ao olho humano, cerca de $0,4\mu m$ a $0,7\mu m$, a vegetação saudável reflete mais luz no comprimento de onda que compreende a banda verde, o que explica por que enxerga-se esse tipo de vegetação dessa cor, enquanto que a reflectância máxima da curva para a vegetação seca ocorre na banda vermelha.

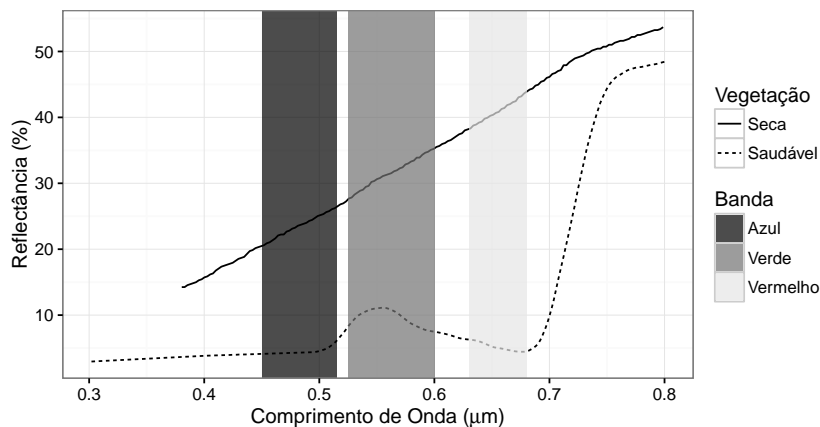


Figura 2.3 – Curvas espectrais para vegetação seca e saudável

Atualmente há diversos tipos de sensores disponíveis, cada um com características próprias, portanto é necessário entender essas características para poder escolher a melhor fonte para uma determinada aplicação. A resolução da imagem é uma das características mais importantes a se ter conhecimento e são subdivididas em 4 tipos.

A resolução espacial define a área de um **pixel** na imagem. Portanto se o sensor tem uma resolução espacial de 20 metros, então o **pixel** terá uma área de $400m^2$, $20m \times 20m$. Quanto maior a resolução espacial menor é o tamanho de cada pixel, o que significa dizer que a imagem é mais detalhada e permite visualizar objetos menores.

Resolução espectral é definida pela quantidade de bandas registradas e pelo tamanho de comprimento de onda de cada banda. Resoluções espectrais maiores indicam mais bandas com comprimentos menores, ou seja, é possível compor mais imagens a partir das bandas existentes.

Uma boa resolução espectral é importante devido às diferenças relativas de reflectância entre os materiais da superfície da terra, que permitem distingui-los uns dos outros, em determinados comprimentos de onda (ALMEIDA *et al.*, 2012).

Resolução radiométrica define o número de níveis de radiância, tons de cinza, que o detector pode discriminar. A quantidade de níveis é expressa em termos do número de dígitos binários (**bits**) necessário para armazenar, em forma digital (DN), o valor de brilho máximo (GONÇALVEZ, 1997). Uma resolução radiométrica de 8 **bits** indica que a imagem possui 256 níveis de cinza (2^8), que variam de 0 a 255 e geram imagens mais ricas em detalhes quando comparadas a imagens de resoluções radiométricas menores como uma imagem com 2 **bits**, preta e branca, como mostra a Figura 2.4

Resolução temporal se relaciona com a frequência em que um sensor revisita uma mesma área. Para aplicações como monitoramento florestal, ciclo fenológico de culturas agricultura ou desastres ambientais, este tipo de resolução é de grande interesse pois se faz necessária a averiguação contínua de uma certa área ao longo do tempo para se detectar mudanças de forma

rápida, eficiente e alocar recursos necessários para a resolução do problema.

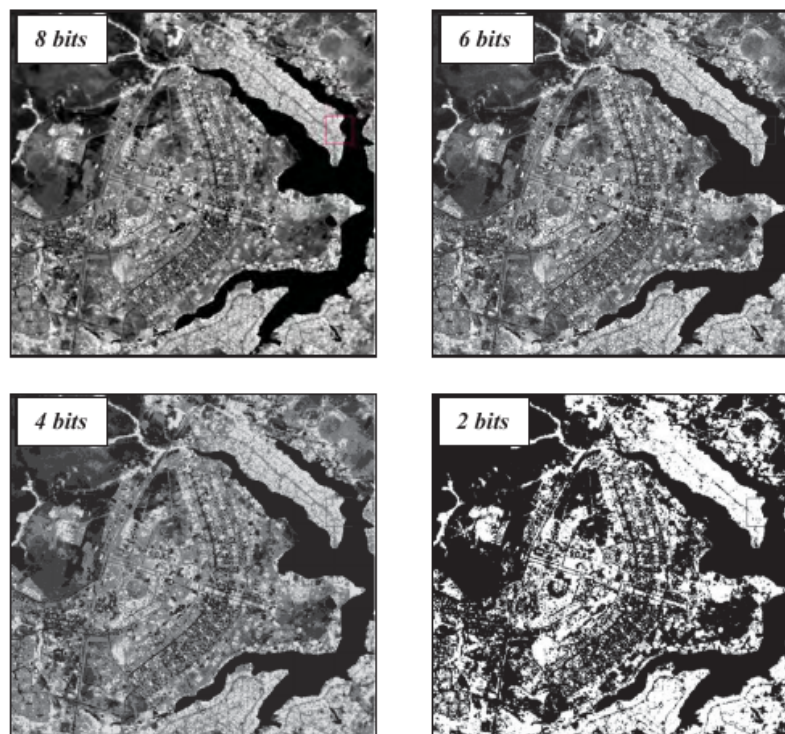


Figura 2.4 – Comparação de imagens a diferentes resolução radiométricas
Fonte: (ALMEIDA *et al.*, 2012)

2.1.2 Landsat 8

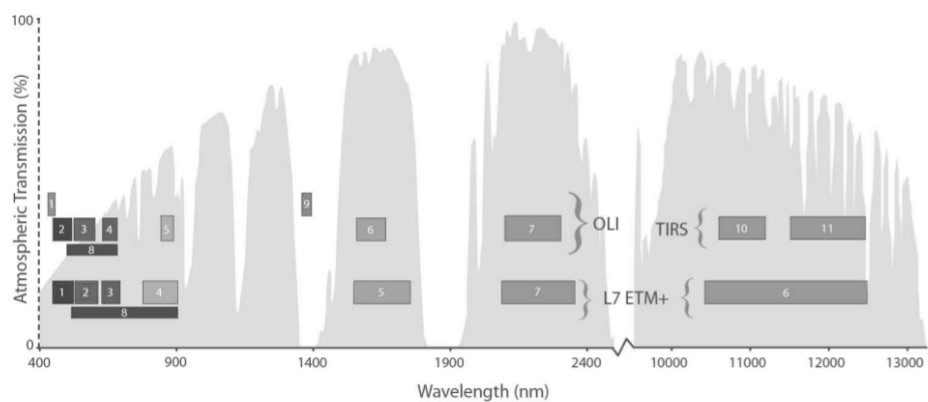
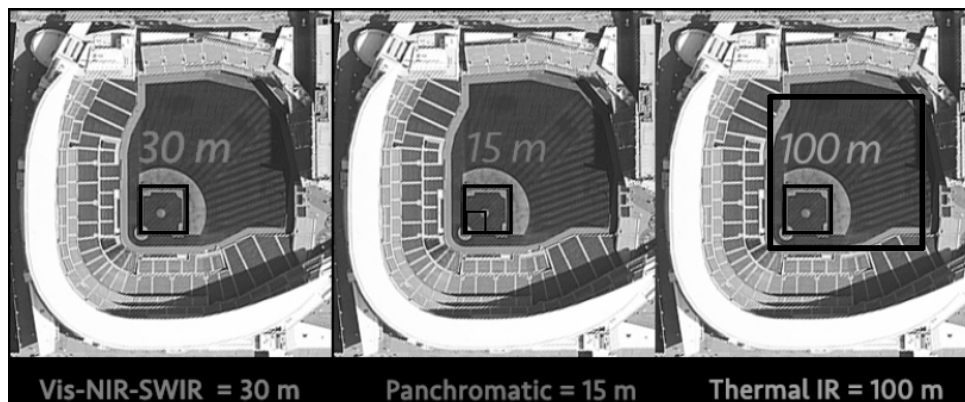
Landsat 8, lançado em 2013, é o oitavo satélite do programa Landsat, produto da colaboração entre NASA e USGS (**U. S. Geological Survey**), que começou em 1965 com a ideia do diretor do USGS, William Pecora, de utilizar satélites de sensoriamento remoto para coletar informações a respeito dos recursos naturais do nosso planeta, mas só em 1970 que o projeto foi autorizado e dois anos mais tarde o primeiro satélite do programa, Landsat 1, foi lançado, abrindo novos horizontes e definindo um marco na história do sensoriamento remoto (NASA, 2006).

O sensor possui uma resolução radiométrica de **16 bits** gerando uma imagem com 65536 níveis de cinza, o que proporciona maior suavidade e riqueza de detalhes. Sua resolução temporal é de 16 dias e de acordo com o site da NASA¹ tem produzido cerca de 725 imagens por dia, aumentando a chance de obter cenas livres de nuvens. As resoluções espacial e espectral podem ser visualizadas na Tabela 2.1, uma comparação com a versão anterior do satélite é apresentada na Figura 2.5 na qual pode-se notar uma melhoria na resolução espectral da nova versão, que possui mais bandas espectrais com comprimentos de onda mais precisos. A Figura 2.6 exemplifica as dimensões dos **pixels** para cada uma das resoluções espaciais.

¹ <http://landsat.gsfc.nasa.gov/>

Tabela 2.1 – Resolução espacial e espectral do satélite Landsat 8

Banda	Comprimento de onda (μm)	Resolução Espacial (m)
Banda 1	0.433 - 0.453	30
Banda 2	0.450 - 0.515	30
Banda 3	0.525 - 0.600	30
Banda 4	0.630 - 0.680	30
Banda 5	0.845 - 0.885	30
Banda 6	1.560 - 1.660	30
Banda 7	2.100 - 2.300	30
Banda 8	0.500 - 0.680	15
Banda 9	1.360 - 1.390	30
Banda 10	10.60 - 11.20	100
Banda 11	11.50 - 12.50	100

Figura 2.5 – Comparação de resoluções entre Landsat 7 e Landsat 8²Figura 2.6 – Dimensão de um **pixel** a diferentes resoluções³

² Adaptado de http://landsat.usgs.gov/L8_band_combos.php

³ Adaptado de <http://landsat.gsfc.nasa.gov/landsat-8/landsat-8-overview>

2.1.2.1 Características das bandas do satélite Landsat 8

É importante conhecer as bandas espectrais do sensor utilizado, pois estas podem ser escolhidas de forma a melhor a interpretação sobre o terreno estudado bem como as análises realizadas sobre o conjunto de atributos escolhido ou ainda serem utilizadas na criação de outros índices relevantes, como por exemplo o índice de vegetação por diferenças normalizadas (**Normalized Difference Vegetation Index - NDVI**)(ROUSE *et al.*, 1974), que quantifica o volume e a condição em que a vegetação se encontra em uma determinada região de estudo.

A banda 1 registra as frequências relacionadas ao azul escuro e violeta e possui dois usos principais: expor águas rasas e localizar pequenas partículas como poeira e fumaça.

As bandas 2, 3 e 4 (azul, verde e vermelho, respectivamente) compreendem a porção do espectro visível ao olho humano e podem ser utilizadas para compor a imagem em cores reais facilitando a visualização e elementos da área para o pesquisador.

A região espectral próxima ao infravermelho (**Near Infrared - NIR**) está presente na banda 5 e é particularmente útil para evidenciar vegetação sadia. Pode ser usada isolada ou combinada com a banda vermelha, de acordo com a eq. (2.2), para calcular o NDVI e outros índices úteis, o que permite medir de forma mais precisa a vegetação do que apenas olhando-se a porção verde da imagem.

$$NDVI = \frac{\text{Banda 5} - \text{Banda 4}}{\text{Banda 5} + \text{Banda 4}} \quad (2.2)$$

As bandas 6 e 7 cobrem partes distintas do infravermelho de ondas curtas (**Short Wave Infrared - SWIR**). Usada para mapear diferentes tipos de solo e rochas também consegue distinguir solo úmido e seco.

A banda 8 é uma camada pancromática, ou seja, é sensível a todas as cores da porção visível do espectro o que significa que objetos de uma determinada cor não terão o seu brilho alterado por conta do comprimento de onda incidente. Por conseguir captar mais luz de uma só vez esta é a banda com maior resolução espacial e portanto a imagem proveniente dessa banda é mais detalhada.

A presença de nuvens em imagens de satélite impossibilita ou interfere na leitura da radiação refletida pela superfície gerando dados de baixa qualidade. Em vista desse problema a banda 9 foi projetada para detectar objetos com reflectância muito intensa ou acima da atmosfera terrestre uma vez que esta absorve, quase que completamente, a energia no comprimento de onda compreendido pela banda.

As bandas 10 e 11 são camadas termais de infravermelho (**Thermal Infrared - TIR**) e registram calor emanado do solo ao invés de medir temperatura do ar como os satélites climatológicos.

2.1.3 Espaço de atributos e métodos de classificação

O conjunto de p bandas espectrais forma um espaço de atributos de forma que um **pixel** é interpretado como um ponto neste espaço p -dimensional, cujas coordenadas são expressas em termo da reflectância registrada em cada banda.

Se as bandas promovem uma boa discriminação então espera-se que haja a formação de agrupamento de **pixels** que sejam correspondentes aos tipos de cobertura terrestre presentes na imagem, como mostra a Figura 2.7a. Entretanto um cenário mais condizente com a realidade seria o representado pela Figura 2.7b, em que as classes apresentam subdivisões causadas por fatores como: diferentes umidades, tipos de solo sob a vegetação e influências topográficas (GONÇALVEZ, 2009).

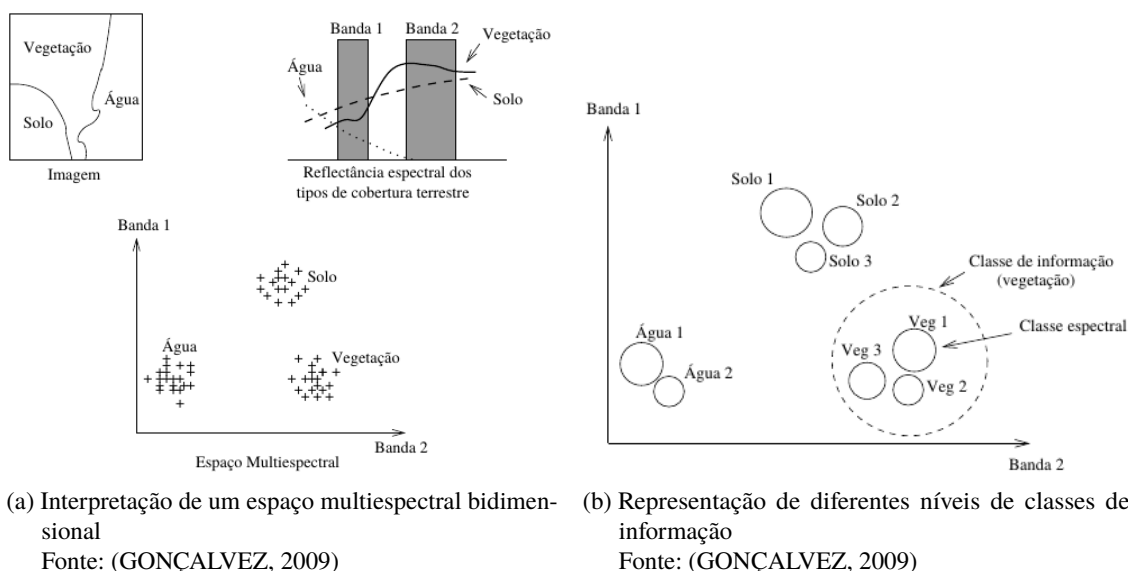


Figura 2.7 – Representação de classes no espaço de atributos

Neste sentido, os métodos de classificação buscam criar regiões afim de separar as classes de acordo com os atributos de cada elemento de modo que, elementos de uma mesma classe são mais próximos entre si em relação a elementos pertencentes as demais classes.

Os métodos de classificação podem ser divididos em três grupos: supervisionados, não supervisionados e híbridos.

2.1.3.1 Métodos supervisionados

Para a utilização desse tipo de método é necessário conhecer a qual classe pertence cada um dos elementos de uma amostra, chamada de conjunto de treinamento. O conjunto de treinamento é utilizado para calcular uma ou mais medidas que por sua vez são utilizadas para rotular um elemento com o código da classe a que deve pertencer.

Os métodos supervisionados podem ainda ser subdivididos em paramétricos ou não paramétricos. Os métodos paramétricos assumem que os dados possuem alguma distribuição de probabilidade específica, geralmente a distribuição normal multivariada, e os parâmetros são estimados para serem utilizados pelo algoritmo de classificação (MATHER; KOCH, 2011). Os métodos não-paramétricos, por outro lado, não assumem nenhuma distribuição de probabilidade e são mais robustos, pois são mais eficientes em espaços com maior variabilidade de distribuições de classes (GONÇALVEZ, 1997).

De forma geral os métodos supervisionados são mais precisos uma vez que se conhece a classe real a qual os elementos pertencem na fase de treinamento, além disso a comparação de imagens classificadas por estes métodos, seja com imagens de regiões próximas ou a mesma imagem em tempos distintos, é facilitada por se conhecer e definir previamente a quantidade e as classes presentes.

Por outro lado, as desvantagens podem ser enumeradas da seguinte maneira (CAMPBELL; WYNNE, 2011):

- O analista que seleciona e classifica as regiões para o treinamento impõem, de certa maneira, uma estrutura que pode não combinar com a classificação natural que existe nos dados e portanto pode não ser distinta ou bem definida no espaço multidimensional.
- Por conta da dificuldade em se observar subdivisões mais refinadas diretamente da imagem, o conjunto de treinamento é usualmente selecionado utilizando como referencia as classes de informação. Entretanto uma área que é facilmente identificada como floresta pode possuir muita diversidade com respeito a densidade, idade e sombreamento e portanto formar uma área de treinamento pobre.
- Apesar da experiência ou esforço do analista é possível que o conjunto de treinamento não represente de forma adequada as condições presentes em toda a imagem, especialmente quando a área estudada é grande, complexa ou inacessível.
- A seleção e identificação de classes para construir o conjunto de treinamento é um processo demorado, desgastante e custoso.
- Classificação supervisionada não permite reconhecer e rotular classes que não foram apresentadas na fase de treinamento, que não foram especificadas pelo analista por falta de conhecimento ou por representar uma porção muito pequena da imagem por exemplo.

Fica claro que a construção do conjunto de treinamento é um fator primordial no uso de métodos supervisionados, de forma que o conjunto deve ser representativo, isto é, represente de forma fidedigna as características espectrais da imagem. Para isso é necessário que as áreas sejam homogenias em relação as classes de informação; não devem incluir regiões atípicas e nem regiões de fronteiras entre categorias.

2.1.3.2 Métodos não supervisionados

O uso desse tipo de método tem crescido nos últimos tempos frente ao grande aumento no volume de dados a disposição nas mais diversas áreas, sejam provenientes de mídias sociais ou qualquer outra fonte. Entretanto esses dados geralmente não são classificados e rotulá-los pode ser um processo custoso, como nos casos de indexação de vídeos, diagnósticos médicos, entre outros (SANCHES, 2003) e portanto existe a necessidade de utilizar métodos que consigam extrair os padrões naturais presentes nesses dados.

Métodos não supervisionados buscam essas estruturas no espaço multidimensional de atributos por meio de métricas de similaridade, como medidas de distância por exemplo. Diferentemente dos métodos supervisionados, estes não necessitam de nenhum conhecimento anterior a classificação, seja sobre distribuição de probabilidade ou sobre o grupo dos elementos a serem classificados, portanto o erro humano é minimizado. As premissas equivocadas que o analista possa ter em relação a região de estudo pouco ou nada influenciará na classificação.

Como dito na seção anterior, classes que representam porções pequenas da imagem podem permanecer irreconhecíveis durante o processo de classificação supervisionada sendo incorporados a classes que não pertencem gerando erro e imprecisão em toda a classificação. Isto não ocorre na classificação não supervisionada, classes únicas são reconhecidas como unidades distintas. Além disso, métodos não supervisionados são geralmente mais uniformes com relação a composição espectral quando comparados com os métodos supervisionados (CAMPBELL; WYNNE, 2011).

As desvantagens desses métodos são, de forma geral, relacionadas a combinação entre as classes fornecidas pelos métodos e as classes que são de interesse para o usuário de maneira que raramente existirá uma correspondência biunívoca entre as duas (CAMPBELL; WYNNE, 2011).

2.1.3.3 Métodos híbridos

Métodos que incorporam ambas as classificações supervisionada e não supervisionada afim de combinar suas vantagens e eliminar os problemas individuais.

Usualmente a primeira etapa do processo é não supervisionada, para que todas os padrões naturais da imagem sejam representados e facilitar a construção do conjunto de treinamento, posteriormente é realizado o procedimento supervisionado para refinar os resultados (AI, 2011).

2.2 Redes Neurais Artificiais

Proposta em 1943 pelo matemático Walter Pitts e o psiquiatra Warren McCulloch, redes neurais artificiais são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes, que adquirem conhecimento através da experiência por meio da composição de seus neurônios e conexões entre eles (MCCULLOCH; PITTS, 1943).

Tudo o que um ser humano faz, sente, lembra, pensa ou aprende é resultado do sistema nervoso, que é constituído por mais de 100 bilhões de neurônios interconectados. Toda a informação é transmitida entre os neurônios por meio de impulsos elétricos ou químicos e a medida que experiências ocorrem na vida de uma pessoa essas conexões são fortalecidas, enfraquecidas ou novas conexões podem ser criadas.

O mesmo ocorre com as redes neurais artificiais. Inicialmente são apresentados sinais de entrada que são processados pelos neurônios artificiais e transmitidos através da rede modificando os pesos sinápticos entre as conexões dos neurônios.

Como uma rede interconectada de neurônios, uma RNA pode ser vista basicamente como um grafo conexo, ou seja, um conjunto de vértices chamados de neurônios, que são unidades de processamento de dados, conectados através de camadas por arestas como pode ser visto na Figura 2.8

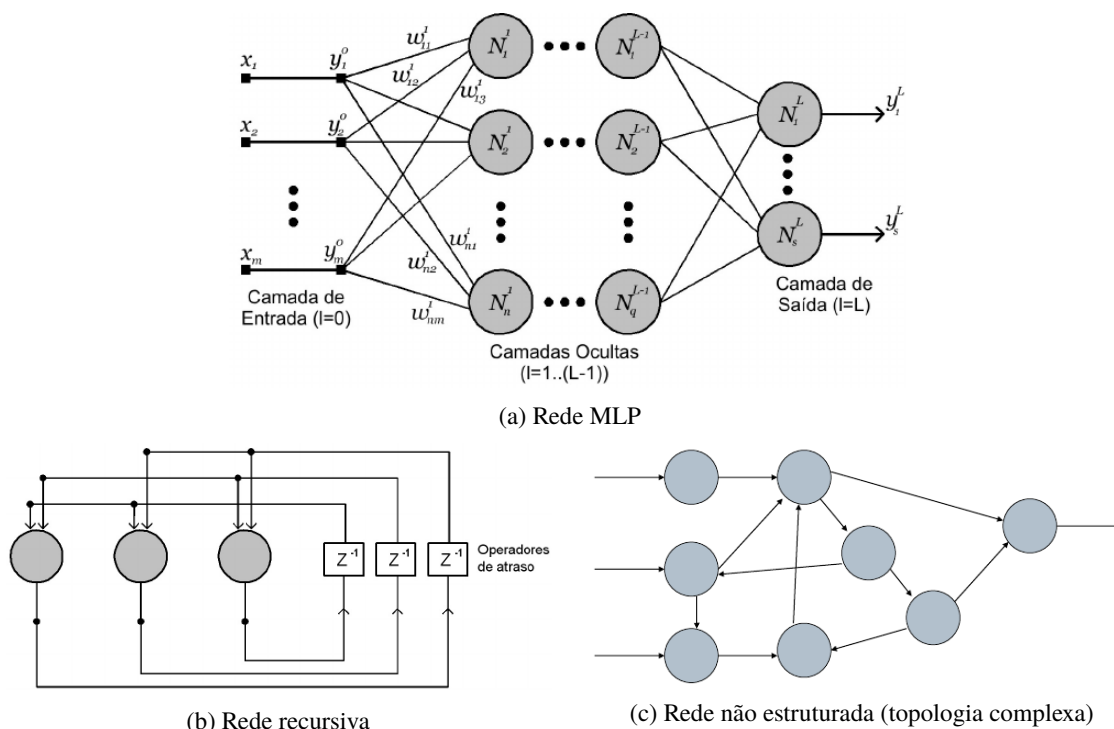


Figura 2.8 – Exemplo de redes neurais artificiais

2.2.1 Neurônio artificial

Um neurônio artificial pode ser representado de acordo com a Figura 2.9.

