

### **5.1 Considerações Iniciais**

A análise de componentes principais (*PCA*) [TURK, M. A. & PENTLAND, A. P. (1991)] é uma transformação linear ortogonal de um espaço  $q$ -dimensional para um espaço  $n$ -dimensional, com  $n \leq q$ . As coordenadas dos dados no novo espaço não são correlacionadas e a maior quantidade de variância dos dados originais é preservada usando-se poucas coordenadas.

Portanto, dadas  $p$  variáveis  $X_1, X_2, \dots, X_p$  encontram-se combinações lineares para produzir novas variáveis  $Z_1, Z_2, \dots, Z_p$  que não sejam correlacionadas, onde  $Z_i$  são as componentes principais e são ordenadas de forma que  $\text{var}(Z_1) \geq \text{var}(Z_2) \geq \dots \geq \text{var}(Z_p)$  e  $\text{var}(Z_i)$  representa a variância de  $Z_i$ .

As variâncias resultantes podem ser tão pequenas ao ponto de serem ignoradas, então a variação no conjunto de dados pode ser descrita pelas poucas variáveis  $Z$  que retenham as maiores variâncias.

O método PCA [DUNTEMAN, H. G. (1989)] consiste em encontrar os autovalores e os correspondentes auto-vetores da matriz de covariância do conjunto de dados considerado, assumindo que os autovalores estão ordenados por valores decrescentes,  $I_1 \geq I_2 \geq \dots \geq I_p \geq 0$ , onde  $I_i$  corresponde à  $i$ -ésima componente

principal. Ou seja,  $Z_i = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$ , sendo  $\text{var}(Z_i) = 1$  e  $a_{i1}, a_{i2}, \dots, a_{ip}$  os elementos do auto-vetor correspondente.

## 5.2 Variância e Covariância

A estatística está baseada no conceito de amostragem, sendo que em uma amostra representa um conjunto de dados, com a média dada pela equação 5.1.

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \quad (5.1)$$

e o desvio padrão, que informa quanto os elementos variam em relação à média, dada na equação 5.2.

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)}} \quad (5.2)$$

A medida de dispersão do conjunto de dados é a variância, dada na equação 5.3.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n-1)} \quad (5.3)$$

O desvio padrão e a variância, possuem uma característica em comum: ambos são métricas de dispersão de conjuntos, mas infelizmente analisam somente uma única dimensão dos dados. Para os conjuntos que são multidimensionais, existem outros métodos estatísticos que analisam o relacionamento entre dimensões. Por exemplo, para esse tipo de análise, utiliza-se a covariância, que mede o grau de correlação entre dois conjuntos bi-dimensionais de dados, ou seja, fornece a informação da existência ou não de alguma possível relação entre esses dois conjuntos. A covariância é dada pela equação 5.4.

$$\text{cov}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)} \quad (5.4)$$

Uma propriedade fundamental da covariância é que ela indica qual o relacionamento entre as dimensões em questão. Se o sinal resultante for positivo, indica que, se o valor de uma dimensão cresce o valor da outra dimensão também cresce, caso contrário, se a covariância for negativa, indica que, se o valor de uma dimensão cresce a outra decresce. Caso a covariância seja zero, isso significa que não existe relação alguma entre as dimensões que estão sendo calculadas. Essa técnica tem sido amplamente utilizada na análise de conjuntos de dados com grandes dimensões.

### 5.3 A Matriz de Covariância

Normalmente a covariância é medida entre duas dimensões. Para maiores dimensões, ou seja, para dados  $n$ -dimensionais, a quantidade de valores de covariância pode ser calculada pela equação 5.5.

$$\frac{n!}{(n-2)!*2} \quad (5.5)$$

Para que seja possível representar valores distintos deve se utilizar a matriz de covariância, conforme mostra a equação 5.6.

$$C^{n \times n} = (c_{ij}; c_{ij} = \text{cov}(Dim_i, Dim_j)) \quad (5.6)$$

A matriz de covariância  $C$  é uma matriz simétrica com  $n$  linhas e  $n$  colunas, dado o conjunto de dados de dimensão  $n$ , sendo que cada entrada é o valor de covariância entre duas dimensões. A equação 5.7 mostra como exemplo a matriz de covariância de um conjunto de três dimensões.

$$C = \begin{pmatrix} \text{cov}(x, x) & \text{cov}(x, y) & \text{cov}(x, z) \\ \text{cov}(y, x) & \text{cov}(y, y) & \text{cov}(y, z) \\ \text{cov}(z, x) & \text{cov}(z, y) & \text{cov}(z, z) \end{pmatrix} \quad (5.7)$$

#### 5.4 Auto-vetores e Auto-valores

Para uma matriz de covariância  $C$ , vale a expressão  $C * u = \lambda * u$ , onde os vetores  $u$  são chamados de auto-vetores e os valores  $\lambda$ , são os auto-valores. Caso a matriz de covariância  $C$  seja quadrada e de ordem  $n$ , então esta possui exatamente  $n$  auto-vetores e todos esses auto-vetores são ortonormais a todos os outros, definindo um espaço de dimensão  $n$ .

Os auto-vetores disponibilizam informações sobre padrões nos dados. Após o cálculo dos auto-vetores, deve-se ordená-los a partir dos auto-valores associados, formando um grau decrescente de representatividade no conjunto, os maiores auto-valores estão associados com auto-vetores que descrevem padrões no conjunto de dados.

#### 5.5 Redução de Dimensionalidade

A redução de dimensionalidade significa que se a maioria das variáveis apresentarem variâncias tão pequenas ao ponto de serem ignoradas, a variação no conjunto de dados pode ser apropriadamente descrita por poucas variáveis  $Z$  que retenham as maiores variâncias.

Da perspectiva do reconhecimento de padrões, o valor real da análise de componentes principais (PCA) [HAYKIN, S. (1999)] é prover uma técnica que seja realmente efetiva para a redução de dimensionalidade. Em resumo, pode-se reduzir o número de características para a representação de dados de uma imagem que seja eficiente, descartando algumas combinações lineares

Conforme mostrado na figura 5.1, temos um sistema de codificação com o objetivo de representação, de forma aproximada, dos dados do vetor  $x$ .

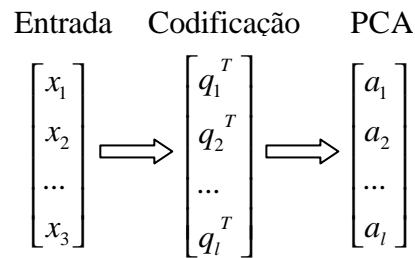


Figura 5.1 – Codificação do PCA.

Na figura 5.2, pode-se notar o processo de mapeamento do espaço de características para o espaço de dados, realizando o processo de decodificação, para que seja possível a reconstrução dos valores utilizados como entrada, ou seja, o vetor original  $x$ .

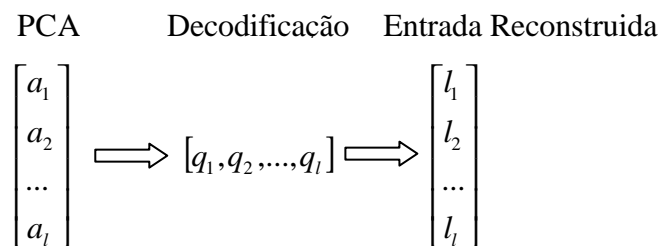


Figura 5.2 – Decodificação do PCA.

## 5.6 Aplicação do PCA no Reconhecimento de Imagens de Faces Humanas

Uma imagem de face  $I(x,y)$ , é uma matriz de  $N$  linhas e  $M$  colunas que pode ser considerada um vetor de  $N \times M$  dimensões, ou seja, uma imagem pode ser representada por um ponto em um espaço de  $N \times M$  dimensões.

Para a aplicação da Análise de Componentes Principais (PCA) em imagens de faces humanas, deve-se primeiramente mapear um conjunto de imagens em um conjunto de pontos nesse espaço. Imagens de faces que sejam de uma mesma pessoa são mapeadas próximas entre si, nesse espaço. O objetivo do PCA [SMITH, L I (2002)] é encontrar vetores que melhor descrevam a variação nessas imagens.