O processamento de dados dentro de um neurônio j em uma iteração t começa com a apresentação de atributos x_1, \dots, x_m para a camada de entrada. Estes são ponderados por pesos sinápticos $w_{j1}(t), \dots, w_{jm}(t)$ que indicam a influência dos atributos na saída da unidade, a soma ponderada produz um nível de atividade $v_j(t)$ e se este exceder um certo limite dado pela função de transferência $\varphi(\cdot)$, a unidade produz uma resposta de saída $y_j(t)$.

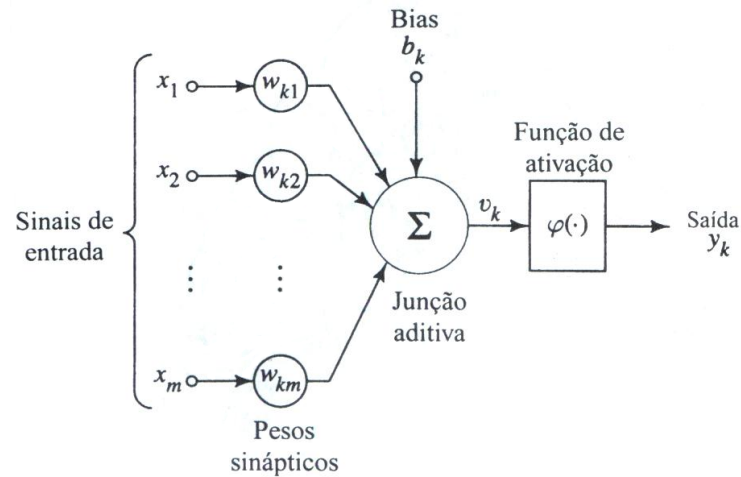


Figura 2.9 – Neurônio Artificial
Fonte: (HAYKIN, 2008)

Em um neurônio artificial, também é comum a presença de um parâmetro conhecido como **bias**, o qual tem o efeito de aumentar ou diminuir a entrada líquida da função de ativação, dependendo se ele é positivo ou negativo (BASGALUPP, 2005).

Escrevendo-se em termos matemáticos a soma ponderada é obtida pela eq. (2.3).

$$u_j(t) = \sum_{i=1}^m w_{ji}(t)x_i \quad (2.3)$$

Acrescentando o **bias** obtêm-se a expressão para o nível de atividade v_j .

$$v_j(t) = u_j(t) + b_j(t) \quad (2.4)$$

Por fim a saída é dada pela eq. (2.5).

$$y_j(t) = \varphi(v_j(t)) \quad (2.5)$$

A função de transferência φ tem o papel de limitar a amplitude da saída do neurônio, geralmente no intervalo normalizado de $[0, 1]$, e em casos específicos $[-1, 1]$. A seguir são dados alguns exemplos de funções de transferência comumente utilizadas por neurônios artificiais segundo Braga, Ludermir e Carvalho (2000).

A função linear, representada pela Figura 2.10 é definida pela eq. (2.6)

$$\varphi(v) = \alpha v \quad (2.6)$$

em que $\alpha \in \mathbb{R}$

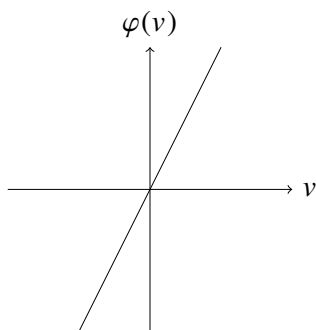


Figura 2.10 – Função de transferência linear

A função rampa, visualizada na Figura 2.11 e definida pela eq. (2.7), é uma função linear limitada a valores de saída $[-\gamma, \gamma]$

$$\varphi(v) = \begin{cases} \gamma & \text{se } v \geq \gamma \\ \alpha v & \text{se } -\gamma < v < \gamma \\ -\gamma & \text{se } v \leq -\gamma \end{cases} \quad (2.7)$$

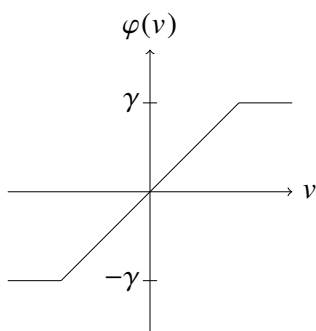


Figura 2.11 – Função de transferência rampa

A função degrau, observada na Figura 2.12, possui uma saída binária, γ ou $-\gamma$, para todos os valores da variável v . Esta função é representada pela eq. (2.8)

$$\varphi(v) = \begin{cases} \gamma & \text{se } v > 0 \\ -\gamma & \text{se } v \leq 0 \end{cases} \quad (2.8)$$

A função mais utilizada, segundo Haykin (2008), é a função sigmoidal Figura 2.13, pois é diferenciável e não decrescente, características que atendem a possíveis pré-requisitos de métodos numéricos. Dentre as diversas funções sigmoidais pode-se destacar a função tangente hiperbólica e a função logística, definidas pela eq. (2.9) e eq. (2.10) respectivamente.

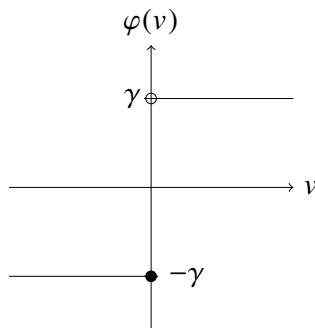


Figura 2.12 – Função de transferência degrau

$$\varphi(v) = \tanh(v) \quad (2.9)$$

$$\varphi(v) = \frac{1}{1 + e^{-\beta v}} \quad (2.10)$$

na qual β é o parâmetro que determina a suavidade da curva.

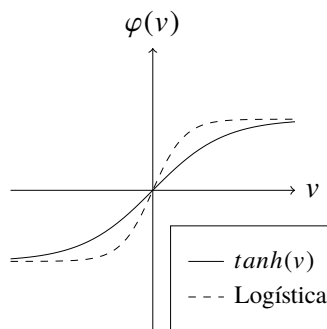


Figura 2.13 – Função de transferência sigmoidal

2.2.2 Perceptron

O perceptron é uma rede com neurônios dispostos em camadas, criada no fim da década de 1950 por Rosenblatt e pode ser considerado o primeiro modelo de redes neurais. Essa é forma mais básica de rede neural sendo composta por uma camada de entrada, que recebe os atributos apresentados a rede e uma camada de saída onde o resultado é calculado e exibido.

Como qualquer outra rede neural o perceptron tem a capacidade de aprender com os dados de entrada a partir de um treinamento e mudança nos pesos sinápticos entretanto possui a limitação de resposta binária, além disso Minsky e Papert (1987) provaram que esta rede pode classificar apenas grupos linearmente separáveis, como mostra a Figura 2.14.

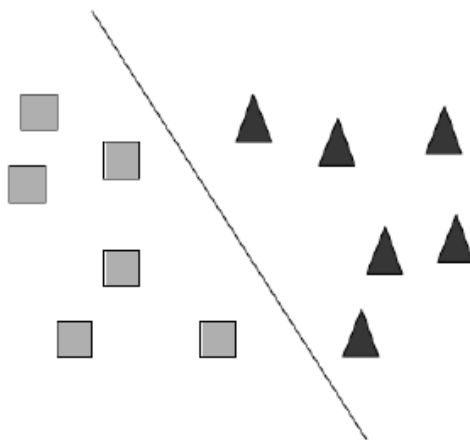


Figura 2.14 – Classificação de espaço linearmente separável via perceptron

2.2.3 Multilayer perceptron

Em virtude das limitações da rede perceptron os pesquisadores da época perderam o interesse na área pois não viam muita utilidade nas redes neurais. Estas ficaram estagnadas por muitos anos até que Rumelhart, Hinton e Williams (1986) desenvolveram o algoritmo de treinamento **backpropagation** e mostraram que é possível treinar eficientemente redes com camadas intermediárias, resultando no modelo de redes neurais artificiais mais utilizado até os dias atuais, a rede MLP, cuja estrutura pode ser vista na Figura 2.15.

Além de herdar os componentes da rede perceptron, uma camada de entrada e uma de saída, a nova rede conta com camadas intermediárias cujo propósito é o refinamento e maior capacidade de extrair padrões dos dados de entrada possibilitando a classificação de espaços não linearmente separáveis como mostra a Figura 2.16.

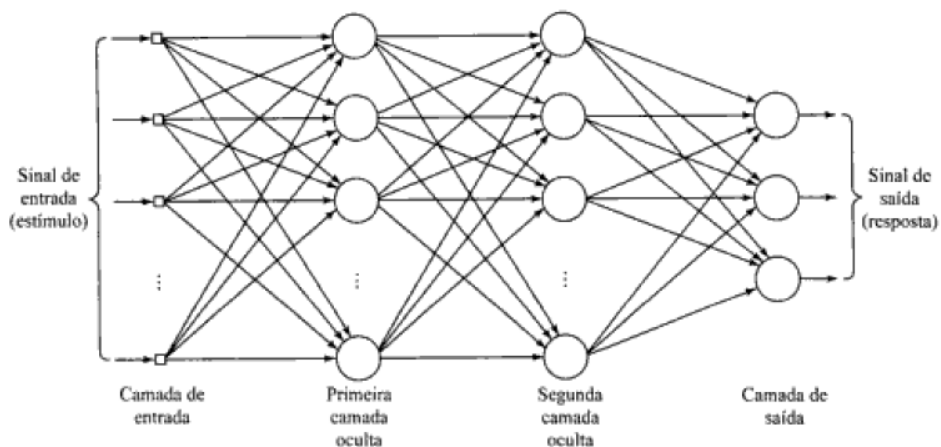


Figura 2.15 – Exemplo de rede MLP com duas camadas intermediárias
Fonte: (HAYKIN, 2008)

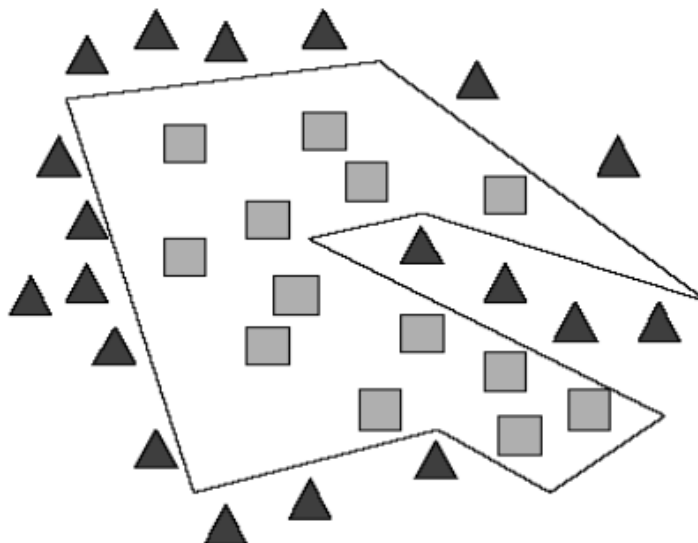


Figura 2.16 – Classificação de espaço não linearmente separável via MLP

Uma dificuldade nas redes MLP está em definir sua estrutura, isto é, determinar a quantidade de camadas intermediárias e a quantidade de neurônios em cada camada. Estudos como os de Cybenko (1989), Hornik, Stinchcombe e White (1989) e Funahashi (1989) investigaram o número de camadas intermediárias necessárias para a implementação de classes de funções em uma RNA. Os resultados indicam que:

- Uma camada intermediária é suficiente para aproximar qualquer função contínua (HERTZ; KROGH; PALMER, 1991).
- Duas camadas intermediárias são o suficiente para aproximar qualquer função matemática (CYBENKO, 1988).

O número de neurônios em cada uma das camadas intermediárias são determinados empiricamente e dependem de vários fatores, tais como quantidade de elementos para treinamento, quantidade de ruído nos dados, complexidade da função a ser aprendida.

Uma quantidade muito elevada de neurônios ocasiona a memorização dos padrões e perda de generalidade para dados não apresentados a rede, o que é chamado de **overfitting**. Em contra partida, uma quantidade reduzida de neurônios pode causar **underfitting**, isto é, ajuste ruim e pouca adaptação da rede além de gasto de tempo excessivo em busca de uma representação ótima, sobrecarregando as unidades por trabalhar com muitas restrições.

Uma regra prática é a de se utilizar cerca de 10% do número de elementos apresentados a rede, com isso a chance de ocorrência de **overfitting** é minimizada, porém não exclui a possibilidade de **underfitting** de forma que o algoritmo pode não convergir durante a fase de treinamento.

2.2.3.1 Processo de Aprendizagem

O algoritmo de treinamento **backpropagation** é uma técnica supervisionada de correção de erros que compara a saída fornecida pela rede com a saída esperada contida no conjunto de treinamento. A partir da comparação entre esses dois valores é calculado um erro, que será retornado camada por camada, ajustando os pesos sinápticos e o **bias** para a iteração seguinte como é mostrado na Figura 2.17 (LIPPMANN, 1987).

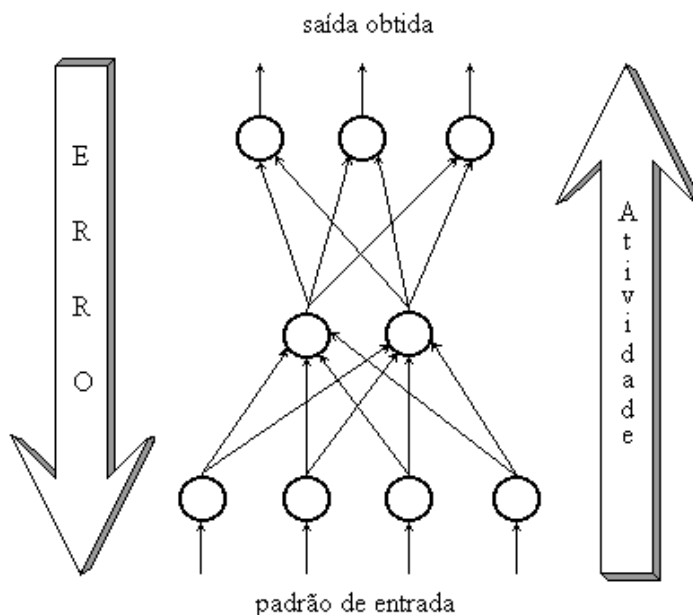


Figura 2.17 – Fluxo de propagação de erros

Dessa forma, o erro de cada neurônio j da camada de saída na iteração t pode ser definido como a diferença entre os valores calculados e esperados, eq. (2.11)

$$e_j(t) = d_j(t) - y_j(t) \quad (2.11)$$

e pode-se obter a soma de erros quadrados de acordo com a eq. (2.12)

$$\mathcal{E}(t) = \frac{1}{2} \sum_{j \in S} e_j^2(t) \quad (2.12)$$

em que S é o conjunto que agrega todos os N neurônios da camada de saída.

Pela regra da cadeia pode-se expressar o gradiente de acordo com eq. (2.13):

$$\frac{\partial \mathcal{E}(t)}{\partial w_{ji}(t)} = \frac{\partial \mathcal{E}(t)}{\partial e_j(t)} \frac{\partial e_j(t)}{\partial y_j(t)} \frac{\partial y_j(t)}{\partial v_j(t)} \frac{\partial v_j(t)}{\partial w_{ji}(t)} \quad (2.13)$$

Diferenciando ambos os lados da eq. (2.12) em relação a $e_j(t)$ e a eq. (2.11) com relação a $y_j(t)$, obtém-se:

$$\frac{\partial \mathcal{E}(t)}{\partial e_j(t)} = e_j(t) \quad (2.14)$$

$$\frac{\partial e_j(t)}{\partial y_j(t)} = -1 \quad (2.15)$$

Agora, diferenciando a eq. (2.5) com respeito a $v_j(t)$ e por fim a eq. (2.4) com respeito a $w_{ji}(t)$.

$$\frac{\partial y_j(t)}{\partial v_j(t)} = \varphi'_j(v_j(t)) \quad (2.16)$$

$$\frac{\partial v_j(t)}{\partial w_{ji}(t)} = y_i(t) \quad (2.17)$$

Substituindo-se as derivadas na eq. (2.13).

$$\frac{\partial \mathcal{E}(t)}{\partial w_{ji}(t)} = -e_j(t)\varphi'(v_j(t))y_i(t) \quad (2.18)$$

A correção $\Delta w_{ji}(t)$ aplicada ao peso sináptico w_{ji} é definida pela regra delta, eq. (2.19) (HAYKIN, 2008).

$$\Delta w_{ji}(t) = -\eta \frac{\partial \mathcal{E}(t)}{\partial w_{ji}(t)} \quad (2.19)$$

em que η é a taxa de aprendizado do método **backpropagation** e o sinal negativo representa o gradiente descendente, ou seja, busca-se a direção para correção dos pesos que minimiza o valor $\mathcal{E}(t)$.

Pode-se reescrever a eq. (2.19) utilizando-se a eq. (2.18), o que leva a eq. (2.20),

$$\Delta w_{ji}(t) = \eta \delta_j(t) y_i(t) \quad (2.20)$$

em que o gradiente local $\delta_j(t)$ é definido abaixo:

$$\begin{aligned} \delta_j(t) &= -\frac{\partial \mathcal{E}(t)}{\partial v_j(t)} \\ &= \frac{\partial \mathcal{E}(t)}{\partial e_j(t)} \frac{\partial e_j(t)}{\partial y_j(t)} \frac{\partial y_j(t)}{\partial v_j(t)} \\ &= e_j(t)\varphi'_j(v_j(t)) \end{aligned}$$

De forma análoga é possível mostrar que para as camadas intermediárias o gradiente local é descrito pela eq. (2.21).

$$\delta_j(t) = \varphi'_j(v_j(t)) \sum_{i=1}^N \delta_i(t) w_{ij}(t) \quad (2.21)$$

A escolha de uma taxa adequada é fundamental para encontrar o ponto de mínimo da superfície de erro, Figura 2.18, pois se for pequena demais o algoritmo terá passos curtos demorando para convergir. Por outro lado, se for grande demais oscilará em torno do mínimo e o algoritmo pode não convergir. A fim de diminuir esse efeito, é introduzido um termo denominado **momentum**, uma constante α de compensação do tamanho do passo, influenciada pelos valores passados de pesos sinápticos. O que leva a regra delta generalizada, eq. (2.22).

$$w_{ji}(t+1) = w_{ji}(t) + \eta y_i(t) \delta_j(t) + \alpha (w_{ji}(t) - w_{ji}(t-1)) \quad (2.22)$$

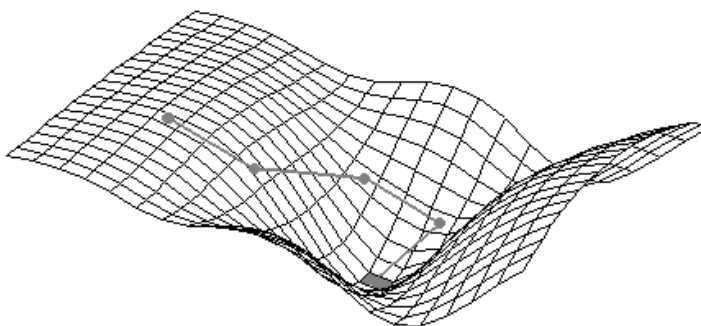


Figura 2.18 – Exemplo de caminho percorrido na superfície de erro pelo algoritmo **backpropagation**

2.2.4 Growing Neural Gas

Um mapa auto-organizável (**self-organizing map** - SOM) é um tipo de RNA não supervisionada que produz uma representação discretizada de baixa dimensão (tipicamente bidimensional) do espaço de atributos, chamado de mapa. Mapas auto-organizáveis são diferentes de outras redes neurais pois utilizam aprendizado competitivo ao invés de aprendizado por correção de erros (como o caso do método **backpropagation** com gradiente descendente) e também por utilizar neurônios vizinhos para preservar propriedades topológicas do espaço de atributos.

Diferentemente das redes apresentadas nas seções anteriores, a rede neural GNG não possui estrutura fixa, ao invés ela busca representar a estrutura topológica do espaço de atributos a partir de uma estrutura mínima e um processo incremental não supervisionado. Para isso a rede proposta por Fritzke (1995) se baseia em dois procedimentos: aprendizagem competitiva

Hebbiana (**Competitive Hebbian Learning - CHL**) (MARTINETZ, 1993) e o método de quantização vetorial **neural gas** (NG) (MARTINETZ; SCHULTEN, 1991).

A técnica CHL assume uma quantidade de centros no \mathbb{R}^p e insere conexões topológicas entre eles avaliando os elementos retirados do conjunto de dados com distribuição $P(\xi)$. Para cada entrada $\mathbf{x} \in P(\xi)$, conecta-se os dois centros mais próximos a este (de acordo com a distância euclidiana) por meio de uma aresta (FRITZKE, 1995).

Martinetz e Schulten (1994) demonstram que esse processo pode resultar na conexão de todos os centros, triangulação de Delaunay, ou em um subgrafo da mesma, denominada triangulação de Delaunay induzida, dependendo de certas características do espaço de atributos. A Figura 2.19 ilustra um exemplo de triangulação de Delaunay, que é gerada com o propósito de maximizar os ângulos das conexões da rede de forma que dado quaisquer 3 neurônios conexos entre si, não há nenhum outro neurônio contido na circunferência formada por estes 3. Os autores explicam ainda que os centros são conectados apenas se a borda em comum dos seus polígonos de Voronói reside pelo menos parcialmente em uma região em que $P(\xi) > 0$, área sombreada da Figura 2.19b.

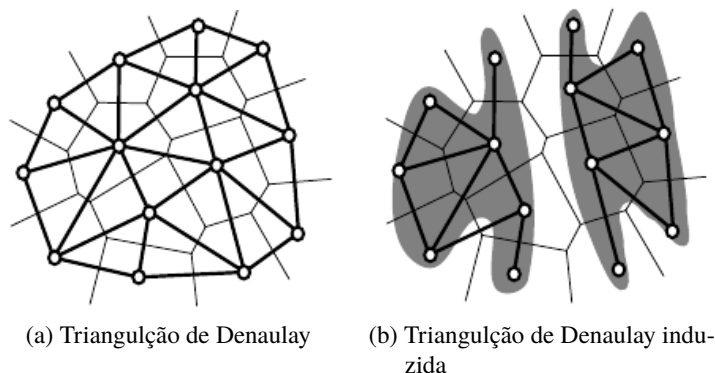


Figura 2.19 – Forma de definir similaridade entre um conjunto de pontos
Fonte: (FRITZKE, 1995)

Os centros sem ligação, conhecidos como unidades mortas, não contribuem para a formação da rede, para se fazer uso deles é necessário movê-los para a região do \mathbb{R}^p em que $P(\xi) \neq 0$. Isso pode ser feito utilizando-se qualquer procedimento de quantização vetorial tal como NG, cujo princípio é adaptar os k centros mais próximos de cada sinal de entrada \mathbf{x} de forma que o valor de k decaia de um valor inicialmente grande para um valor final pequeno. Um valor inicial elevado de k é escolhido para causar a adaptação de grande parte ou até mesmo de todos os centros, movendo-os para regiões não vazias do espaço. Conforme o valor de k diminui, o procedimento passa a reforçar a especificidade de um grupo e o refinamento do ajuste de centros próximos aos dados de entrada.