Sendo  $T_1, T_2, T_3, \dots, T_m$ , o conjunto de  $M$  imagens de um individuo, pode-se descrever a face média desse conjunto de acordo com a equação 5.8.

$$\Psi = \frac{1}{M} \sum_{n=1}^M T_n \quad (5.8)$$

Cada imagem difere da média através do vetor  $\mathbf{f}_i = T_i - \Psi$ , sendo que  $1 \leq i \leq M$ . A técnica consiste em submeter o conjunto de vetores  $\mathbf{f}_i$  ao algoritmo do PCA, no qual irá realizar a busca de  $M$  vetores ortonormais  $\mathbf{m}$  que melhor descrevam a distribuição dos dados. Os vetores  $\mathbf{m}$  são os auto-vetores da matriz de covariância, de acordo com a equação 5.9, sendo que  $A = \{\mathbf{f}_1, \mathbf{f}_2, \mathbf{f}_3, \dots, \mathbf{f}_M\}$  é de ordem  $N^2 \times M$ . Com isso, caracteriza-se um grande problema para o processamento dessa informação, em termos computacionais, pois a matriz  $C$  é de ordem  $N^2 \times N^2$ .

$$C = \frac{1}{M} \sum_{n=1}^M \mathbf{f}_n \mathbf{f}_n^T = AA^T \quad (5.9)$$

Pode-se afirmar que o método consiste em encontrar os  $p$  auto-vetores com os maiores auto-valores associados em uma matriz de covariância  $C = A^T A$ , que é de ordem  $M \times M$ . Deve-se, então, escolher os  $p$  auto-vetores de maior representatividade, sendo que esses vetores pelo fato de representarem faces ortogonais, são denominados de auto-faces, como mostra a figura 5.3.



Figura 5.3 – Exemplos de auto-vetores gerados pela técnica PCA [KIRBY, M. & SIROVICH, I. (1990)].

O próximo passo será encontrar os vetores de características de cada classe de pessoas [ANDRADE, S. S. (2003)]. Cada imagem de face de uma pessoa é projetada para o novo espaço de dimensão ( $p \leq M$ ).

Para o processo de identificação de uma nova imagem obtém-se o vetor de características dessa nova imagem e compara-se esse vetor com os vetores de características de todas as classes de pessoas.

A imagem é então classificada através da distância euclidiana entre o vetor de características da classe e o da imagem da pessoa em questão. É necessário definir um limiar  $\Theta$ , como a distância mínima para que uma imagem de face humana seja classificada como reconhecida pelo método PCA.

O trabalho de [THOMAZ (1999)] apresenta uma descrição relativo à interpretação geométrica do PCA:

*“As componentes principais podem ser entendidas como resultados de duas operações geométricas básicas:*

- 1. Uma translação na origem do sistema de eixos de coordenadas para o ponto definido pelo vetor-média  $\bar{x}$  e*
- 2. Uma rotação nos eixos coordenados originais de tal forma que os mesmos se alinhem segundo as direções de máxima variância dos dados.”*

Um exemplo de interpretação geométrica do PCA, no qual o auto-vetor com o maior auto-valor associado possui a maior representatividade do conjunto, de acordo com a figura 5.4.

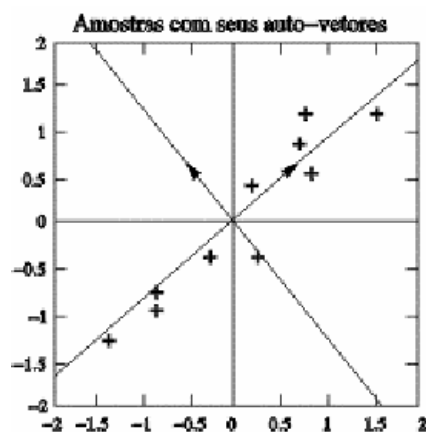


Figura 5.4 – Amostras de auto-vetores. [ANDRADE, S. S. (2003)].

### 5.7 IMPCA (PCA 2D)

O algoritmo do *IMPCA* [YANG, J., et al. (2004)] baseia-se no conceito do *PCA* [TURK, M. A. & PENTLAND, A. P. (1991)]. Dado  $X$  um vetor coluna unitário  $n$ -dimensional, projeta-se a imagem  $A$ , uma matriz  $m \times n$ , fazendo-se a transformação linear:

$$Y = AX \quad (5.10)$$

Obtém-se um vetor projetado  $Y$   $m$ -dimensional, que é chamado de vetor de características da imagem  $A$ . Para se fazer uma boa projeção deste vetor  $X$  de fato, a difusão das amostras projetadas podem ser introduzidas para medir a discriminação de energia da projeção do vetor  $X$ . A difusão das amostras projetadas podem ser caracterizadas pelo traço da matriz de covariância do vetor de características projetado.