2.2.4.1 Algoritmo de aprendizagem

A rede GNG se adapta a topologia do espaço de atributos combinando a CHL, para gerar as conexões da grade de neurônios, e o método NG, para distribuir os centros no espaço de entrada. A ideia principal do algoritmo de aprendizagem é adicionar novas unidades a uma rede inicialmente pequena. Esta adição é baseada em medidas de erro calculadas durante os passos de adaptação anteriores, de forma que os novos neurônios são inseridos entre aqueles que possuem o maior erro acumulado. Esta forma incremental de se construir a estrutura da rede aliada ao fato de não precisar conhecer o número total de neurônios a priori tornam a rede GNG especial (LIMA, 2011).

Abaixo segue o algoritmo de treinamento como proposto por Fritzke (1995):

Considere:

- Um conjunto A de unidades, ou nós. Cada unidade $c \in A$ está associada a um vetor de referência $\mathbf{w}_c \in \mathbb{R}^p$. Os vetores de referência podem ser interpretados como posições das unidades no espaço de atributos.
- Um conjunto N de conexões (ou arestas) entre pares de nós. Essas conexões não possuem peso sendo o seu único propósito definir estruturas topológicas de vizinhança

1. Inicializar a rede com duas unidades a e b em posições aleatórias \mathbf{w}_a e \mathbf{w}_b no \mathbb{R}^p .
2. Apresentar um elemento de entrada ξ pertencente a distribuição $P(\xi)$.
3. Encontrar o unidade mais próxima s_1 e a segunda mais próxima s_2 .
4. Incrementar a idade de todas as arestas diretamente ligadas a s_1 .
5. Adicione o quadrado da distancia entre o elemento de entrada e a unidade s_1 a uma variável local:

$$\Delta error(s_1) = \|\mathbf{w}_{s_1} - \xi\|^2$$

6. Mova s_1 e todos os seus vizinhos topológicos em direção a ξ por frações da distância total ϵ_b e ϵ_n , respectivamente:

$$\Delta \mathbf{w}_{s_1} = \epsilon_b(\xi - \mathbf{w}_{s_1})$$

$$\Delta \mathbf{w}_n = \epsilon_n(\xi - \mathbf{w}_n) \quad \text{para todos os vizinhos diretos, } n, \text{ de } s_1$$

7. Se s_1 e s_2 estão conectados por uma aresta, atribua zero para a idade desta aresta. Se esta aresta não existir, crie-a.

8. Remova arestas com idade maior que a_{max} . Se isso resultar em nós sem arestas, remova-os também.
9. Se a quantidade de elementos apresentados a rede for um múltiplo inteiro de λ , insira um novo nó da seguinte maneira:
 - a) Determine a unidade q com o maior erro acumulado.
 - b) Insira uma nova unidade r na metade do caminho entre q e seu vizinho com a maior variável de erro:

$$\mathbf{w}_r = \frac{1}{2}(\mathbf{w}_q + \mathbf{w}_f)$$
 - c) Insira uma aresta conectando a nova unidade r em q e f e remova a aresta original entre q e f .
 - d) Diminua as variáveis de erro de q e f multiplicando-as por uma constante α . Inicialize a variável de erro de r com o novo valor da variável de erro de q .
10. Diminua todas as variáveis de erro multiplicando-as por uma constante d .
11. Se o critério de parada (e.g., tamanho máximo da rede ou alguma medida de performance) ainda não foi satisfeito volte ao passo 2.

O passo 3 do algoritmo apresenta a característica de aprendizado competitivo presente em um mapa auto-organizável. Juntamente com o procedimento de quantização vetorial NG, presente no passo 6, buscam posicionar os neurônios em locais que reflitam o espaço de atributos ($P(\xi) > 0$) a partir do deslocamento do neurônio vencedor e seus vizinhos diretos em direção ao ponto amostrado.

O procedimento CHL é executado pelo passo 7, que insere uma aresta entre as duas unidades mais próximas do sinal de entrada e portanto adiciona uma conexão na triangulação de Delaunay induzida. Com a inserção e remoção de arestas, o algoritmo busca construir essa triangulação sobre a distribuição $P(\xi)$, que por sua vez determina a topologia da rede, isto é, a relação de vizinhança dos centros por meio de suas arestas.

Conforme os neurônios se deslocam, convergindo para **clusters** distintos, é necessário remover as arestas que ligam neurônios de grupos diferentes (passo 8). Este resultado é alcançado por conta do sistema de envelhecimento de arestas em torno da unidade mais próxima (passo 4) combinado com a reinicialização das arestas já existentes entre as duas unidades mais próximas (passo 7).

A soma do quadrado da distância (passo 5) ajuda a identificar unidades em regiões onde o mapeamento entre o espaço de atributos e neurônios ainda apresenta um erro elevado.

Para reduzir o erro nessas regiões novas unidades são inseridas (passo 9), o que permite maior cobertura da rede sobre o espaço de atributos.

A Figura 2.20 elucida o progresso da rede GNG aplicada a um espaço genérico, isto é, a rede parte de uma conexão mínima de 2 neurônios e se adapta ao espaço de atributos por meio da adição de neurônios e conexões entre estes, sendo que ao final do processo cada classe é representado por um subconjunto de conexões.

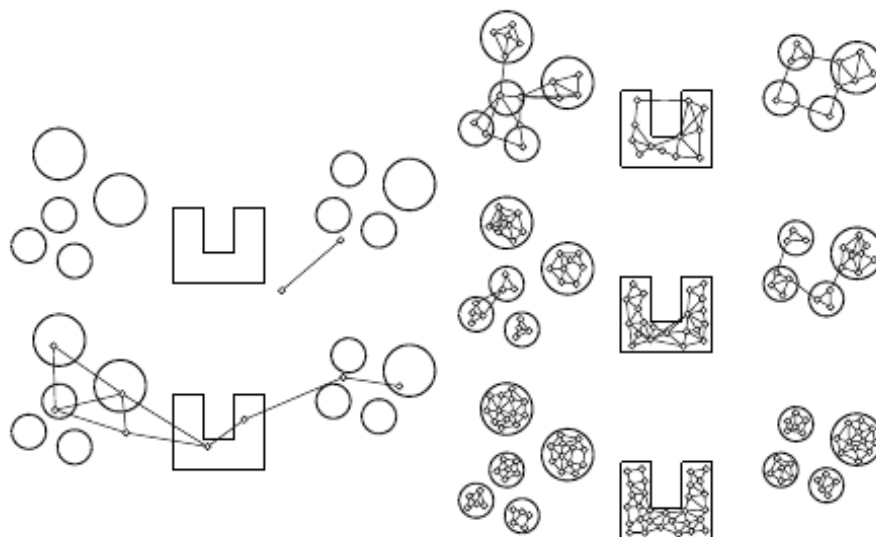


Figura 2.20 – Exemplo de adaptação e crescimento da rede GNG aplicada a um conjunto de dados genérico

2.3 Máxima Verossimilhança

O método estatístico paramétrico supervisionado de classificação mais utilizado em sensoriamento remoto é o método de classificação por máxima verossimilhança (MASELLI *et al.*, 1992; FOODY; MCCULLOCH; YATES, 1995; RODRIGUES; QUEIROZ; GÓMEZ, 2004).

Como classificador supervisionado é necessário fornecer um conjunto de treinamento para o método, entretanto, para que haja resultados satisfatórios este conjunto deve ser bem definido, o que requer um conhecimento da área analisada para que sejam apresentadas informações de todas as classes e, em alguns casos, uma grande quantidade de píxeis.

Quanto maior a dimensão do problema mais píxeis são necessários e o método pode se tornar inviável uma vez que o tamanho das amostras coletadas disponíveis em situações reais pode não ser suficiente. Para contornar este problema pesquisadores utilizam a própria imagem para tentar identificar e definir classes de forma visual, auxiliados por procedimentos de falsa cor, filtros, ou porções de imagem de forma aumentada.

A qualidade da classificação supervisionada é tão boa quanto o conjunto de treinamento escolhido. Dessa forma deve-se amostrar áreas que não apresentem mistura de classes e obter

amostras completas e representativas, isto é, amostras que representem univocamente uma classe e amostras de todas as classes, caso contrário estas podem ser confundidas ou mesmo desconsideradas. Um número igual de elementos de cada classe também é desejável para se evitar um viés na classificação.

No escopo de sensoriamento remoto, o método de classificação da máxima verossimilhança é uma técnica estatística paramétrica de classificação supervisionada que utiliza as amostras de treinamento para estimar os parâmetros da distribuição dos dados. Assume-se que cada uma das classes possui distribuição normal multivariada no conjunto das bandas espectrais, calcula-se a probabilidade de um **pixel** pertencer a cada uma das classes e então este é atribuído ao grupo que apresenta a maior probabilidade de ocorrência, Figura 2.21, condição expressa pela eq. (2.23) (RICHARDS; JIA, 2006).

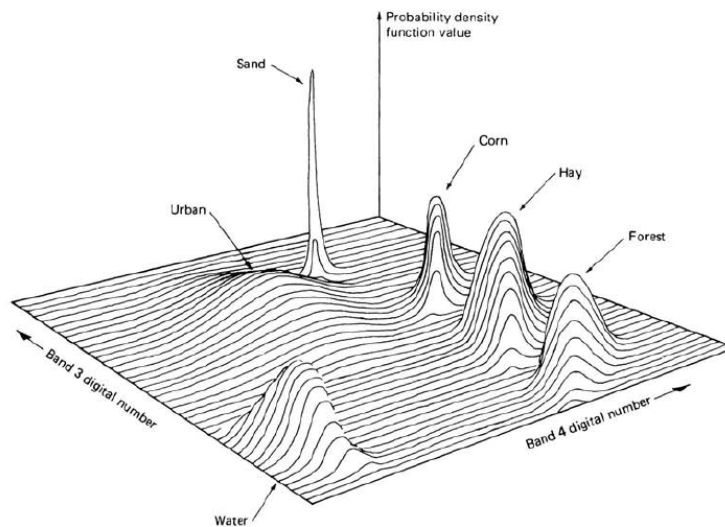


Figura 2.21 – Exemplo de superfície de distribuição normal bivariada para classes distintas no espaço formado por duas bandas espectrais

Fonte: (LILLESAND; KIEFER; CHIPMAN, 2004)

$$\mathbf{x} \in w_i, \quad \text{se } p(w_i|\mathbf{x}) > p(w_j|\mathbf{x}) \quad \forall i \neq j \in \{1, 2, \dots, k\} \quad (2.23)$$

em que $p(w_i|\mathbf{x})$ é a probabilidade da classe w_i ser a correta para um **pixel** na posição \mathbf{x} e k é quantidade total de classes.

De acordo com o teorema de Bayes tem-se que,

$$p(w_i|\mathbf{x}) = \frac{p(\mathbf{x}|w_i)p(w_i)}{p(\mathbf{x})} \quad (2.24)$$

e portanto obtém-se a seguinte expressão:

$$\mathbf{x} \in w_i, \quad \text{se } p(\mathbf{x}|w_i)p(w_i) > p(\mathbf{x}|w_j)p(w_j) \quad \forall i \neq j \in \{1, 2, \dots, k\} \quad (2.25)$$

em que: $p(\mathbf{x}|w_i)$ é a probabilidade de \mathbf{x} pertencer a classe w_i ; $p(w_i)$ é a probabilidade da classe w_i ocorrer na imagem.

Por conveniência matemática é utilizada a função discriminante, definida pela eq. (2.26), como regra de decisão.

$$g_i(\mathbf{x}) = \ln[p(\mathbf{x}|w_i)p(w_i)] = \ln[p(\mathbf{x}|w_i)] + \ln[p(w_i)] \quad (2.26)$$

Logo,

$$\mathbf{x} \in w_i, \quad \text{se} \quad g_i(\mathbf{x}) > g_j(\mathbf{x}) \quad \forall i \neq j \in \{1, 2, \dots, k\} \quad (2.27)$$

Pela condição de normalidade imposta,

$$p(\mathbf{x}|w_i) = \frac{1}{\sqrt{(2\pi)^n |\boldsymbol{\Sigma}_i|}} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i)\right) \quad (2.28)$$

e portanto,

$$g_i(\mathbf{x}) = \ln[p(w_i)] - \frac{n}{2} \ln[2\pi] - \frac{1}{2} \ln[|\boldsymbol{\Sigma}_i|] - \frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \quad (2.29)$$

Ao remover os elementos constantes a todas as classes, a forma final da função discriminante fica determinada pela eq. (2.30) e é usada na regra de decisão enunciada na eq. (2.27).

$$g_i(\mathbf{x}) = \ln[p(w_i)] - \ln[|\boldsymbol{\Sigma}_i|] - (\mathbf{x} - \boldsymbol{\mu}_i)' \boldsymbol{\Sigma}_i^{-1} (\mathbf{x} - \boldsymbol{\mu}_i) \quad (2.30)$$

2.4 K-Means

Método de classificação não supervisionada que busca particionar os n elementos de um conjunto de dados em k grupos. Estes grupos por sua vez são representados por pontos centrais, chamados de centroides, de forma que a distância dos grupos aos seus respectivos centroides seja mínima, como é representado na eq. (2.31) (MACQUEEN, 1967).

$$\min_{C_i, \boldsymbol{\mu}_i} \left(\sum_{i=1}^k \sum_{\mathbf{x} \in C_i} \|\mathbf{x} - \boldsymbol{\mu}_i\|^2 \right) \quad (2.31)$$

em que C_i e $\boldsymbol{\mu}_i$ representam o i -ésimo grupo e centroide respectivamente.

Uma deficiência deste método é a exigência de que se conheça o número k de classes a serem agrupadas e portanto, apesar de ser um método não supervisionado, exige um conhecimento prévio da área a ser estudada por parte do analista.

Outra desvantagem do método é o fato de que a solução do algoritmo pertença a classe dos problemas **NP-Hard (Non-deterministic Polynomial-time hard)**(NANDY *et al.*, 2012), ou seja, não necessariamente existem algoritmos para verificar se a resposta para um problema de decisão dessa classe é correta em tempo polinomial, tão pouco possuem solução determinística em tempo polinomial.

Apesar da complexidade do método **k-means** existem algoritmos, como o algoritmo iterativo de Lloyd (LLOYD, 1982), que convergem rapidamente após poucos passos. Entretanto essa convergência atinge apenas um mínimo local e essa solução local é altamente dependente da inicialização dos centroides no espaço multidimensional de atributos.

O algoritmo de Lloyd é dividido em duas partes que são repetidas até que a modificação dos centroides seja menor que um valor ϵ , matematicamente expressas pela eq. (2.32).

1. Dado k centroides os pontos são classificados de acordo com o centroide mais próximo.
2. A partir dos grupos formados os centroides são recalculados como a média da classe que representam.

$$C_i = \{\mathbf{x} : \|\mathbf{x} - \boldsymbol{\mu}_i\| \leq \|\mathbf{x} - \boldsymbol{\mu}_j\|\} \quad (2.32a)$$

$$\boldsymbol{\mu}_i = \frac{1}{\#C_i} \sum_{\mathbf{x} \in C_i} \mathbf{x} \quad (2.32b)$$

2.5 Análise de Componentes Principais

Análise de componentes principais é uma técnica matemática proposta por Pearson (1901) que tem por objetivo reduzir a dimensionalidade do espaço de atributos por meio de combinações lineares ortogonais. Por ser um método estritamente matemático independe de pressuposições estatísticas tal como distribuições de probabilidade e portanto atende a uma classe maior de problemas e aplicações.

No ambiente de sensoriamento remoto, em que se trabalha com um volume de dados grande o suficiente a ponto do tempo de processamento dos métodos poder se tornar um problema grave, essa técnica pode ser utilizada para reduzir a carga computacional além de facilitar entendimento e interpretação dos dados para tomada de decisões.

A Figura 2.22 mostra um conjunto de pontos no espaço \mathbb{R}^3 , em que é possível notar que duas dimensões possuem grande variabilidade em quanto que as coordenadas da dimensão restante não varia muito. A questão então é a seguinte: É possível remover esta terceira dimensão sem que haja uma perda significativa de informação?

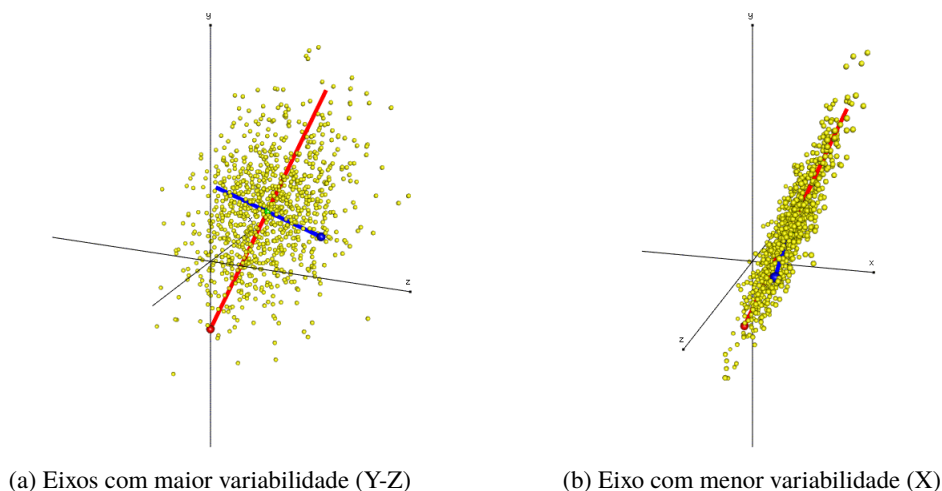


Figura 2.22 – Exemplo de aplicação ACP

Tentar responder a essa questão analisando apenas a matriz de covariância dos dados (eq. (2.33)) não é o suficiente, pois, uma vez que as variáveis são correlacionadas, a diagonal principal da matriz de covariâncias representa a variância da variável projetada no seu respectivo eixo e essa projeção pode gerar uma inflação ou uma diminuição na variância, mascarando o seu efeito real.

$$\mathbf{S} = \begin{bmatrix} 1.01 & 2.01 & 0.77 \\ 2.01 & 4.93 & 0.73 \\ 0.77 & 0.73 & 2.95 \end{bmatrix} \quad (2.33)$$

O que a análise de componentes principais propõem é criar um novo sistema de eixos, chamados de componentes principais, a partir da combinação linear das variáveis originais, de forma que este novo sistema de eixos esteja centrado na direção de maior variabilidade dos dados (representados pelos eixos de cor vermelho, azul e verde da Figura 2.22) (BISHOP, 1995).

Seja \mathbf{x} o vetor de variáveis X_1, \dots, X_p . O vetor \mathbf{z} de componentes principais Z_1, \dots, Z_p é dado pela eq. (2.34),

$$\mathbf{z} = \mathbf{A}\mathbf{x} \quad (2.34)$$

em que \mathbf{A} é uma matriz de coeficientes e sua i -ésima linha, \mathbf{a}_i , satisfaz a seguinte condição:

$$\mathbf{a}_i' \mathbf{a}_i = 1, \quad \forall i \in \{1, \dots, p\}$$

Para o cálculo dos coeficientes toma-se a matriz de covariância $\mathbf{\Sigma}$ das variáveis X_1, \dots, X_p e encontram-se os autovetores normalizados correspondentes aos autovalores $\lambda_1, \dots, \lambda_p$, como indica a eq. (2.35).

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \sigma_{pp} \end{bmatrix}$$

$$\mathbf{a}_i = \frac{\mathbf{v}_i}{\|\mathbf{v}_i\|}, \quad \forall i \in \{1, \dots, p\} \quad (2.35)$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}'_1 \\ \mathbf{a}'_2 \\ \vdots \\ \mathbf{a}'_p \end{bmatrix} \quad (2.36)$$

Assumindo que os autovalores estejam ordenados de forma que $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, então λ_i se relaciona ao i -ésimo componente principal (CP) de forma que $\text{Var}(Z_i) = \lambda_i$, como mostrado a seguir:

$$\begin{aligned} \text{Var}(Z_i) &= \text{Var}(\mathbf{a}'_i \mathbf{x}) \\ &= \mathbf{a}'_i \mathbf{\Sigma} \mathbf{a}_i \\ &= (\mathbf{\Sigma}' \mathbf{a}_i)' \mathbf{a}_i \end{aligned}$$

Como a matriz de covariância é simétrica então $\mathbf{\Sigma}' = \mathbf{\Sigma}$, logo

$$\begin{aligned} \text{Var}(Z_i) &= (\mathbf{\Sigma}' \mathbf{a}_i)' \mathbf{a}_i \\ &= (\mathbf{\Sigma} \mathbf{a}_i)' \mathbf{a}_i \end{aligned}$$

Pela definição de autovetor tem-se que $\mathbf{\Sigma} \mathbf{a}_i = \lambda_i \mathbf{a}_i$, portanto

$$\begin{aligned} \text{Var}(Z_i) &= (\mathbf{\Sigma} \mathbf{a}_i)' \mathbf{a}_i \\ &= (\lambda_i \mathbf{a}_i)' \mathbf{a}_i \\ &= \lambda_i \mathbf{a}'_i \mathbf{a}_i \\ &= \lambda_i \langle \mathbf{a}_i, \mathbf{a}_i \rangle \\ &= \lambda_i \|\mathbf{a}_i\|^2 \\ &= \lambda_i \end{aligned}$$

Uma característica importante do método baseia-se na propriedade de que o traço de uma matriz é igual a soma de seus autovalores, como descreve a eq. (2.37). De acordo com Manly (2008) isto significa que os componentes principais representam, em certo sentido, toda a variabilidade dos dados originais.

$$\text{tr}(\mathbf{\Sigma}) = \sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \lambda_1 + \lambda_2 + \dots + \lambda_p \quad (2.37)$$

Na prática é preferível padronizar os dados de forma que cada uma das variáveis tenha média 0 e variância 1 a fim de não promover nenhum tipo de influência nos componentes por parte de algumas variáveis (JOLLIFFE, 2002). Nessa instância a matriz de covariância e a matriz de correlação \mathbf{C} são idênticas e iguais a

$$\mathbf{C} = \begin{bmatrix} 1 & c_{12} & \dots & c_{1p} \\ c_{21} & 1 & \dots & c_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ c_{p1} & c_{p2} & \dots & 1 \end{bmatrix}$$

e ainda

$$\text{tr}(\mathbf{C}) = p = \lambda_1 + \lambda_2 + \dots + \lambda_p \quad (2.38)$$

Outra característica importante é a de que os componentes principais são não correlacionados, como é mostrado a seguir:

$$\begin{aligned} \text{Cov}(Z_i, Z_j) &= \text{Cov}(\mathbf{a}_i \mathbf{x}, \mathbf{a}_j \mathbf{x}) \\ &= \mathbf{a}_i' \boldsymbol{\Sigma} \mathbf{a}_j \\ &= (\boldsymbol{\Sigma}' \mathbf{a}_i)' \mathbf{a}_j \\ &= (\boldsymbol{\Sigma} \mathbf{a}_i)' \mathbf{a}_j \\ &= (\lambda_i \mathbf{a}_i)' \mathbf{a}_j \\ &= \lambda_i \mathbf{a}_i' \mathbf{a}_j \\ &= \lambda_i \langle \mathbf{a}_i, \mathbf{a}_j \rangle \end{aligned}$$

Como autovetores associados a autovalores distintos são ortogonais então o produto interno entre eles é zero e portanto

$$\text{Cov}(Z_i, Z_j) = 0$$

Aplicando estes procedimentos ao exemplo anterior pode-se chegar na matriz de covariância dos componentes principais abaixo. Como demonstrado, os componentes são não correlacionados, de maneira que os valores da diagonal principal deixam de ser uma projeção e representam a variância real dos dados neste novo espaço de atributos, sendo possível, por meio de algum critério, decidir se o eixo de menor variância pode ser desconsiderado, de forma a reduzir a dimensão para o \mathbb{R}^2 .

$$\underline{\mathbf{S}} = \begin{bmatrix} 6.08 & 0.00 & 0.0 \\ 0.00 & 2.72 & 0.0 \\ 0.00 & 0.00 & 0.1 \end{bmatrix}$$

Existem dois critérios básicos para definir a quantidade k de componentes que não serão descartados. O critério de Kaiser determina que sejam utilizados os k componentes cujos autovalores sejam maiores ou iguais a 1 (KAISER, 1960). Outro critério largamente utilizado na prática é o critério da variância acumulada que determina o uso dos k primeiros componentes que representem uma variância acumulada maior ou igual a um nível pré-determinado de interesse, usualmente estabelecido como 80%. Este último geralmente leva a escolha de uma quantidade menor de componentes, o que é preferível pois além de reduzir a carga computacional ainda gera resultados interpretáveis.