A partir deste ponto de vista é adotado o seguinte critério:

$$J(X) = tr(S_x), \quad (5.11)$$

Aonde  $S_x$  denota a matriz de covariância do vetor de características projetado do conjunto de treinamento e  $tr(S_x)$  denota o traço de  $S_x$ . O significado físico da maximização do critério na equação 5.11 é encontrar a projeção  $X$ , sobre todas as amostras projetadas, no qual a difusão total das amostras resultantes é maximizada. A matriz de covariância  $S_x$  pode ser definida como:

$$S_x = E(Y - EY)^T = E[AX - E(AX - E(AX))][AX - E(AX)]^T = E[(A - EA)X][(A - EA)X]^T \quad (5.12)$$

então,

$$tr(S_x) = X^T [E(A - EA)^T (A - EA)]X \quad (5.13)$$

Considerando-se a matriz  $G_1$  dada na equação 5.14 como a



$$G_t = E[(A - EA)^T (A - EA)] \quad (5.14)$$

matriz de covariância da imagem, é fácil verificar que  $G_t$  é uma matriz definida e não negativa  $n \times n$ . Pode-se, então avaliar  $G_t$  diretamente usando as amostras de imagens de treinamento. Supondo-se que as amostras da matriz de treinamento sejam  $M$  no total, a  $j$ th imagem de treinamento é definida como uma matriz  $m \times n$   $A_j (j = 1, 2, \dots, M)$  e a média de todas as amostras de imagem de treinamento é definida por  $\bar{A}$ . Então  $G_t$  pode ser avaliada como

$$G_t = \frac{1}{M} \sum_{j=1}^M (A_j - \bar{A})^T (A_j - \bar{A}) \quad (5.15)$$

Alternativamente, o critério na equação 5.11 pode ser expressado por

$$J(X) = X^T G_t X \quad (5.16)$$

no qual  $X$  é um vetor coluna unitário. O critério é chamado de critério de espalhamento total. O vetor unitário  $X$  que maximiza o critério é chamado de eixo de projeção ótimo. Intuitivamente, este significa que a difusão total das amostras projetadas é maximizado antes da projeção da matriz imagem sobre  $X$ .

O eixo de projeção ótima  $X_{opt}$  é um vetor unitário que maximiza  $J(X)$ , isto é, o auto-vetor de  $G_t$  correspondente ao maior auto-valor. No geral é necessário selecionar um grupo de eixos de projeção,  $X_1, \dots, X_d$ , sujeito a restrições ortonormais e maximizando o critério  $J(X)$ , dado que,

$$\begin{cases} \{X_1, \dots, X_d\} = \arg \max J(X) \\ X_i^T X_j = 0, i \neq j, i, j = 1, \dots, d \end{cases} \quad (5.17)$$

De fato, os eixos de projeção ótimos,  $X_1, \dots, X_d$  são auto-vetores de  $G_t$  correspondendo aos primeiros  $d$  maiores auto-valores.

## 5.8 Considerações Finais

Neste capítulo foram apresentadas duas técnicas para redução de dimensionalidade, o *PCA* [KIM, K. A., et al (2004)] e o *IMPCA* [WANG, L. et al. (2004)]. O *IMPCA* [YANG J., et al. (2004)] como o próprio nome sugere (Image Principal Component Analysis) foi desenvolvido para extração de características em imagens ou seja em um domínio 2D. Ao contrário do *PCA* [DUNTEMAN, H. G. (1989)] o *IMPCA* [YANG, J. & YANG, Y. S. (2001)] trabalha com matrizes em vez de vetores como acontece com o *PCA*, ou seja, para se utilizar *PCA* em imagens, deve-se transformar a matriz de imagem em um vetor. Com o *IMPCA* [WANG, L. et al (2004)] isto não é necessário, pois a matriz de covariância pode ser construída diretamente usando a matriz original da imagem.

Em contraste, a matriz de covariância da *IMPCA* é muito menor do que a da técnica *PCA* [CHEN, F. C. et al. (2003)]. Como resultado tem-se duas principais vantagens do *IMPCA* sobre o *PCA* [KIM, K. A., et al. (2004)]. Primeiro, a matriz de covariância é mais precisa e segundo [YANG, J., et al. (2004)] o cálculo de tempo de processamento é menor para o *IMPCA*, pois a matriz de covariância do *IMPCA* é menor que a do *PCA*, fazendo com que o sistema ganhe desempenho.