Para que a análise de componentes principais tenha um desempenho satisfatório, isto é, reduza a dimensionalidade para um tamanho k em que se consiga obter um resultado de forma viável, é importante que as variáveis X_1, \dots, X_p sejam bem correlacionadas, de maneira que quanto mais correlacionadas as variáveis forem, maior será a redução da dimensão do problema.

3 MATERIAL E MÉTODOS

O que se vê em muitos estudos que comparam métodos de classificação aplicados a imagens de satélite é a falta de uma análise mais rigorosa dos resultados em termos gerais, pois estes são limitados usualmente a:

- Dados simulados.
- Conjunto de treinamento e teste não independentes, uma vez que são extraídos da mesma população.
- Comparação de valores absolutos de estatísticas sem medidas de desvio ou teste de hipótese.
- Utilização de um método de classificação inicial para determinar uma “verdade de referência” (já que imagens de satélite não possuem rótulos reais relacionando a qual classe cada pixel pertence) e então os métodos de interesse são comparados utilizando estas classes pré fabricadas.

Estas abordagens impossibilitam conclusões gerais ou super estimam os valores de performance dos métodos utilizados. Isso ocorre porque os conjuntos de treinamento e teste são geralmente extraídos de apenas uma imagem e portanto os objetos são coletados no mesmo tempo, sob as mesmas condições físicas e o mesmo sensor, de forma que qualquer fator exógeno é removido (SCHNEIDER; MURPHY; MELKUMYAN, 2014). Esta metodologia funciona mas, como colocado em seções anteriores, exige um esforço significativo já que requer a extração manual de conjunto de treinamento de imagens com grande quantidade de pixels, ademais torna a automação do processo de classificação impossível já que na prática é necessário classificar diversas imagens de diferentes locais e tempos distintos.

Nesse sentido o uso de bibliotecas espectrais permite que métodos sejam aplicados de forma consistente e automática. Somente assim é possível avaliar e tirar conclusões reais sobre a qualidade dos métodos (SCHNEIDER; MURPHY; MELKUMYAN, 2014).

Para tanto, o conjunto de dados utilizado afim de comparar os métodos propostos foram criados a partir da biblioteca espectral ASTER 2.0, composta por mais de 2400 curvas espectrais de materiais naturais e criados pelo homem (BALDRIDGE *et al.*, 2009). Dessa biblioteca foram selecionados apenas objetos das classes solo, vegetação e água, que apresentavam medições de reflectância para todos os comprimentos de onda registrados pelas bandas do satélite Landsat 8 com resolução espectral de 30m, Figura 3.1.

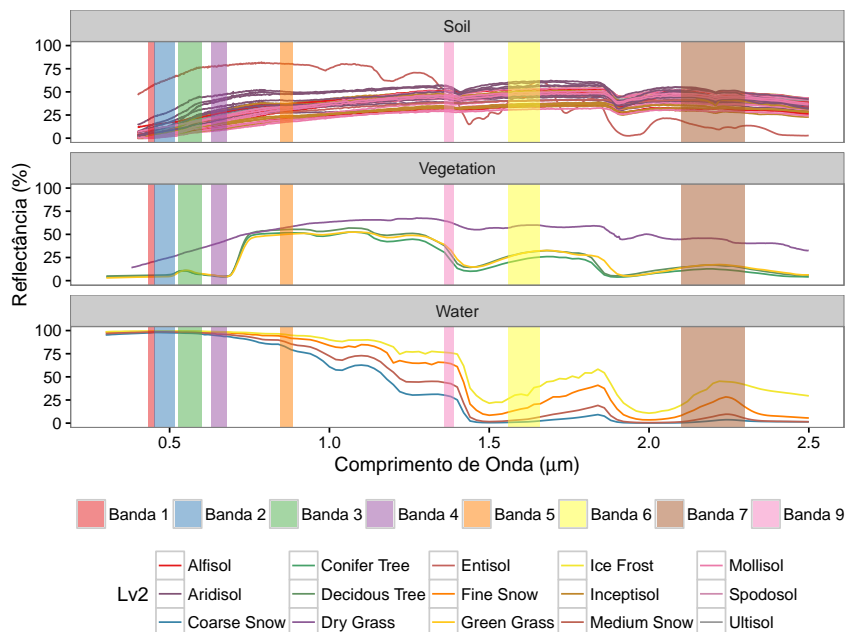


Figura 3.1 – Visualização dos dados utilizados para comparação dos métodos

As bandas termais 10 e 11 não foram utilizadas por recomendação do manual do usuário Landsat 8 (USGS; NASA, 2015), que sugere não usá-las para análises quantitativas devido a grande variabilidade da calibração do sensor e propõem ainda que caso seja feito o uso da banda 10, este deve ser em conjunto com um modelo atmosférico para estimar temperatura de brilho na superfície. A banda 8 não foi utilizada pois é necessário que todos os atributos sejam do mesmo tamanho e portanto seria necessário algum tipo de operação para agrupar os **pixels** com resolução de 15m em **pixels** de 30m ou então quebrar os **pixels** de 30m em **pixels** de 15m. Como este procedimento pode causar uma transformação indesejada dos valores oficiais registrados nas imagens e portanto levar a resultados incorretos optou-se por não usar a informação dessa banda.

Os dados foram rotulados em 2 níveis sendo que o primeiro representa as classes de informação em um contexto mais amplo, enquanto o segundo nível apresenta subcategorias do nível 1, como pode-se ver na Tabela 3.1.

Optou-se por utilizar mais de um nível de classificação para testar a capacidade dos métodos distinguirem classes cada vez mais refinadas e com isso verificar como se comportam a medida que se aumenta o grau de complexidade do problema.

Uma vez realizado esse pré-processamento de seleção e organização da biblioteca espectral foi feita uma amostragem aleatória de 230 mil observações. Essa amostra foi separada de forma aleatória em dois conjuntos independentes, um para o treinamento dos métodos e outro para testá-los. Para verificar a capacidade de aprendizado e generalização dos métodos, estes dois conjuntos foram separados considerando-se as seguintes proporções para a parte de treinamento: 10%, 30%, 50%, 70%, e 90%.

Tabela 3.1 – Classes dos dados a diferentes níveis

<i>Level 1</i>	<i>Level 2</i>
<i>Vegetation</i>	<i>Green Grass</i>
	<i>Dry Grass</i>
	<i>Conifer Tree</i>
	<i>Deciduous Tree</i>
<i>Water</i>	<i>Ice Frost</i>
	<i>Coarse Snow</i>
	<i>Medium Snow</i>
	<i>Fine Snow</i>
<i>Soil</i>	<i>Spodosol</i>
	<i>Ultisol</i>
	<i>Entisol</i>
	<i>Alfisol</i>
	<i>Inceptisol</i>
	<i>Aridisol</i>
	<i>Mollisol</i>

Com o intuito de verificar se a possível redução do esforço computacional pela análise de componentes principais pode prejudicar em algum sentido a classificação dos dados, o procedimento de classificação foi aplicado aos dados em sua forma original e sobre os componentes principais formados pela ACP.

Para aumentar a confiabilidade dos resultados cada um dos tratamentos foi repetido 5 vezes obtendo um total de 100 medições de acurácia, tempo de execução e índice kappa para cada um dos 3 métodos, como mostra a Figura 3.2.

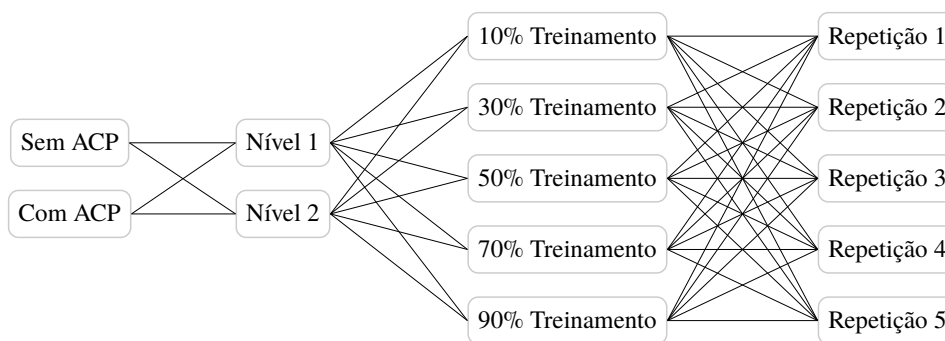


Figura 3.2 – Diagrama representando o delineamento do experimento

Acurácia é a medida de proporção de resultados classificados corretamente em relação ao total e pode ser calculado, como mostra a eq. (3.1), por meio da matriz de confundimento \mathbf{E} , também conhecida como matriz de erros ou tabela de contingência.

$$\mathbf{E} = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1k} \\ e_{21} & e_{22} & \cdots & e_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ e_{k1} & e_{k2} & \cdots & e_{kk} \end{bmatrix}$$

$$\pi_o = \frac{\text{tr}(\mathbf{E})}{\sum_{i=1}^k \sum_{j=1}^k e_{ij}} \quad (3.1)$$

Em que e_{ij} é quantidade de **pixels** predita como pertencentes a classe i sendo que pertencem realmente a classe j . Portanto o caso ótimo, em que o classificador agrupa os dados de forma perfeita, seria um matriz de confundimento diagonal e conseqüentemente com acurácia igual a 1.

O valor da acurácia observada pode levar a uma conclusão equivocada sobre a qualidade da classificação. Um método que tem uma acurácia observada igual a 0.8 quando a acurácia esperada é igual a 0.75, caso os dados fossem classificados de forma aleatória, pode ser menos interessante em relação a um método que apresente acurácia observada igual a 0.5 sendo que a esperada é igual a 0.05, por exemplo.

O índice kappa, dado pela eq. (3.5), busca corrigir esse problema e se torna uma medida menos ilusória uma vez que não depende de nenhum outro valor para ser interpretado.

Seja o total marginal da j -ésima coluna da matriz de confundimento denotado pela eq. (3.2) e o total marginal da i -ésima linha dado pela eq. (3.3) então a acurácia esperada pode ser calculada de acordo com a eq. (3.4).

$$e_{+j} = \sum_{i=1}^k e_{ij} \quad (3.2)$$

$$e_{i+} = \sum_{j=1}^k e_{ij} \quad (3.3)$$

$$\pi_e = \frac{\sum_{i=1}^k e_{+i}e_{i+}}{\left(\sum_{i=1}^k \sum_{j=1}^k e_{ij} \right)^2} \quad (3.4)$$

$$\kappa = \frac{\pi_o - \pi_e}{1 - \pi_e} \quad (3.5)$$

Foram calculados intervalos de confiança via **bootstrap** não paramétrico para as médias dos tratamentos considerando-se um nível de significância igual a 5%. As médias foram então comparadas de maneira que se há interseção entre intervalos de confiança para as médias de dois tratamentos, então estas podem ser consideradas iguais.

A técnica de **bootstrap** é muito comum para avaliar a variabilidade de uma estatística. Teoricamente é possível calcular o erro padrão e o intervalo de confiança para qualquer estatística gerada a partir dos dados entretanto as equações para determinadas medidas podem não estar disponíveis ou depender da distribuição dos dados, que pode ser desconhecida, ou ainda ser muito difícil de determinar (PENG, 2016).

O **bootstrap** não paramétrico é uma técnica independente de modelo ou distribuição e baseia-se na re-amostragem dos dados com reposição, de forma que a estatística de interesse é calculada para cada uma das amostras. Com os r valores calculados obtêm-se uma densidade para a estatística de interesse e pode-se tomar os percentis 2.5% e 97.5% dessa densidade como meio de determinar o intervalo de confiança desejado (HASTIE; TIBSHIRANI; FRIEDMAN, 2008).

Os critérios de parada adotados para a rede GNG foram: erro menor que 0.001; número máximo de neurônios igual a 1% da quantidade total de dados; valor máximo de iterações igual 10 mil.

Para fins de aplicação os métodos foram utilizados para classificar uma imagem real da região que corresponde a divisa entre o Paraná e o Paraguai compreendida entre as coordenadas presentes na Tabela 3.2 e registrada no dia 09 de Agosto de 2015.

Tabela 3.2 – Coordenadas da região de estudo

Extremo	Latitude (Graus)	Longitude (Graus)
Superior Esquerda	-23.50323	-55.73266
Superior Direita	-23.46791	-53.44905
Inferior Esquerda	-25.63777	-55.71100
Inferior Direita	-25.59881	-53.38851

As imagens foram coletas por meio do catálogo da divisão de geração de imagens¹ do Instituto Nacional de Pesquisas Espaciais (INPE) e os níveis de cinza foram convertidos para reflectância de acordo com as regras contidas no manual do usuário do Landsat 8 (USGS; NASA, 2015).

Os códigos foram implementados utilizando o software R, (R Core Team, 2015), e compilados em uma máquina virtual Linux com memória RAM de 12GB e 7 núcleos de processamento Intel(R) Xeon(R) CPU X5570, cada um com clock de 2.93GHz.

¹ <http://www.dgi.inpe.br/catalogo/>

4 RESULTADOS E DISCUSSÕES

4.1 Biblioteca Espectral

A Figura 4.1 apresenta um gráfico de dispersão dos dados provenientes da biblioteca espectral. Conclusões a respeito da separabilidade das classes e suas distribuições geralmente são o objetivo ao analisar um gráfico desses. Nesse sentido pode-se argumentar, por exemplo, que existe uma grande sobreposição das classes de vegetação e solo, enquanto que a classe água encontra-se mais afastada destas, entretanto deve-se tomar cuidado com esse tipo de análise uma vez que estão sendo observadas projeções bidimensionais do espaço de atributos e não necessariamente essas colocações se mantêm verdadeiras no espaço original.

Uma solução para o problema acima encontra-se em reduzir a dimensão do espaço via análise de componentes principais. A Tabela 4.1 apresenta o desvio padrão de cada um dos componentes principais e, pelo critério de Kaiser, seria necessário utilizar os 2 primeiros componentes para reduzir a dimensionalidade dos dados sem que haja nenhuma perda significativa de informação. O mesmo se observa pela proporção acumulada da variância e conclui-se que os dois primeiros componentes explicam cerca de 95% da variabilidade total dos dados, sendo assim suficientes para representá-los.

Tabela 4.1 – Variância explicada pelos componentes principais

	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8
Desvio padrão	2.3471	1.4278	0.4975	0.3645	0.2130	0.1325	0.0884	0.0400
Proporção da variância	0.6886	0.2548	0.0309	0.0166	0.0057	0.0022	0.0010	0.0002
Proporção acumulada	0.6886	0.9434	0.9744	0.9910	0.9966	0.9988	0.9998	1.0000

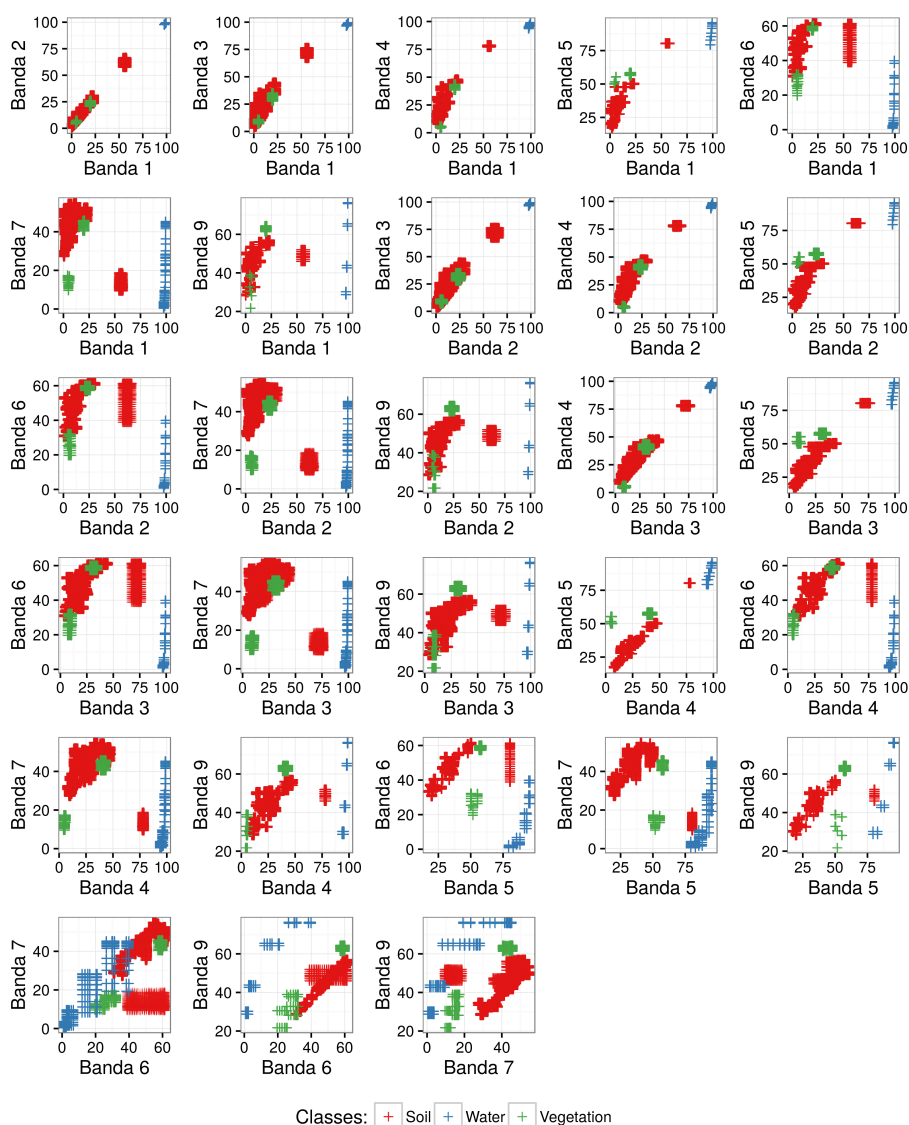
A Figura 4.4 representa os dados no novo espaço de atributos em que pode-se tirar conclusões mais acertadas uma vez que os pontos se encontram no plano bidimensional e, portanto, mais fácil de se interpretar os resultados. Com respeito ao nível 1, todos os pontos da classe água ainda estão separados das demais classes enquanto alguns pontos de vegetação se misturam aos pontos de solo. No nível 2 percebe-se uma mistura ainda maior entre as classes, ademais, ao se atentar às elipses da distribuição normal de cada classe, é possível verificar a influência que pontos atípicos podem ter sobre os parâmetros da distribuição e como isso pode atrapalhar a precisão de métodos paramétricos.

A densidade dos dados originais e transformados estão representadas na Figura 4.2 e na Figura 4.5, respectivamente. Nelas é possível ver que a distribuição dos dados apresentam características que não condizem com a distribuição normal, ou seja, apresentam multimodalidade, assimetria e caudas pesadas.

Os **boxplots**, Figura 4.3 e Figura 4.6, mostram a presença de diversos pontos atípicos,

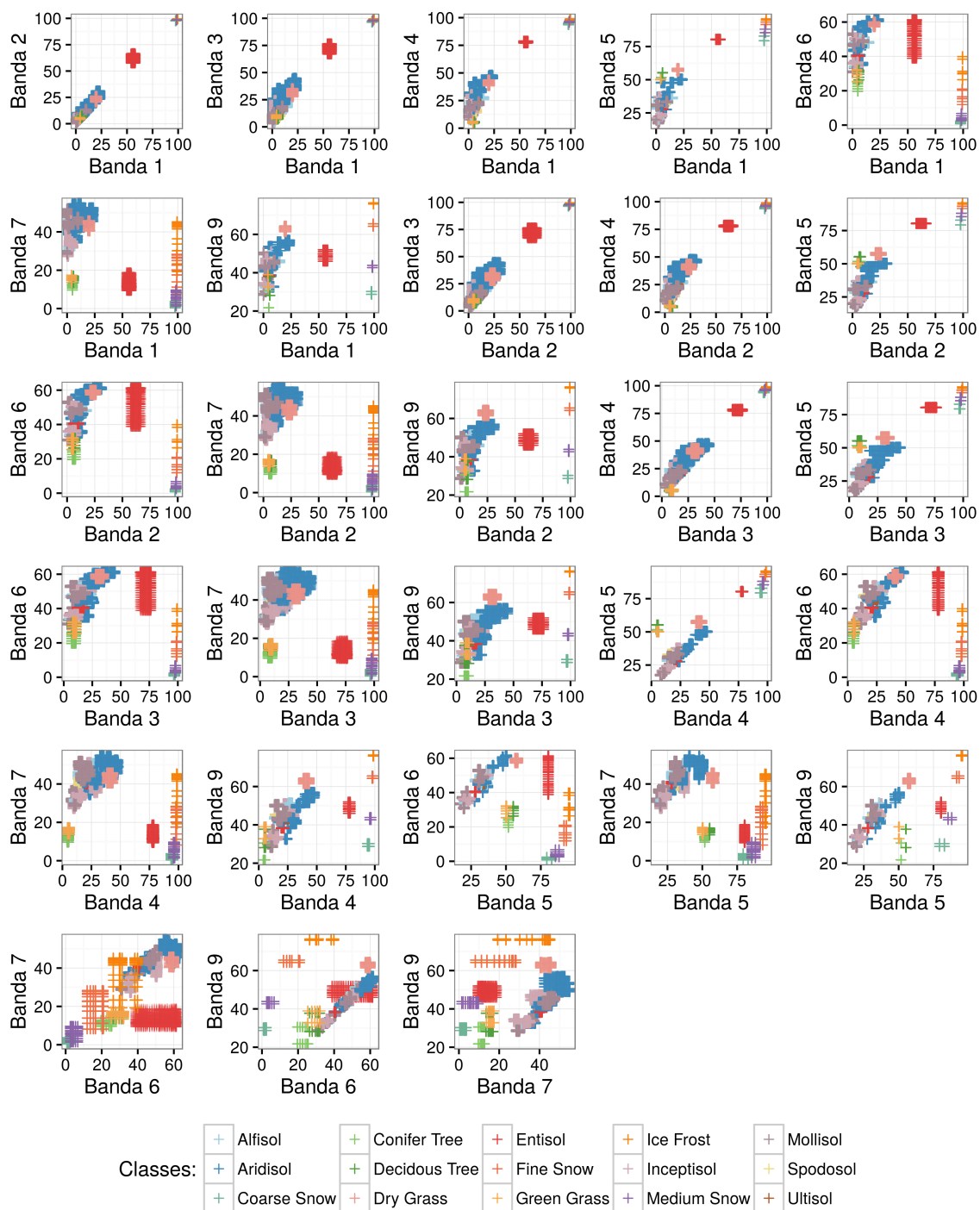
isto é, pontos distantes dos quartis mais extremos de sua distribuição e ainda ajudam a evidenciar a assimetria das classes com relações a determinadas bandas espectrais, pois para diversos casos a mediana encontra-se deslocada em direção ao primeiro ou ao terceiro quartil, pode-se citar os seguinte exemplos: Na Figura 4.3a este comportamento é observado mais claramente na classe de vegetação para as bandas 1, 2, 3, 4 e 7; na Figura 4.3b para a classe **entisol** banda 6, classe **inceptisol** bandas 4 a 9, entre outros.

Esta análise descritiva leva a crer que as pressuposições exigidas pelo método de classificação por máxima verossimilhança (distribuição normal multivariada no espaço de atributos para cada uma das classes) não são satisfeitas o que teoricamente inviabilizaria o método.



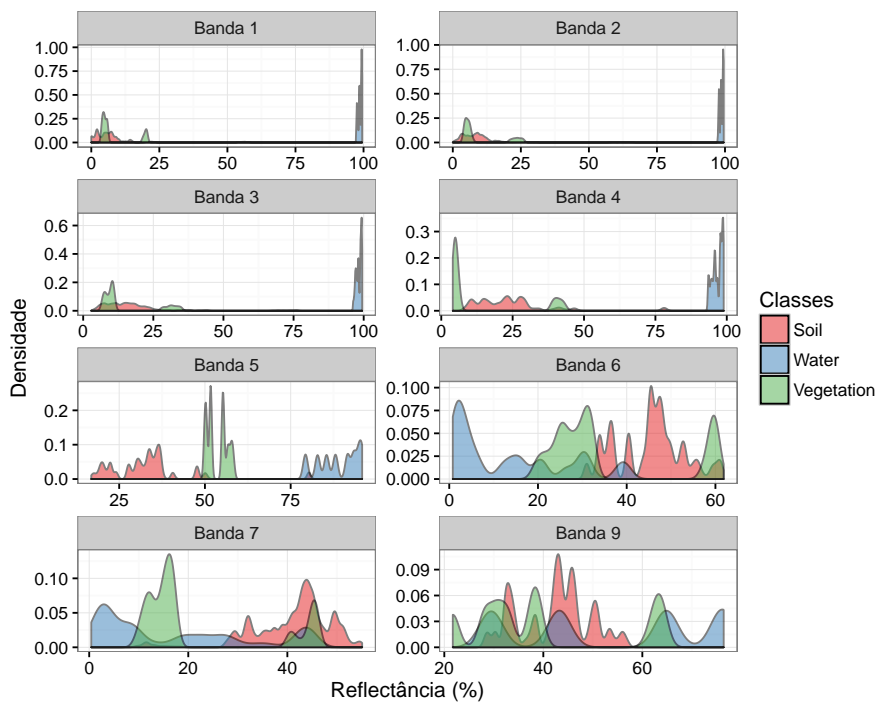
(a) Nível 1

Figura 4.1 – Gráfico de dispersão dos pontos representados no espaço bidimensional formado pelos pares de bandas (continua)

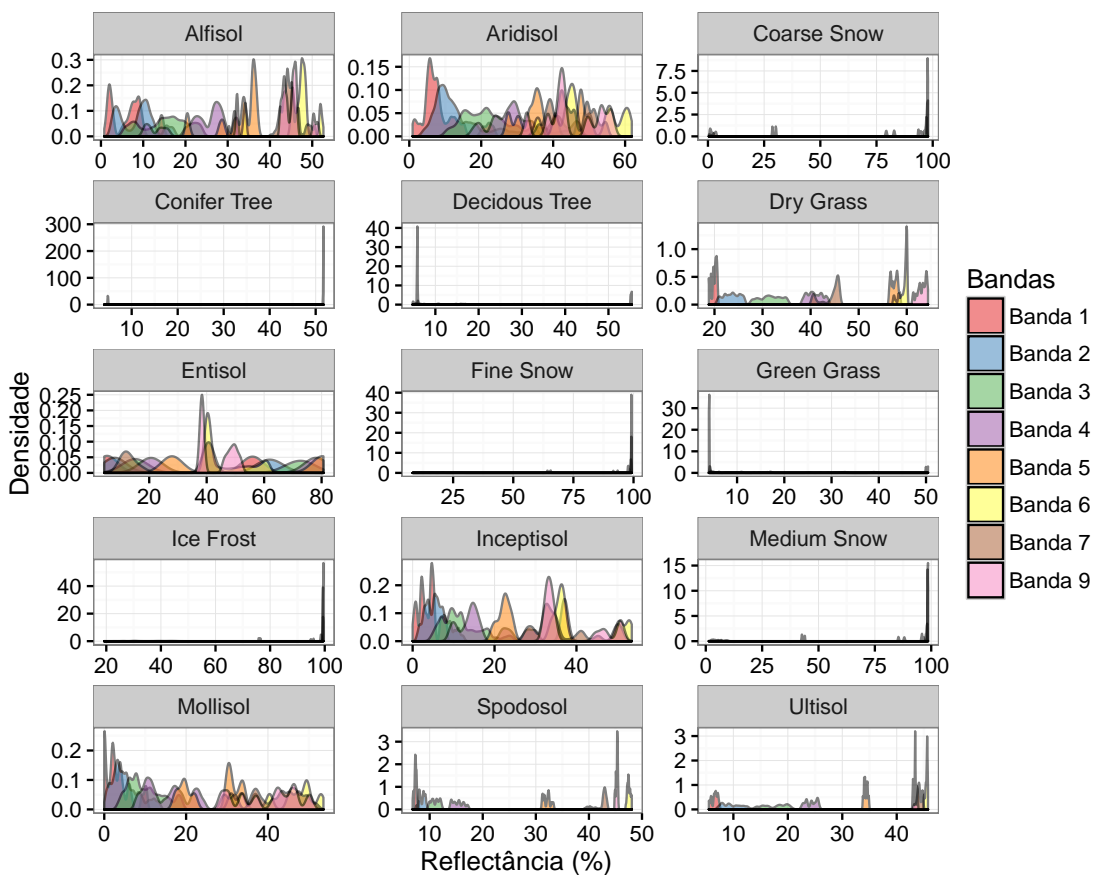


(b) Nível 2

Figura 4.1 – Gráfico de dispersão dos pontos representados no espaço bidimensional formado pelos pares de bandas (conclusão)

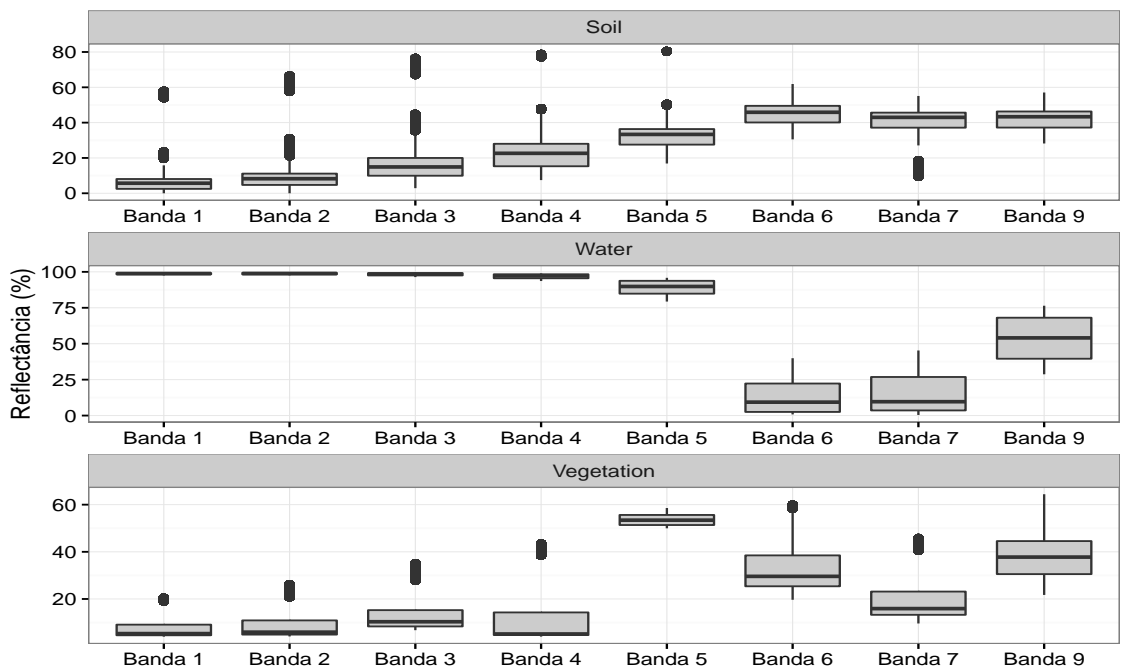


(a) Nível 1

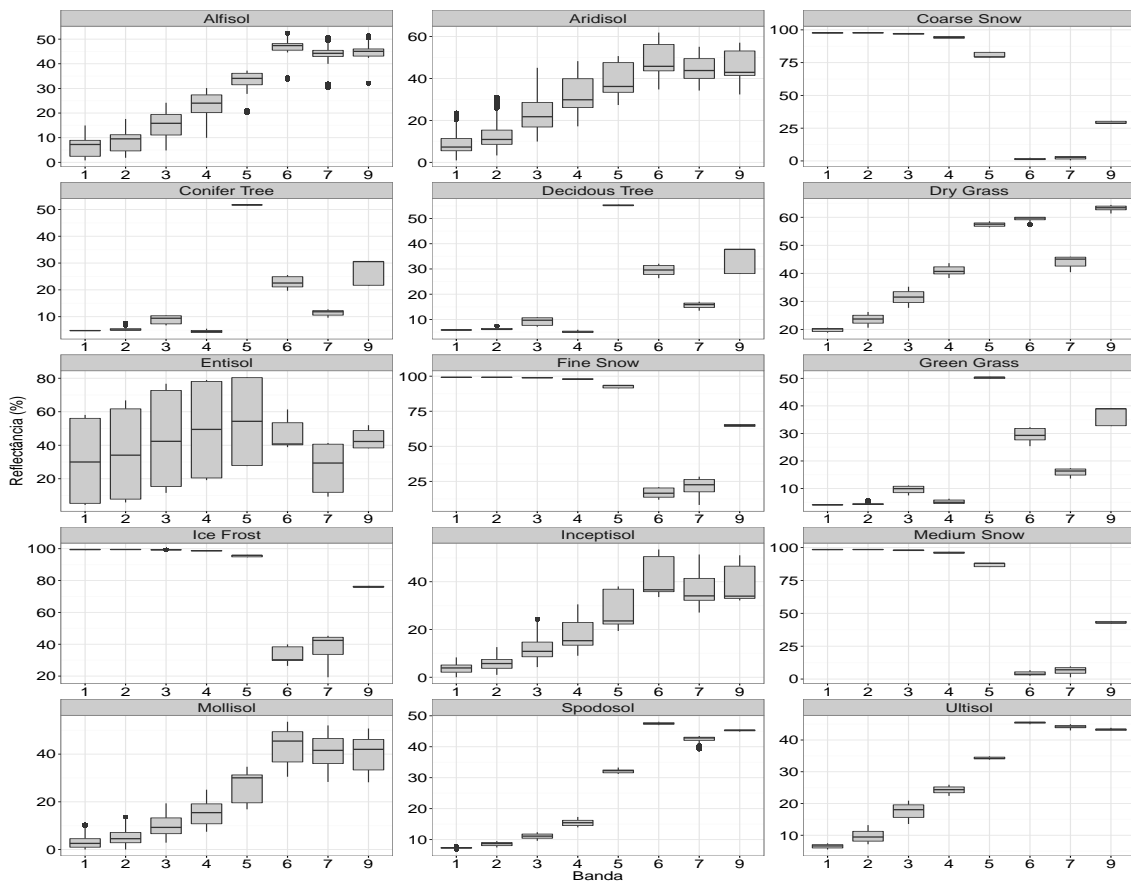


(b) Nível 2

Figura 4.2 – Densidade dos pontos para todas as bandas e classes

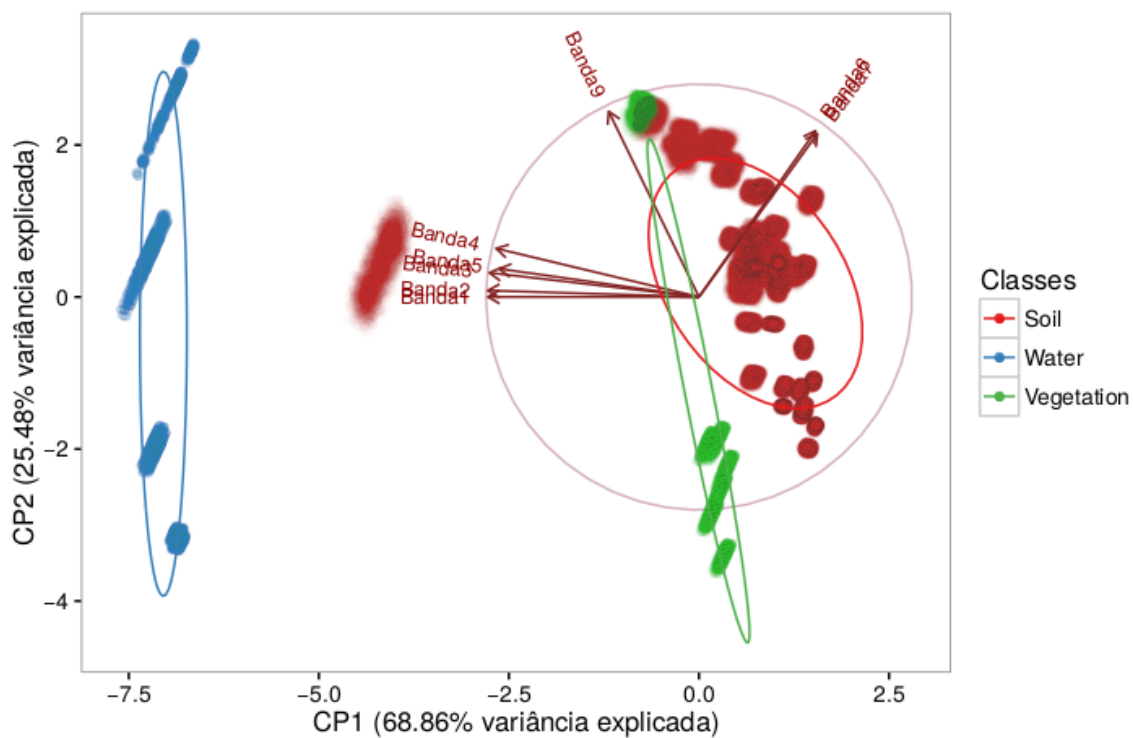


(a) Nível 1

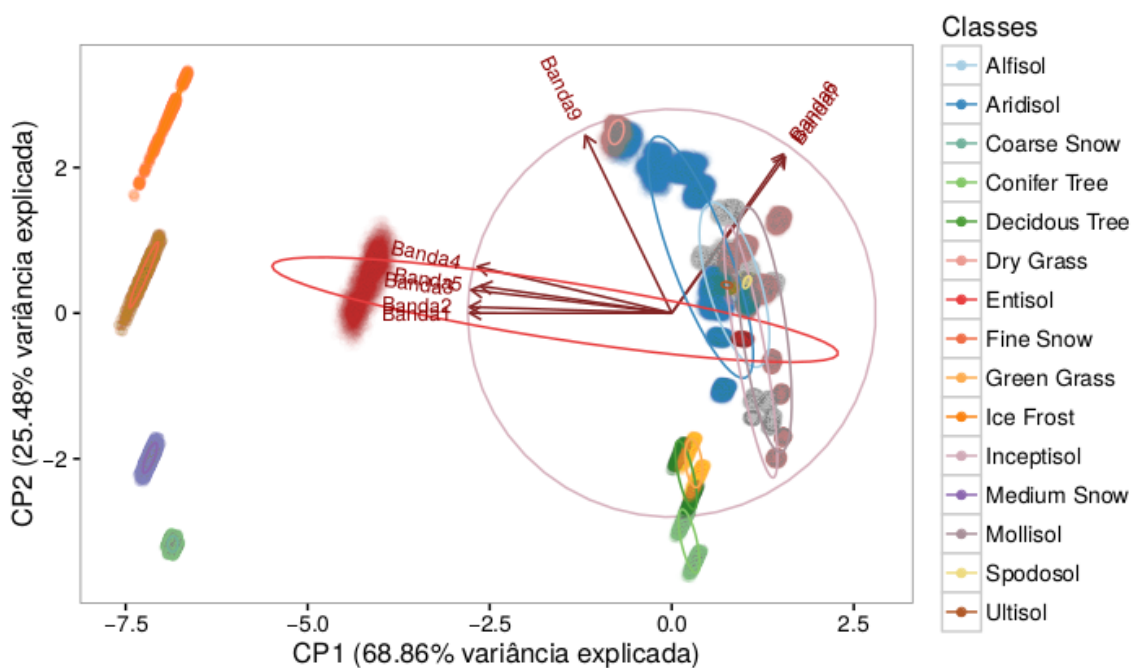


(b) Nível 2

Figura 4.3 – Boxplot dos pontos para todas as bandas e classes

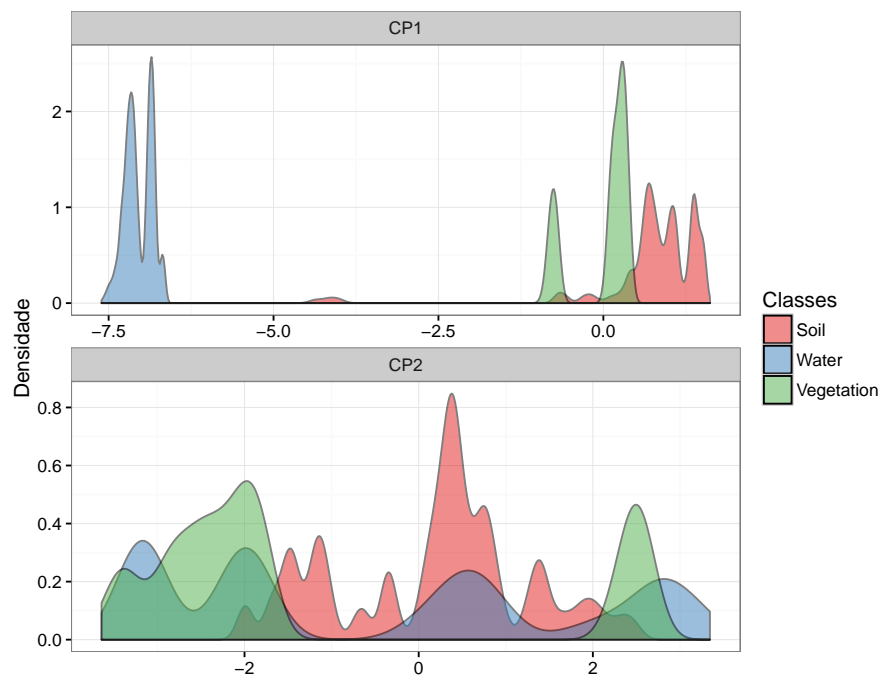


(a) Nível 1

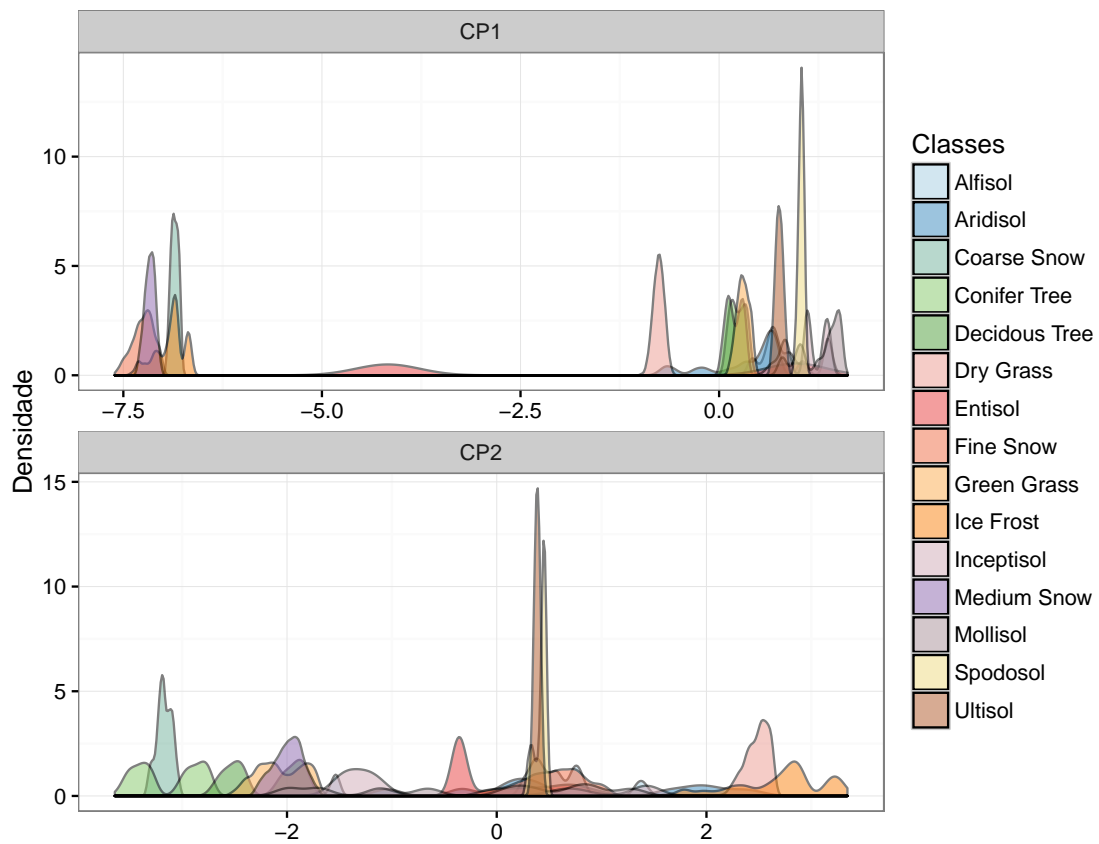


(b) Nível 2

Figura 4.4 – Biplot dos pontos transformados pela análise de componentes principais

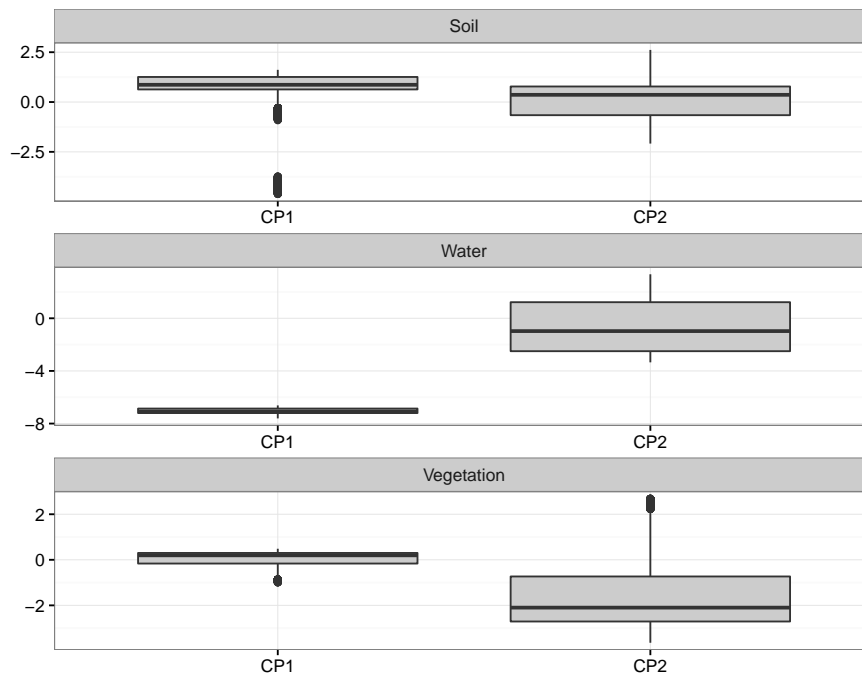


(a) Nível 1

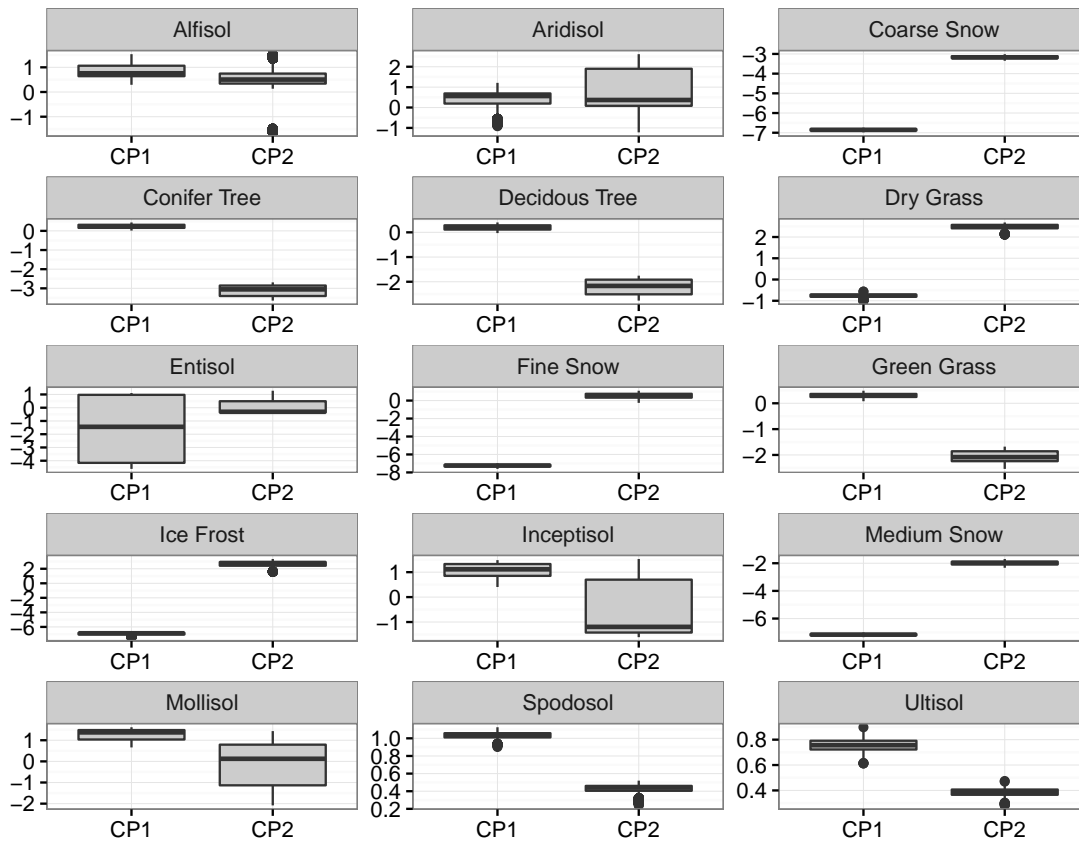


(b) Nível 2

Figura 4.5 – Densidade dos pontos para os componentes principais e classes



(a) Nível 1



(b) Nível 2

Figura 4.6 – Boxplot dos pontos para os componentes principais e classes

A Figura 4.7 mostra o índice kappa obtido para cada uma das repetições. Nota-se pela pouca variabilidade das repetições que os métodos são bastante estáveis, o que explica a pequena amplitude dos intervalos de confiança para as médias, observadas na Figura 4.8.

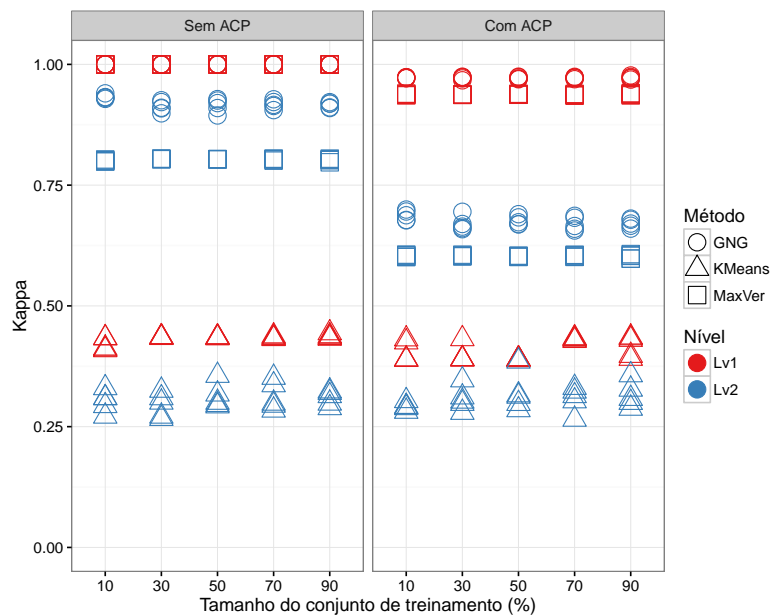


Figura 4.7 – Índice kappa obtido pela análise

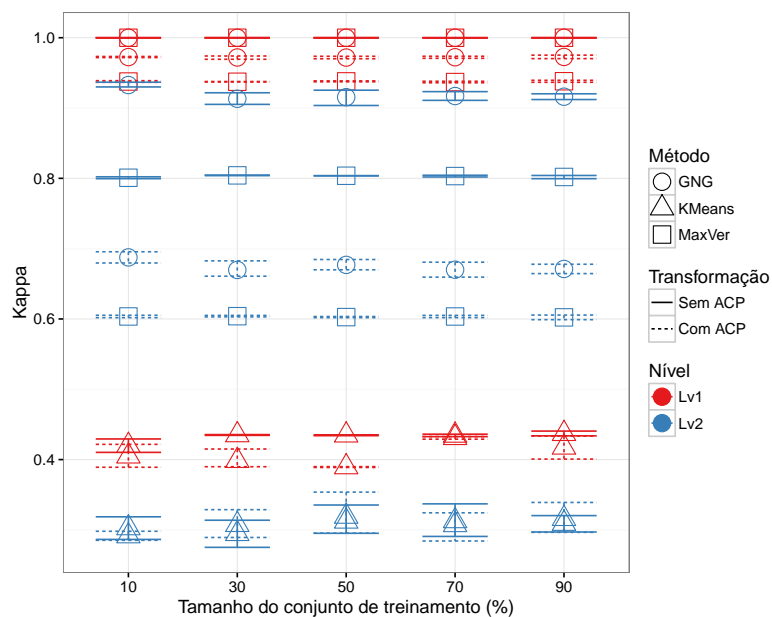


Figura 4.8 – Média do índice kappa com intervalo de confiança para os tratamentos

Pela comparação das médias do índice kappa e seus intervalos de confiança é possível verificar que o único fator que não exprime nenhuma influência sobre as análises é o tamanho do conjunto de treinamento dos métodos, para todos os outros existe uma diferença significativa nas médias.

A performance do método **k-means** é inferior aos demais para todos os casos. A rede neural artificial por outro lado apresenta um desempenho superior ao método de classificação por máxima verossimilhança, exceto para a classificação dos dados no nível 1 sem redução de dimensionalidade, em que os resultados dos dois métodos são similares.

Com exceção do **k-means**, o desempenho dos métodos cai quando aplicado a transformação de componentes principais e essa diminuição é mais expressiva quando aplicados ao nível com maior número de classes, nível 2. De forma análoga, todos os métodos apresentam resultados superiores no nível 1 quando comparados com o nível 2.

Pode-se perceber pela Figura 4.9 que a variabilidade do tempo de execução dos métodos é um pouco maior e consecutivamente a amplitude dos intervalos de confiança também aumenta, Figura 4.10.

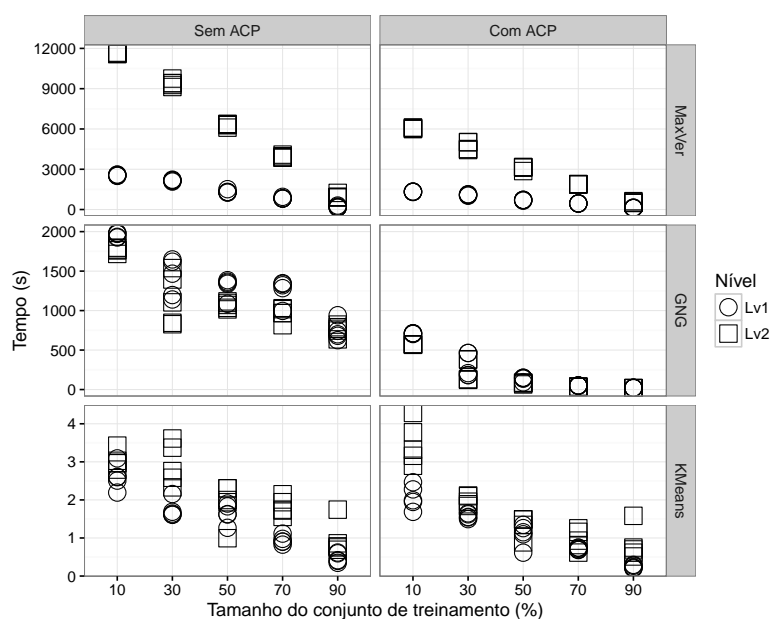


Figura 4.9 – Tempo de execução dos métodos

Para o tempo de execução o tamanho dos conjuntos de treinamento e teste tem um papel importante, pois a medida que se aumenta a proporção de pontos no conjunto de treinamento e, consecutivamente, diminui-se os elementos do conjunto de teste, os métodos tendem a executarem de forma mais rápida, principalmente o método da máxima verossimilhança que, no caso mais extremo, apresenta uma queda de mais de 3 horas, o que indica que a parte de teste dos métodos exige maior esforço computacional.

Apesar do baixo desempenho do algoritmo **k-means**, este é extremamente rápido, demorando poucos segundos para concluir sua execução, enquanto a RNA demora minutos e o método de classificação por máxima verossimilhança pode demorar horas.

Para o método **k-means** o único fator relevante para a pequena diminuição de tempo

de execução é o tamanho dos conjuntos de treinamento e teste, de forma que não há mudança significativa entre os tempos para diferentes níveis ou para os dados com e sem transformação. Já para a rede GNG existe uma distinção clara no tempo quando aplicada a transformação, sendo que o algoritmo executa aproximadamente 3 vezes mais rápido com a redução de dimensão e essa taxa se mantém constante para todos os tamanhos de amostra. Também para a rede GNG existe uma pequena diferença entre as médias dos níveis 1 e 2 para alguns dos tamanhos estudados.

Para o método de classificação por máxima verossimilhança todos os fatores geram mudanças expressivas. A diferença entre níveis, por exemplo, resulta aproximadamente em um tempo 4 vezes maior do nível 1 para o nível 2 no caso mais extremo, 90% dos elementos para conjunto de teste, e metade disso quando considerando os dados com e sem transformação.

Existe também uma diferença significativa para os tempos de execução entre os métodos de classificação sendo que, de forma geral, o método da máxima verossimilhança apresenta os piores tempos, seguido pelo GNG e por fim o **k-means**. O método de classificação por máxima verossimilhança supera a velocidade da rede neural artificial apenas no nível 1 quando se tem poucos elementos para se classificar, 30% ou menos de elementos no conjunto de teste, sendo a diferença entre eles inferior a 10 minutos.

Para informações mais específicas os valores médios e intervalos de confiança para o índice kappa e o tempo de execução dos métodos encontram-se no ANEXO A e ANEXO B, respectivamente.

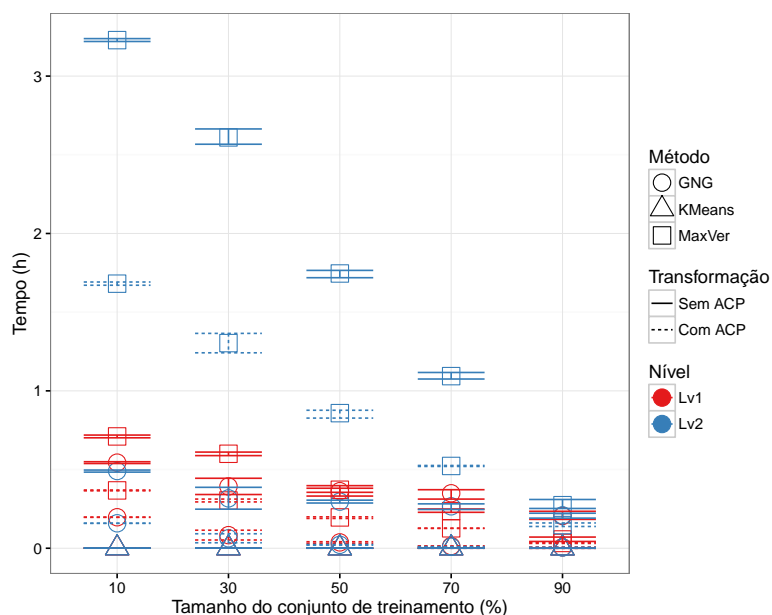


Figura 4.10 – Média do tempo de execução com intervalo de confiança para os tratamentos

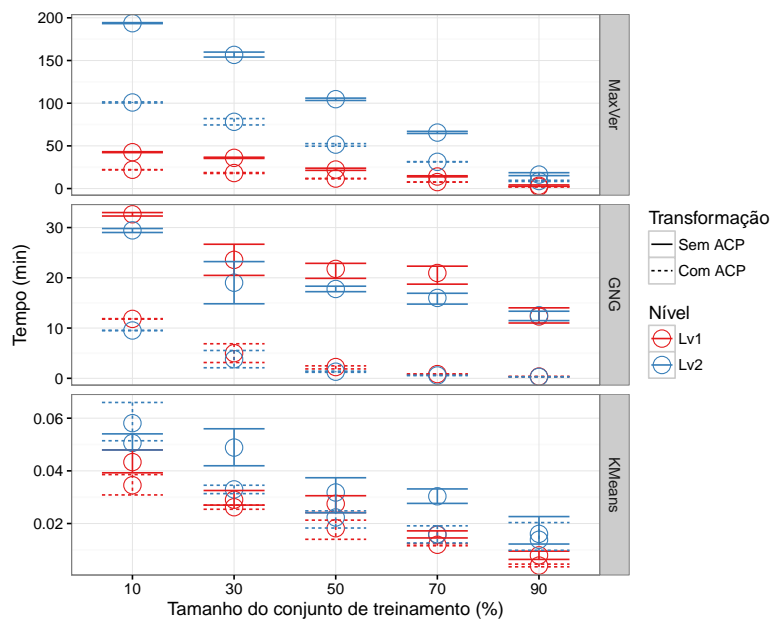
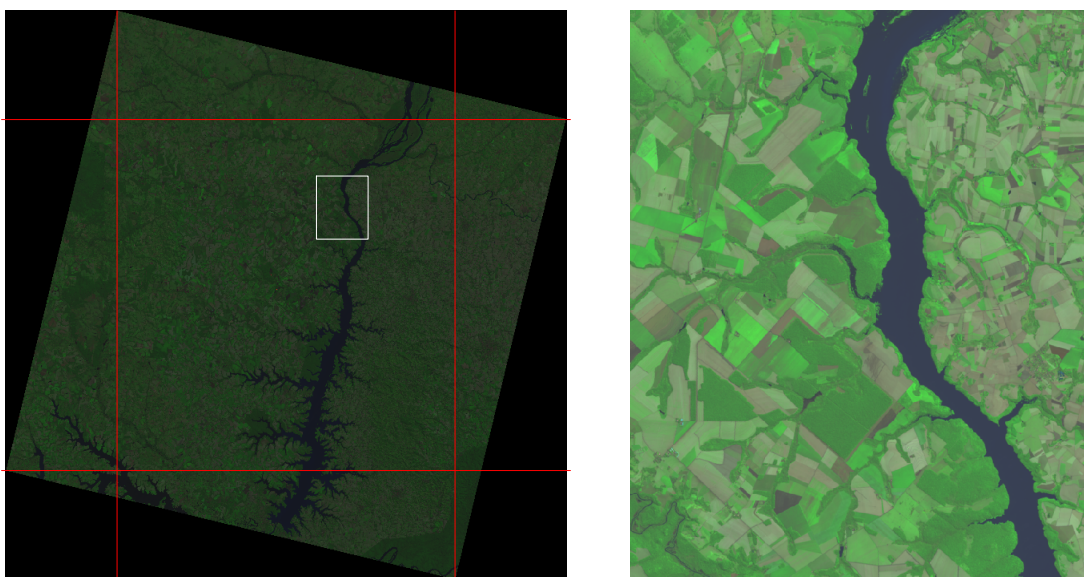


Figura 4.11 – Separação por método para melhor visualização dos tempos de execução

4.2 Imagem de Satélite

Por questões de capacidade do computador utilizado nas análises foi realizado um corte na imagem de satélite descrita na seção 3 e trabalhou-se apenas com a parte central da imagem, representado pelo quadrado vermelho no centro da Figura 4.12a. Para melhor visualização e comparação dos resultados também foi selecionada uma pequena área, localizada no interior do retângulo branco e apresentada na Figura 4.12b.



(a) Imagem completa

(b) Porção ampliada

Figura 4.12 – Área de estudo representada em falsa cor pelas bandas 7, 5 e 4

O conjunto de treinamento a ser utilizado pelo método de classificação por máxima verossimilhança foi criado de forma manual e visual a partir da imagem de estudo. Este procedimento realizado por meio da busca seletiva de conjuntos de **pixels**, a partir da ampliação de partes da imagem, que representassem os seguintes rótulos: solo, vegetação e água.

A partir da análise descritiva apresentada na Figura 4.13 pode-se notar pela superposição de pontos (Figura 4.13a) e pela intersecção das classes pré-definidas no conjunto de treinamento (Figura 4.13b) que não existe nenhum padrão visivelmente distinto.

Nota-se também a quantidade elevada de pontos atípicos pela Figura 4.13d e pela Figura 4.13f. Além destes pontos percebe-se que as curvas de densidade, além de apresentarem curtose e assimetria elevadas, indicam pela multimodalidade a possibilidade de existirem mais grupos na imagem além dos 3 sugeridos: vegetação, solo e água. Essa observação é particularmente importante quando se trabalha com uma imagem de satélite pois quase que certamente, dado o tamanho da área presente, existem mais fatores a serem considerados, como por exemplo solos com diferentes umidades; solo com e sem vegetação; vegetação em diferentes estágios de crescimento e diferentes culturas. Portanto é razoável supor que apenas as 3 classes propostas não definem todo o ambiente de forma satisfatória e mais classes podem ser exploradas.

A Figura 4.15 e Figura 4.16 apresentam a imagem classificada por cada um dos três métodos apresentados neste trabalho com e sem o uso de análise de componente principal: rede neural artificial GNG, máxima verossimilhança e **k-means**.

Uma primeira conclusão que observa-se desses resultados é que a análise de componentes principais, apesar de agilizar o tempo de execução dos métodos, gera imagens mais granulares para os métodos não supervisionados. Apesar dessa conclusão não ser verdadeira para o método de classificação por máxima verossimilhança, é possível notar que a maior uniformidade presente na imagem classificada utilizando-se a análise de componentes principais ocorre pela dificuldade em definir regiões de fronteira ou com mistura de classes.

Comparando-se os métodos visualmente, pode-se perceber que a rede GNG gera imagens mais uniformes e com fronteiras mais bem definidas do que o método **k-means**, que mesmo para um objeto bastante uniforme como o rio Paraná e o Lago Itaipu gera resultados bem inconsistentes.

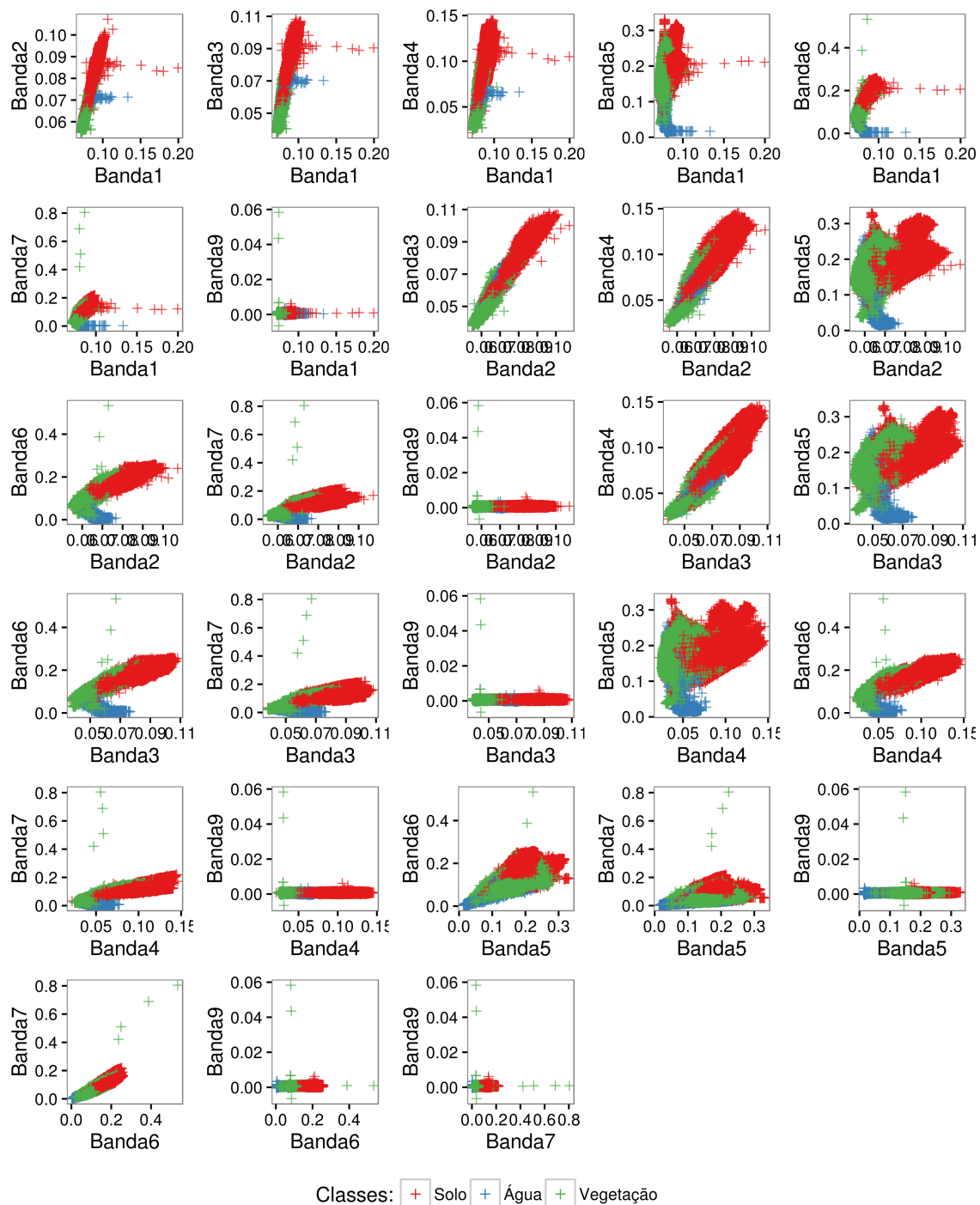
A princípio, quando comparadas as imagens classificadas pela rede neural e pelo método da máxima verossimilhança (Figura 4.15) com a imagem completa (Figura 4.12a), é possível pensar que a classificação por verossimilhança apresenta resultados mais condizentes com a realidade. Isso ocorre por conta da inabilidade do olho humano de enxergar um amplo espectro de cores quando dispostas em pequenas parcelas próximas umas das outras. Contudo, quando observa-se mais atentamente uma porção menor da imagem numa escala mais próxima (Figura 4.12b e Figura 4.16), verifica-se que o método de classificação por máxima verossimilhança agrega todos os diferentes padrões existentes na imagem nos 3 rótulos estabelecidos pelo conjunto de

treinamento enquanto a rede GNG consegue distingui-las.



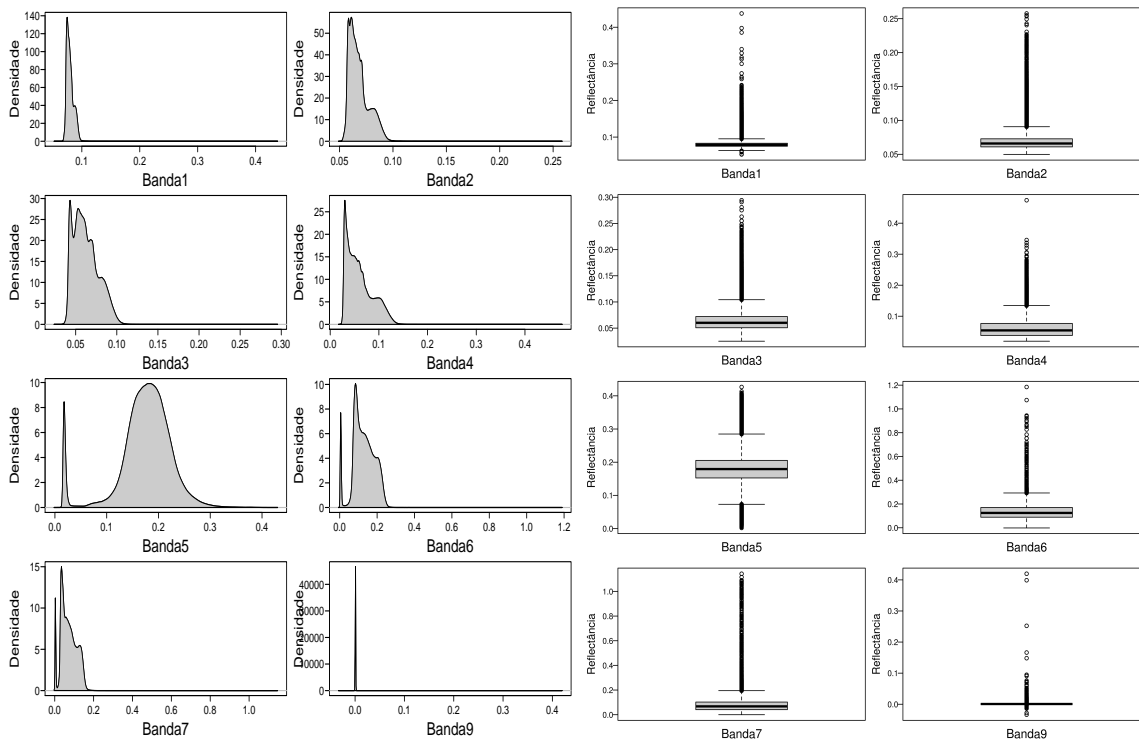
(a) Gráfico de dispersão dos pontos da imagem representados no espaço bidimensional formado pelos pares de bandas

Figura 4.13 – Análise exploratória dos pontos da imagem e do conjunto de treinamento (continua)



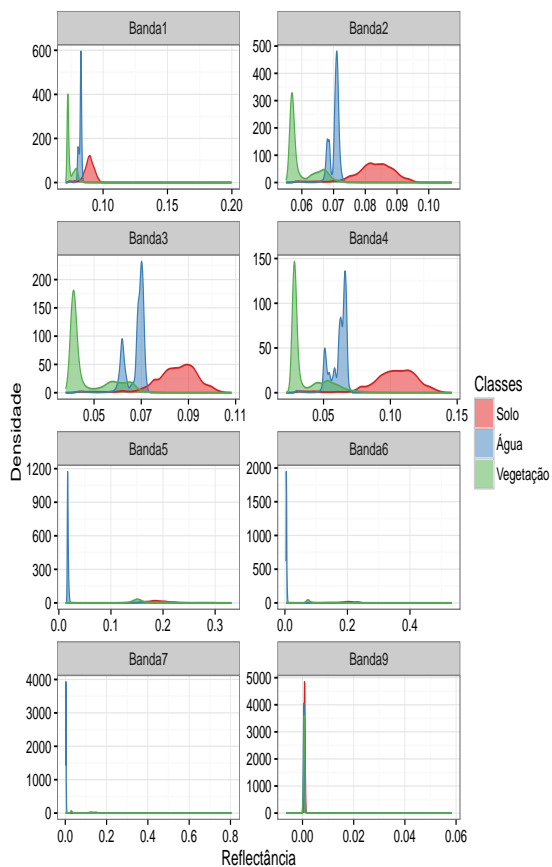
(b) Gráfico de dispersão do conjunto de treinamento representado no espaço bidimensional formado pelos pares de bandas

Figura 4.13 – Análise exploratória dos pontos da imagem e do conjunto de treinamento (continuação)

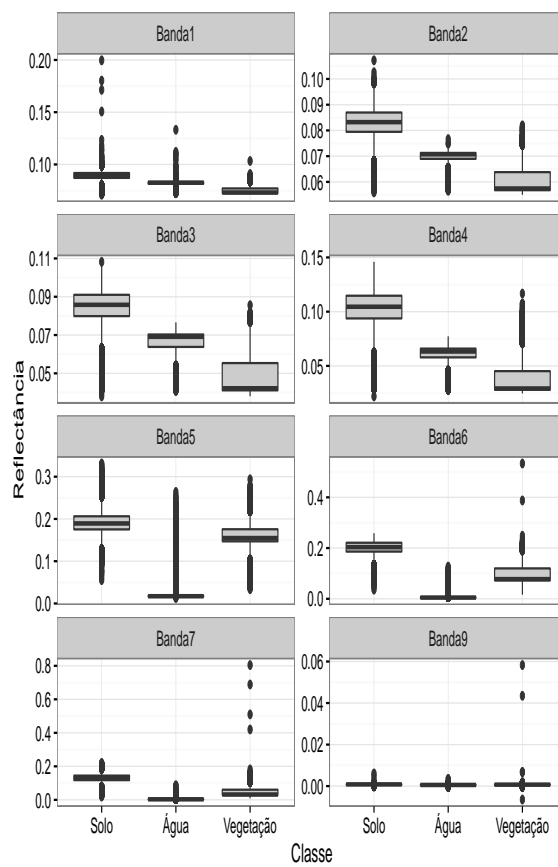


(c) Densidade dos pontos da imagem

(d) Boxplot dos pontos da imagem

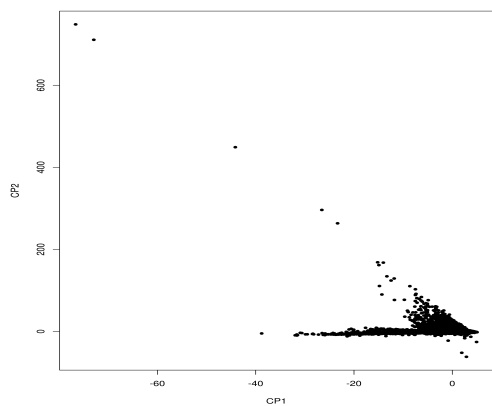


(e) Densidade do conjunto de treinamento

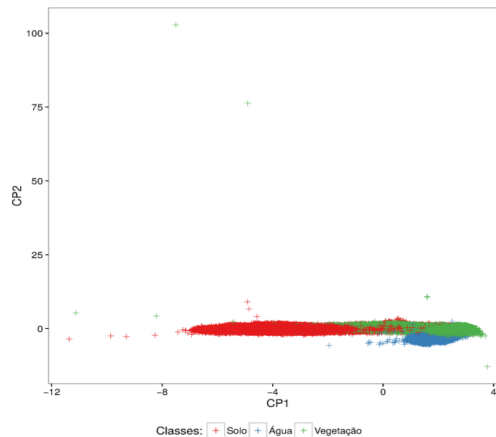


(f) Boxplot conjunto de treinamento

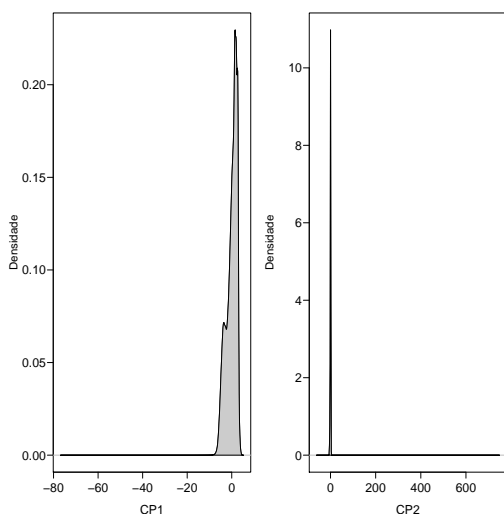
Figura 4.13 – Análise exploratória dos pontos da imagem e do conjunto de treinamento (conclusão)



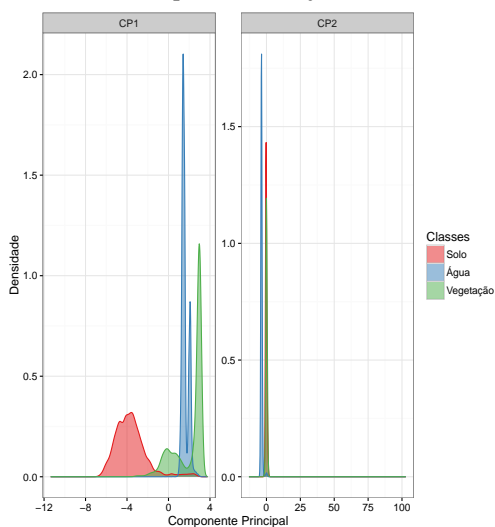
(a) Gráfico de dispersão dos pontos da imagem



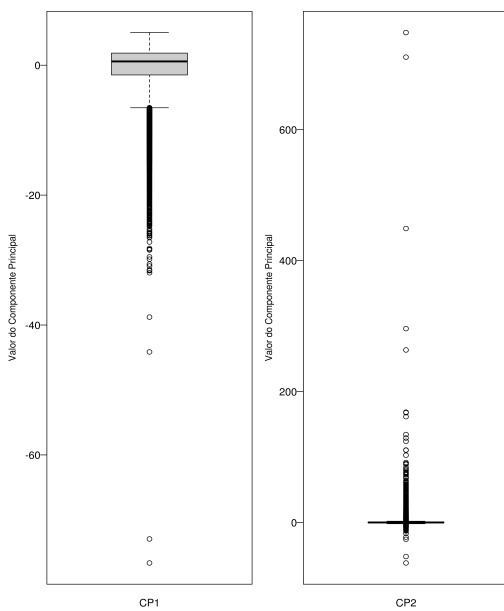
(b) Gráfico de dispersão do conjunto de treinamento



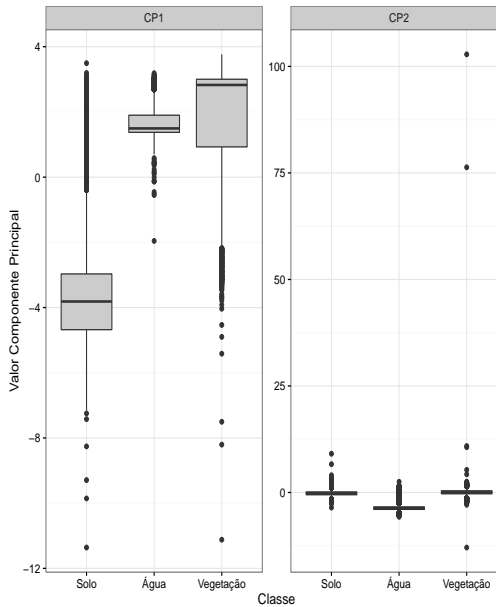
(c) Densidade dos pontos da imagem



(d) Densidade do conjunto de treinamento

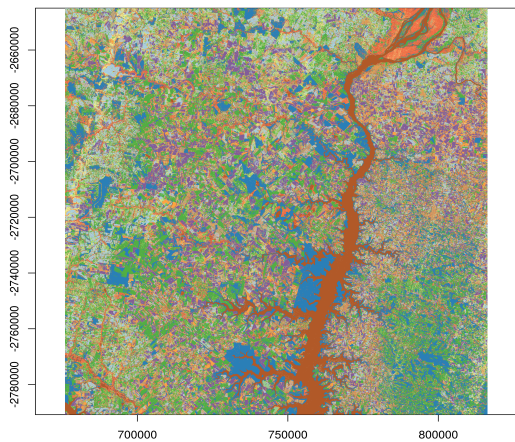


(e) Boxplot dos pontos da imagem

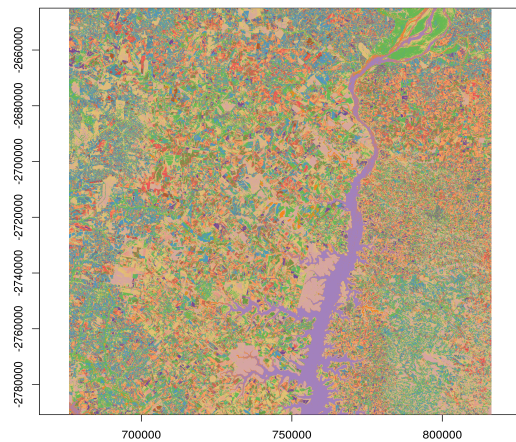


(f) Boxplot do conjunto de treinamento

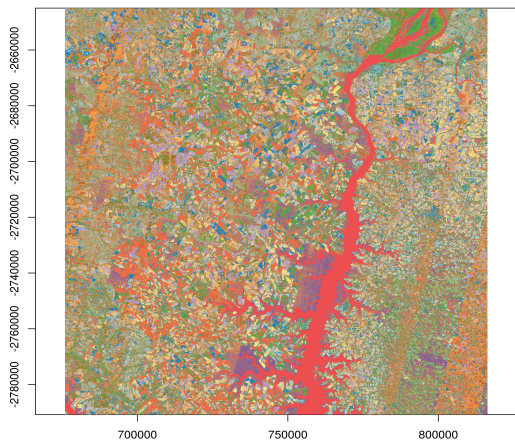
Figura 4.14 – Análise exploratória dos componentes principais da imagem e conjunto de treinamento



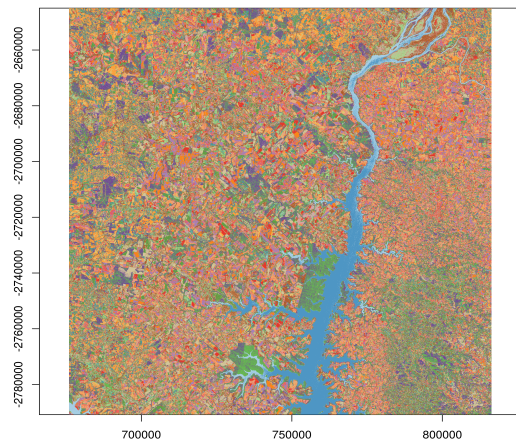
(a) **Growing neural gas** sem ACP



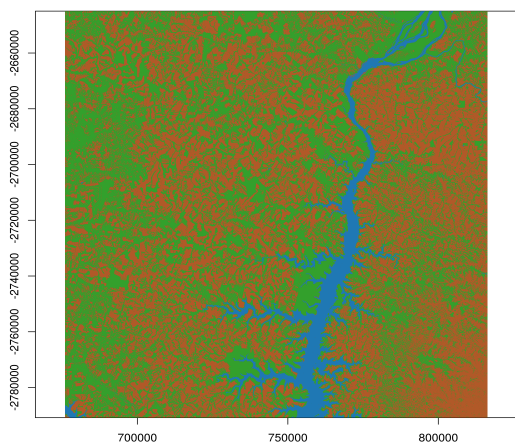
(b) **Growing neural gas** com ACP



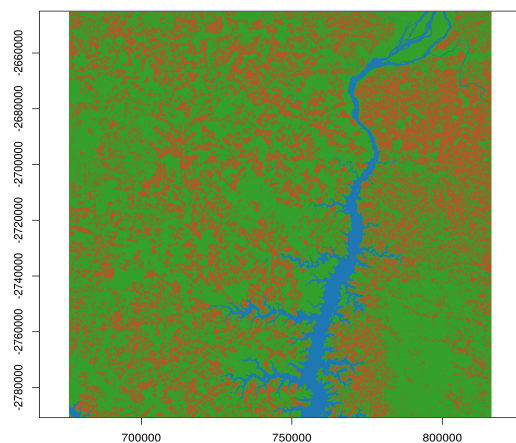
(c) **K-Means** sem ACP



(d) **K-Means** com ACP

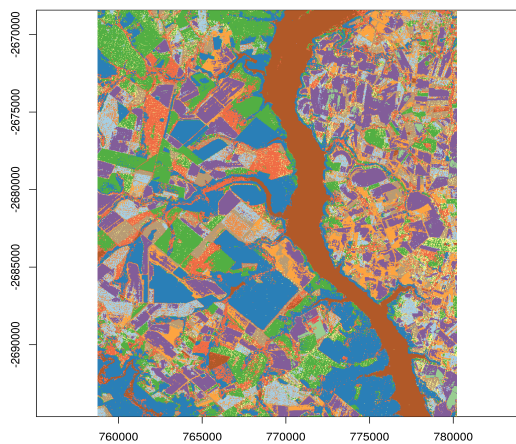


(e) **Máxima verossimilhança** sem ACP

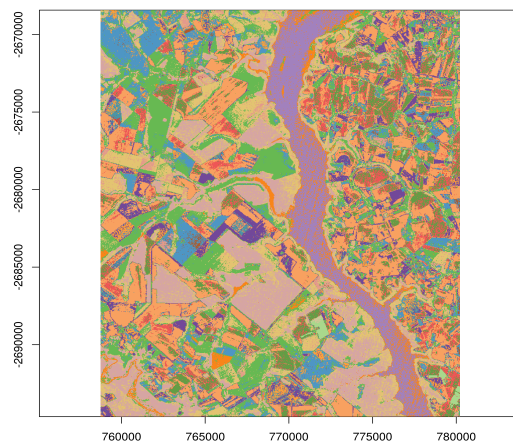


(f) **Máxima verossimilhança** com ACP

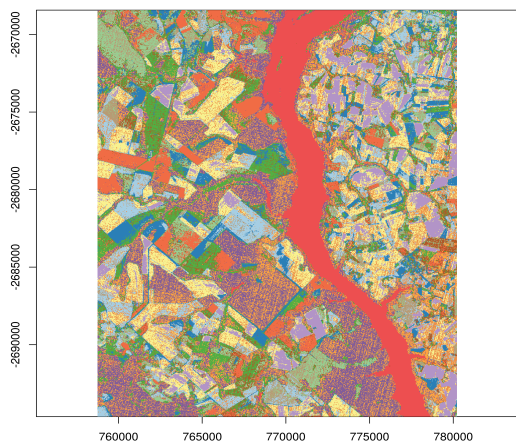
Figura 4.15 – Imagem classificada utilizando os métodos estudados com e sem análise de componentes principais



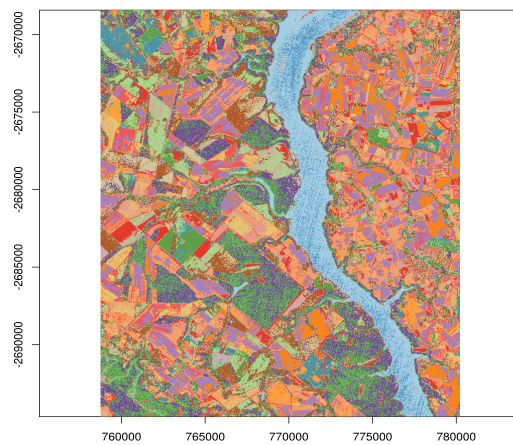
(a) Growing neural gas sem ACP



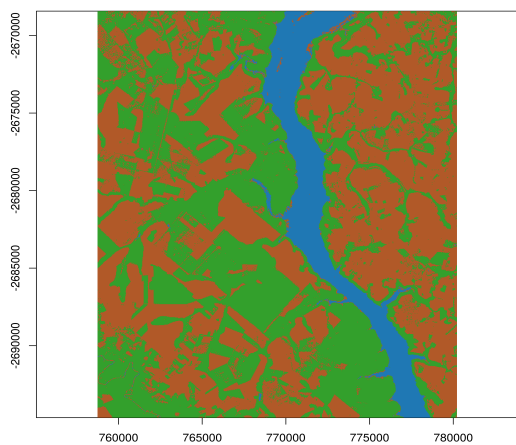
(b) Growing neural gas com ACP



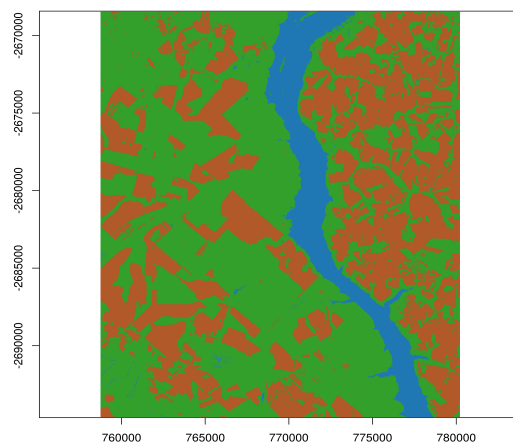
(c) K-Means sem ACP



(d) K-Means com ACP



(e) Máxima verossimilhança sem ACP



(f) Máxima verossimilhança com ACP

Figura 4.16 – Ampliação de parte da imagem

5 CONSIDERAÇÕES FINAIS

A classificação de imagens de satélite possui uma série de aplicações práticas importantes em diversos setores, porém os benefícios em utilizar esta ferramenta vem com o custo computacional necessário para gerar uma imagem informativa para o usuário final. Dessa forma é necessário investigar, da maneira mais fidedigna possível, métodos mais eficientes, com melhor desempenho, práticos e que permitam automatizar ao máximo os processos complexos.

O objeto deste trabalho é propor uma metodologia autônoma, acurada e eficiente para a classificação de imagens provenientes de sensores remotos. Para tanto, realizou-se a comparação entre uma abordagem de inteligencia artificial, até então pouco estudada nesse contexto, denominada **growing neural gas**, com as principais técnicas da área de sensoriamento remoto, denominadas método da máxima verossimilhança e **k-means**. Esta comparação levou em consideração aspectos qualitativos, como medidas de tempo de execução e acurácia, bem como vantagens e desvantagens metodológicas.

O estudo realizado neste trabalho mostrou, a partir de uma análise rigorosa dos resultados de classificação dos dados amostrados da biblioteca espectral ASTER 2.0 e dos dados provenientes de uma imagem real, que a rede neural artificial GNG apresentou resultados de forma mais rápida do que o método supervisionado, atualmente empregado em grande parte dos **softwares** de sensoriamento remoto, e ainda obteve um maior índice de acertos na classificação.

De acordo com os resultados obtidos, a rede GNG se comportou de forma mais robusta em relação aos demais métodos estudados, pois apresentou resultados mais consistentes a diferentes tratamentos e principalmente por ser independente de um conjunto de treinamento. Esse fator é de extrema importância pois: diminui custos, tempo gasto com pré-processamento de dados e possíveis erros humanos; além de aumentar a capacidade de automação, já que os padrões da imagem foram encontrados pelo próprio algoritmo ao invés de serem coletados para uma imagem específica e apresentados previamente ao método, como foi realizado ao se empregar a máxima verossimilhança. Mostrou-se também nesse trabalho um exemplo de aplicação real em que as pressuposições estabelecidas pelo método de classificação por máxima verossimilhança não foram satisfeitas, o que afeta a qualidade dos parâmetros e pode produzir resultados de baixa qualidade.

As conclusões supracitadas mostram a importância e a vantagem de se trabalhar com métodos não paramétricos e robustos como a rede GNG, que se mantém estável na presença de pontos discrepantes e não é afetada por erros de modelagem. Entretanto, a simplicidade do método da máxima verossimilhança permite que seja facilmente implementado em uma linguagem de programação compilada e de baixo nível, como C por exemplo, o que torna o algoritmo mais veloz. Além da fácil implementação, o método de classificação por máxima

verossimilhança ainda é altamente paralelizável, o que pode permitir a execução do método de forma ainda mais rápida. Essas são limitações que ocorrem ao se trabalhar com a rede GNG, o seu algoritmo mais complicado exige estruturas mais complexas, o que se reflete em tempo de desenvolvimento do programa, além de dificultar ou mesmo impossibilitar a paralelização do código.

Neste trabalho foram calculadas e estudadas medidas globais de acurácia, que permitiram a conclusão sobre a superioridade da rede neural estudada em relação aos outros métodos. Entretanto, não foi investigado como os classificadores atuam dentro de cada classe. Este tipo de análise é importante pois pode-se estar interessado em distinguir apenas uma classe específica entre as demais e, portanto, pode-se optar por um método com um índice de acerto global inferior mas que classifique muito bem a classe de interesse. O exemplo a seguir elucida essa questão.

Assumindo-se um conjunto de dados que possui 3 classes, com 50 observações cada, por exemplo, diferentes classificadores podem chegar as seguintes matrizes de erro:

$$\mathbf{E}_1 = \begin{bmatrix} 50 & 25 & 10 \\ 0 & 0 & 40 \\ 0 & 25 & 0 \end{bmatrix} \quad \mathbf{E}_2 = \begin{bmatrix} 10 & 25 & 5 \\ 20 & 25 & 30 \\ 20 & 0 & 15 \end{bmatrix} \quad (5.1)$$

Para as duas matrizes a acurácia é igual a

$$\frac{\text{tr}(\mathbf{E}_1)}{150} = \frac{\text{tr}(\mathbf{E}_2)}{150} = \frac{50}{150} = \frac{1}{3}$$

porém a segunda (\mathbf{E}_2) é mais homogênea, acertando alguns elementos de todas as classes, enquanto a primeira (\mathbf{E}_1) classifica corretamente todas as observações da classe 1 mas não é capaz de identificar as classes 2 e 3.

De forma geral considera-se que os objetivos deste trabalho foram atingidos e espera-se que com o tempo a metodologia aqui proposta seja o **mainstream** na área de classificação de imagens de sensores remotos. Também espera-se que esta pesquisa seja base de referência para os próximos estudos dessa linha.

REFERÊNCIAS

- AI, J. **Supervised and Unsupervised Prudit for Active Insurance Fraud Detection**. Ann Arbor, Michigan: Proquest, Umi Dissertation Publishing, 2011. p 21–22
- AKINTUNDE, J.A.; ADZANDEH, E.A.; FABIYI, O.O. Spatio-temporal pattern of urban growth in jos metropolis, nigeria. **Remote Sensing Applications: Society and Environment**, Mississippi, USA, 30 p., 2016.
- ALMEIDA, T.; MENESES, P.R.; ROSA, A.N.C.S.; SANO, E.E.; SOUZA, E.B.; BAPTISTA, G.M.M.; BRITES, R.S. **Introdução ao processamento de imagens de sensoriamento remoto**. Brasília, DF: Universidade de Brasília–CNPq, 2012. v. 11. 276 p.
- BALDRIDGE, A.M.; HOOK, S.J.; GROVE, C.I.; RIVERA, G. The aster spectral library version 2.0. **Remote Sensing of Environment**, Pasadena, CA, v. 113, p. 711–715, 2009.
- BASGALUPP, M.P. **Aplicações de Redes Neurais Artificiais no Processo de Precificação de Opções**. 2005. 59 p. Monografia (Trabalho de Conclusão de Curso) – Universidade Federal de Pelotas (UFPel), Pelotas, RS, 2005.
- BENEDIKTSSON, J. A.; SWAIN, P. H.; ERSOY, O. K. Neural network approaches versus statistical methods in classification of multisource remote sensing data. **IEEE Transactions on Geoscience and Remote Sensing**, Lafayette, US, v. 28, n. 4, p. 540–552, 1990.
- BISHOP, C. M. **Neural Networks for Pattern Recognition**. New York, NY: Oxford University Press, 1995. p 304–329.
- BRAGA, A.; LUDERMIR, T.; CARVALHO, A. **Redes neurais artificiais: teoria e aplicações**. Rio de Janeiro: Livros Técnicos e Científicos, 2000. 262 p.
- CAMPBELL, J. B.; WYNNE, R. H. **Introduction to Remote Sensing**. 5. ed. New York, USA: The Guilford Press, 2011. 718 p.
- COLWELL, R. N. History and place of photographic interpretation. In: PHILIPSON, W. R. (Ed.). **The Manual of Photographic interpretation**. 2. ed. Bethesda, MD: American Society for Photogrammetry and Remote Sensing, 1997. cap. 1, p. 3–47.
- CRÓSTA, A. P. **Processamento Digital de Imagens de Sensoriamento Remoto**. Campinas, SP: Universidade Estadual de Campinas–UNICAMP, 1992. 171 p.
- CYBENKO, G. **Continuous valued neural networks with two hidden layers are sufficient**. Massachusetts, US: Tufts University, Department of Computer Science, 1988. 14 p. Technical report.
- CYBENKO, G. Approximation by superpositions of a sigmoidal function. **Mathematics of Control, Signals and Systems**, London, UK, v. 2, n. 4, p. 303–314, 1989.
- FOODY, G. M.; MCCULLOCH, M. B.; YATES, W. B. Classification of remotely sensed data by an artificial neural network: Issues related to training data characteristics. **Photogrammetric Engineering & Remote Sensing**, Bethesda, MD, v. 61, n. 4, p. 391–401, 1995.

FOX, D.M.; LAAROUSSI, Y.; MALKINSON, L.D.; C, F. Maselli; ANDRIEU, J.; BOTTAI, L.; WITTENBERG, L. Postfire: A model to map forest fire burn scar and estimate runoff and soil erosion risks. **Remote Sensing Applications: Society and Environment**, Mississippi, USA, v. 4, p. 83–91, 2016.

FRITZKE, B. A growing neural gas network learns topologies. In: **Advances in Neural Information Processing Systems 7**. Cambridge MA: MIT Press, 1995. p. 625–632.

PEARSON, K.F.R.S. Liii. on lines and planes of closest fit to systems of points in space. **Philosophical Magazine Series 6**, London, UK, v. 2, n. 11, p. 559–572, 1901.

FUNAHASHI, K. I. On the approximate realization of continuous mappings by neural networks. **Neural Networks**, USA, v. 2, n. 3, p. 183–192, 1989.

GONÇALVEZ, M. L. **Uma Arquitetura Neural Modular para Classificação de Imagens Multiespectrais de Sensoriamento Remoto**. 1997. 112 p. Dissertação (Mestrado) – Universidade Estadual de Campinas (UNICAMP), Campinas, SP, 1997.

GONÇALVEZ, M. L. **Métodos de Classificação Não-supervisionada de Imagens de Sensoriamento Remoto usando Mapas Auto-organizáveis de Kohonen**. 2009. 455 p. Tese (Doutorado) – Universidade Estadual de Campinas (UNICAMP), Campinas, SP, 2009.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. H. **The Elements of Statistical Learning**. 2. ed. Stanford, California: Springer, 2008. p 219–293.

HAYKIN, S. **Redes Neurais - Princípio e Prática**. 2. ed. Rio Grande do Sul: Prentice Hall, Inc., 2008. p 178–278.

HERTZ, J.; KROGH, A.; PALMER, R. G. Introduction to the theory of neural computation. In: SANTA FE INSTITUTE. **Lecture Notes Volume in the Santa Fe Institute Studies in the Sciences of Complexity**. New Mexico, USA, 1991. 327 p.

HORNIK, K.; STINCHCOMBE, M.; WHITE, H. Multilayer feedforward networks are universal approximators. **Neural Networks**, USA, v. 2, n. 5, p. 359–366, 1989.

JOLLIFFE, I.T. **Principal Component Analysis**. 2. ed. New York, NY: Springer, 2002. 63 p.

JOSEPH, G. **Fundamentals of remote sensing**. 2. ed. Hyderabad, Índia: Hyderabad Universities Press, 2005. 471 p.

KAISER, H. F. The application of electronic computers to factor analysis. **Educational and Psychological Measurement**, Edinburgh, UK, v. 20, n. 1, p. 141–151, 1960.

KANELLOPOULOS, I.; VARFIS, A.; WILKINSON, G. G.; MÉGIER, J. Land-cover discrimination in spot hrv imagery using an artificial neural network—a 20-class experiment. **International Journal of Remote Sensing**, Indiana, US, v. 13, n. 5, p. 917–924, 1992.

LALONDE, K. **Investigations into the analysis of remote sensing images with a growing neural gas pattern recognition algorithm**. In: THE INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 05., 2005, Montreal. **PROCEEDINGS**. Quebec: IEEE. 2005. v. 3. p. 1698–1703.

LILLESAND, T. M.; KIEFER, R. W.; CHIPMAN, J. W. **Remote Sensing and Image Interpretation**. 5. ed. Hoboken, NJ: Wiley, 2004. 812 p.

- LIMA, A. G. **Classificador Neural Hírido para Imagens Obtidas por Sensoriamento Remoto**. 2011. 65 p. Dissertação (Mestrado) – Universidade Federal do Rio Grande do Norte (UFRN), Natal, RN, 2011.
- LIPPMANN, R. P. An introduction to computing with neural nets. **IEEE ASSP Magazine**, Maryland, US, v. 4, n. 2, p. 4–22, 1987.
- LLOYD, S. P. Least squares quantization in pcm. **IEEE Transactions on Information Theory**, IEEE Press, Piscataway, NJ, USA, v. 28, n. 2, p. 129–137, 1982.
- LU, G.; HALIG, L.; WANG, D.; QIN, X.; CHEN, Z. G.; FEI, B. Spectral-spatial classification for noninvasive cancer detection using hyperspectral imaging. **Journal of Biomedical Optics**, Bellingham, WA, v. 19, 18 p., 2014.
- MACQUEEN, J. **Some methods for classification and analysis of multivariate observations**. In: THE FIFTH BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 67., 1967, Berkeley. **PROCEEDINGS**. California: University of California Press. 1967. p. 281–297.
- MANLY, B. J.F. **Métodos estatísticos multivariados: uma introdução**. Porto Alegre, RS: Bookman, 2008. 230 p.
- MARTINETZ, T. **Competitive hebbian learning rule forms perfectly topology preserving maps**. In: THE INTERNATIONAL CONFERENCE ON ARTIFICIAL NEURAL NETWORKS, 93., 1993, Amsterdam. **PROCEEDINGS**. The Netherlands: Springer London. 1993. p. 427–434.
- MARTINETZ, T.; SCHULTEN, K. **A “neural-gas” network learns topologies**. In: THE INTERNATIONAL CONFERENCE ON ARTIFICIAL NEURAL NETWORKS, 91., 1991, Espoo. **PROCEEDINGS**. Finland: Elsevier Science. 1991. p. 397–402.
- MARTINETZ, T.; SCHULTEN, K. Topology representing networks. **Neural Networks**, Illinois, USA, v. 7, n. 3, p. 507–522, 1994.
- MASELLI, F.; CONESE, C.; PETKOV, L.; RESTI, R. Inclusion of prior probabilities derived from a nonparametric process into the maximum-likelihood classifier. **Photogrammetric Engineering & Remote Sensing**, Bethesda, MD, v. 58, n. 2, p. 201–207, 1992.
- MATHER, P. M.; KOCH, M. **Computer Processing of Remotely-Sensed Images: An Introduction**. 4. ed. New Jersey, USA: John Wiley & Sons, 2011. 460 p.
- MCCULLOCH, W. S.; PITTS, W. H. A logical calculus of the ideas immanent in nervous activity. **Bulletin of Mathematical Biophysics**, New York, NY, v. 5, p. 115–133, 1943.
- MINSKY, M. L.; PAPERT, S. A. **Perceptrons: An Introduction to Computational Geometry**. Cambridge, MA: MIT Press, 1987. 308 p.
- MITCHELL, M.; WILSON, R. R.; TWEDT, D. J.; MINI, A. E.; JAMES, J. D. Object-based forest classification to facilitate landscape-scale conservation in the mississippi alluvial valley. **Remote Sensing Applications: Society and Environment**, Mississippi, USA, v. 4, p. 55–60, 2016.
- NANDY, S. C.; SEN, S.; MAHAJAN, M.; NIMBHORKAR, P.; VARADARAJAN, K. The planar k-means problem is np-hard. **Theoretical Computer Science**, Kolkata, India, v. 442, p. 13–21, 2012.

NASA. **History: From the Beginning**. Disponível em:

<<http://landsat.gsfc.nasa.gov/about/history.html>>. Acesso em: 05 de dezembro de 2016.

PENG, R. D. **R Programming for Data Science**. Leanpub, 2016. 182 p. Disponível em:

<<http://leanpub.com/rprogramming>>. Acesso em: 05 de dezembro de 2016.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2015. 3501 p.

RICHARDS, J. A.; JIA, X. **Remote Sensing Digital Image Analysis. An Introduction**.

Germany: Springer, 2006. 453 p.

RODRIGUES, A. G.; QUEIROZ, R. B.; GÓMEZ, A. T. **A comparative study between neural network and maximum likelihood in the satellite image classification**. In: THE FIRST INTERNATIONAL WORKSHOP ON ARTIFICIAL NEURAL NETWORKS: DATA PREPARATION TECHNIQUES AND APPLICATION DEVELOPMENT, 04., 2004, Setúbal. **PROCEEDINGS**. Portugal: SCITEPRESS, 2004. 8 p.

ROUSE, J. W.; HAAS, R. H.; SCHELL, J. A.; DEERING, D. W. Monitoring vegetation systems in the great plains with erts. **NASA Special Publication**, Maryland, USA, v. 351, p. 309–317, 1974.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning Internal Representations by Error Propagation. In: _____. **Parallel distributed processing** Cambridge, MA: MIT Press, 1986. v. 1, cap. 8, p. 326–371.

GONÇALVES, M. L. **Uma Arquitetura Neural Modular para Classificação de Imagens Multiespectrais de Sensoriamento Remoto**. 1997. 112 p. Dissertação (Mestrado) – Universidade Estadual de Campinas (UNICAMP), Campinas, SP, 1997.

SANCHES, M. K. **Aprendizado de máquina semi-supervisionado: proposta de um algoritmo para rotular exemplos a partir de poucos exemplos rotulados**. 2003. 142 p. Dissertação (Mestrado) – Universidade de São Paulo (USP), São Carlos, SP, 2003.

SCHNEIDER, S.; MURPHY, R. J.; MELKUMYAN, A. Evaluating the performance of a new classifier – the gp-oad: A comparison with existing methods for classifying rock type and mineralogy from hyperspectral imagery. **{ISPRS} Journal of Photogrammetry and Remote Sensing**, Hannover, Germany, v. 98, p. 145–156, 2014.

USGS; NASA. **Landsat 8 Data User's Handbook**. 1.0. ed. Sioux Falls, South Dakota, 2015. 106 p.

ZHUANG, X.; ENGEL, B. A.; BAUMGARDNER, M. F.; SWAIN, P. H. Improving classification of crop residues using digital land ownership data and landsat tm imagery. **Photogrammetric Engineering and Remote Sensing**, Bethesda, MD, v. 57, p. 1487–1492, 1991.

ANEXOS

ANEXO A - MÉDIAS E INTERVALOS DE CONFIANÇA PARA ÍNDICE KAPPA

Amostra (%)	Nível	Método	Transformação	2.5% LI	Kappa	97.5% LS
10	Lv1	GNG	Sem ACP	1.00	1.00	1.00
10	Lv1	GNG	Com ACP	0.97	0.97	0.97
10	Lv1	KMeans	Sem ACP	0.41	0.42	0.43
10	Lv1	KMeans	Com ACP	0.39	0.40	0.42
10	Lv1	MaxVer	Sem ACP	1.00	1.00	1.00
10	Lv1	MaxVer	Com ACP	0.94	0.94	0.94
10	Lv2	GNG	Sem ACP	0.93	0.93	0.94
10	Lv2	GNG	Com ACP	0.68	0.69	0.70
10	Lv2	KMeans	Sem ACP	0.29	0.30	0.32
10	Lv2	KMeans	Com ACP	0.29	0.29	0.30
10	Lv2	MaxVer	Sem ACP	0.80	0.80	0.80
10	Lv2	MaxVer	Com ACP	0.60	0.60	0.61
30	Lv1	GNG	Sem ACP	1.00	1.00	1.00
30	Lv1	GNG	Com ACP	0.97	0.97	0.97
30	Lv1	KMeans	Sem ACP	0.43	0.43	0.44
30	Lv1	KMeans	Com ACP	0.39	0.40	0.42
30	Lv1	MaxVer	Sem ACP	1.00	1.00	1.00
30	Lv1	MaxVer	Com ACP	0.94	0.94	0.94
30	Lv2	GNG	Sem ACP	0.91	0.91	0.92
30	Lv2	GNG	Com ACP	0.66	0.67	0.68
30	Lv2	KMeans	Sem ACP	0.28	0.29	0.31
30	Lv2	KMeans	Com ACP	0.29	0.31	0.33
30	Lv2	MaxVer	Sem ACP	0.80	0.80	0.80
30	Lv2	MaxVer	Com ACP	0.60	0.60	0.61
50	Lv1	GNG	Sem ACP	1.00	1.00	1.00
50	Lv1	GNG	Com ACP	0.97	0.97	0.97
50	Lv1	KMeans	Sem ACP	0.43	0.43	0.44
50	Lv1	KMeans	Com ACP	0.39	0.39	0.39
50	Lv1	MaxVer	Sem ACP	1.00	1.00	1.00
50	Lv1	MaxVer	Com ACP	0.94	0.94	0.94
50	Lv2	GNG	Sem ACP	0.90	0.92	0.93
50	Lv2	GNG	Com ACP	0.67	0.68	0.68
50	Lv2	KMeans	Sem ACP	0.30	0.31	0.33
50	Lv2	KMeans	Com ACP	0.29	0.32	0.35
50	Lv2	MaxVer	Sem ACP	0.80	0.80	0.80
50	Lv2	MaxVer	Com ACP	0.60	0.60	0.60
70	Lv1	GNG	Sem ACP	1.00	1.00	1.00
70	Lv1	GNG	Com ACP	0.97	0.97	0.97
70	Lv1	KMeans	Sem ACP	0.43	0.43	0.44
70	Lv1	KMeans	Com ACP	0.43	0.43	0.43
70	Lv1	MaxVer	Sem ACP	1.00	1.00	1.00
70	Lv1	MaxVer	Com ACP	0.94	0.94	0.94
70	Lv2	GNG	Sem ACP	0.91	0.92	0.92
70	Lv2	GNG	Com ACP	0.66	0.67	0.68
70	Lv2	KMeans	Sem ACP	0.29	0.31	0.34
70	Lv2	KMeans	Com ACP	0.29	0.31	0.32
70	Lv2	MaxVer	Sem ACP	0.80	0.80	0.80
70	Lv2	MaxVer	Com ACP	0.60	0.60	0.61
90	Lv1	GNG	Sem ACP	1.00	1.00	1.00
90	Lv1	GNG	Com ACP	0.97	0.97	0.98
90	Lv1	KMeans	Sem ACP	0.43	0.44	0.44
90	Lv1	KMeans	Com ACP	0.40	0.42	0.43
90	Lv1	MaxVer	Sem ACP	1.00	1.00	1.00
90	Lv1	MaxVer	Com ACP	0.94	0.94	0.94
90	Lv2	GNG	Sem ACP	0.91	0.92	0.92
90	Lv2	GNG	Com ACP	0.66	0.67	0.68
90	Lv2	KMeans	Sem ACP	0.30	0.31	0.32
90	Lv2	KMeans	Com ACP	0.30	0.32	0.34
90	Lv2	MaxVer	Sem ACP	0.80	0.80	0.80
90	Lv2	MaxVer	Com ACP	0.60	0.60	0.61

ANEXO B - MÉDIAS E INTERVALOS DE CONFIANÇA PARA TEMPO DE EXECUÇÃO

Amostra (%)	Nível	Método	Transformação	2.5% LI (s)	Tempo (s)	97.5% LS (s)
10	Lv1	GNG	Sem ACP	1938.10	1959.11	1979.13
10	Lv1	GNG	Com ACP	706.55	709.88	712.32
10	Lv1	KMeans	Sem ACP	2.36	2.60	2.88
10	Lv1	KMeans	Com ACP	1.84	2.07	2.31
10	Lv1	MaxVer	Sem ACP	2526.26	2557.55	2588.84
10	Lv1	MaxVer	Com ACP	1313.43	1322.92	1332.41
10	Lv2	GNG	Sem ACP	1741.04	1766.99	1788.74
10	Lv2	GNG	Com ACP	568.61	571.24	573.22
10	Lv2	KMeans	Sem ACP	2.88	3.04	3.24
10	Lv2	KMeans	Com ACP	3.09	3.49	3.96
10	Lv2	MaxVer	Sem ACP	11591.78	11622.96	11657.64
10	Lv2	MaxVer	Com ACP	6013.69	6051.81	6089.93
30	Lv1	GNG	Sem ACP	1236.75	1414.82	1600.76
30	Lv1	GNG	Com ACP	188.69	300.53	412.55
30	Lv1	KMeans	Sem ACP	1.62	1.74	1.95
30	Lv1	KMeans	Com ACP	1.53	1.57	1.62
30	Lv1	MaxVer	Sem ACP	2118.85	2160.96	2200.86
30	Lv1	MaxVer	Com ACP	1059.76	1092.28	1124.80
30	Lv2	GNG	Sem ACP	893.61	1141.75	1393.57
30	Lv2	GNG	Com ACP	125.73	228.45	331.16
30	Lv2	KMeans	Sem ACP	2.52	2.93	3.36
30	Lv2	KMeans	Com ACP	1.88	1.98	2.08
30	Lv2	MaxVer	Sem ACP	9241.27	9397.12	9591.10
30	Lv2	MaxVer	Com ACP	4470.54	4692.63	4917.36
50	Lv1	GNG	Sem ACP	1193.54	1306.55	1371.62
50	Lv1	GNG	Com ACP	112.25	134.86	148.77
50	Lv1	KMeans	Sem ACP	1.45	1.65	1.83
50	Lv1	KMeans	Com ACP	0.85	1.09	1.27
50	Lv1	MaxVer	Sem ACP	1279.09	1337.96	1430.79
50	Lv1	MaxVer	Com ACP	681.69	704.90	718.55
50	Lv2	GNG	Sem ACP	1034.46	1067.08	1099.70
50	Lv2	GNG	Com ACP	71.89	79.47	85.46
50	Lv2	KMeans	Sem ACP	1.44	1.91	2.24
50	Lv2	KMeans	Com ACP	1.10	1.34	1.49
50	Lv2	MaxVer	Sem ACP	6188.39	6284.76	6356.23
50	Lv2	MaxVer	Com ACP	2976.90	3090.27	3156.45
70	Lv1	GNG	Sem ACP	1124.16	1257.75	1337.20
70	Lv1	GNG	Com ACP	49.64	50.77	51.99
70	Lv1	KMeans	Sem ACP	0.87	0.95	1.03
70	Lv1	KMeans	Com ACP	0.69	0.72	0.74
70	Lv1	MaxVer	Sem ACP	821.68	849.47	898.12
70	Lv1	MaxVer	Com ACP	453.46	455.93	459.91
70	Lv2	GNG	Sem ACP	886.09	960.09	1015.24
70	Lv2	GNG	Com ACP	32.90	34.44	36.29
70	Lv2	KMeans	Sem ACP	1.66	1.82	2.02
70	Lv2	KMeans	Com ACP	0.76	0.96	1.15
70	Lv2	MaxVer	Sem ACP	3870.65	3936.81	4020.21
70	Lv2	MaxVer	Com ACP	1870.96	1882.44	1893.30
90	Lv1	GNG	Sem ACP	660.44	738.62	845.25
90	Lv1	GNG	Com ACP	20.13	21.46	22.55
90	Lv1	KMeans	Sem ACP	0.38	0.48	0.57
90	Lv1	KMeans	Com ACP	0.21	0.24	0.28
90	Lv1	MaxVer	Sem ACP	155.81	200.59	255.14
90	Lv1	MaxVer	Com ACP	110.46	121.03	135.90
90	Lv2	GNG	Sem ACP	693.08	752.79	800.93
90	Lv2	GNG	Com ACP	16.88	17.69	18.63
90	Lv2	KMeans	Sem ACP	0.73	0.97	1.37
90	Lv2	KMeans	Com ACP	0.58	0.83	1.22
90	Lv2	MaxVer	Sem ACP	910.38	985.26	1115.08
90	Lv2	MaxVer	Com ACP	498.08	529.92	578.29