

# Princípios de Redes Neurais Artificiais e de Reconhecimento de Padrões

Prof. Antônio de Pádua Braga  
Departamento de Engenharia Eletrônica  
Escola de Engenharia da UFMG

7 de junho de 2017



# Sumário

<b>1</b>	<b>Extração de Características</b>	<b>5</b>
1.1	Introdução . . . . .	5
1.2	Análise de Componentes Principais . . . . .	5
1.2.1	Centralização do conjunto de amostras . . . . .	5
1.2.2	Projeção ótima . . . . .	6
1.2.3	Cálculo da projeção . . . . .	8
1.2.4	Maximizando a projeção . . . . .	9
1.2.5	Exemplo . . . . .	10



# Capítulo 1

## Extração de Características

### 1.1 Introdução

### 1.2 Análise de Componentes Principais

#### 1.2.1 Centralização do conjunto de amostras

Considere um conjunto  $D = \{\mathbf{x}_i, y_i\}_{i=1}^N$ , com  $N$  pares de amostras  $(\mathbf{x}_i, y_i)$ , em que  $\mathbf{x}_i$  é o vetor de entrada de dimensão  $n$  e  $y_i$  indica a sua classe. O centróide  $\mathbf{x}_m$  das amostras de entrada  $\mathbf{x}_i$  corresponde ao vetor cujas distâncias em relação a todos os outros vetores do conjunto  $D_u = \{\mathbf{x}_i\}_{i=1}^N$  seja mínima ou, em outras palavras, que minimize a função de custo da Equação 1.1.

$$J(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{N_u} (\mathbf{x} - \mathbf{x}_i)^2 \quad (1.1)$$

em que  $\mathbf{x}$  é um vetor arbitrário de mesma dimensão de  $\mathbf{x}_i$ .

Com o objetivo de obter o vetor  $\mathbf{x}_m$ , deriva-se  $J(\mathbf{x})$  em relação a  $\mathbf{x}$  e chega-se à expressão apresentada na Equação 1.2.

$$\frac{\partial J(\mathbf{x})}{\partial \mathbf{x}} = N\mathbf{x} - \sum_i \mathbf{x}_i \quad (1.2)$$

Igualando-se a Equação 1.2 a zero chega-se à Equação 1.3 que indica que o vetor  $\mathbf{x}_m$  que minimiza  $J(\mathbf{x})$  equivale ao vetor médio de  $D$ .

$$\mathbf{x}_m = \frac{1}{N} \sum_i \mathbf{x}_i \quad (1.3)$$

A Figura 1.1 apresenta dados amostrados de duas distribuições normais com médias nos pontos  $(2, 2)$  e  $(4, 4)$ , assim como o ponto médio de todo o conjunto de dados, obtido conforme Equação 1.3.

Considerando-se agora  $\mathbf{x}_m$  como a origem de um novo sistema de coordenadas, o vetor  $(\mathbf{x}_i - \mathbf{x}_m)$  corresponde ao vetor arbitrário  $\mathbf{x}_i$  centralizado no novo sistema com origem em  $\mathbf{x}_m$ .

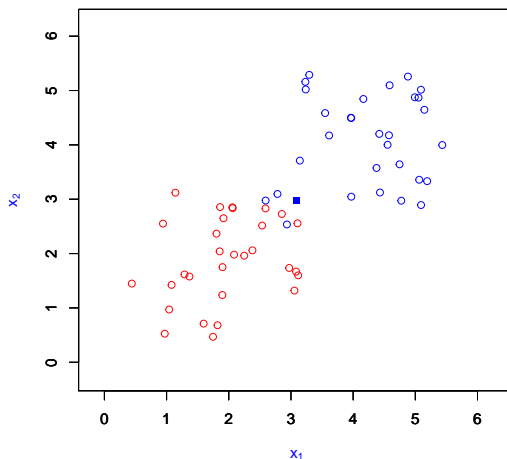


Figura 1.1: Dados amostrados de duas distribuições Normais formando agrupamentos distintos. O ponto médio é apresentado na cor azul no centro da figura.

### 1.2.2 Projeção ótima

Conforme Figura 1.1, o novo sistema de coordenadas tem o ponto  $\mathbf{x}_m$  como origem, pela qual passam infinitas retas de direção arbitrária  $\mathbf{u}$ . Deseja-se encontrar agora qual destas retas maximiza as magnitudes das projeções dos vetores  $(\mathbf{x}_i - \mathbf{x}_m)$  sobre a mesma. As coordenadas dos vetores  $\mathbf{u}$  em relação à origem  $(0, 0)$  podem ser obtidas como  $\mathbf{u}_0 = \mathbf{x}_m + \mathbf{u}_{x_m}$  em que  $\mathbf{u}_{x_m} = [\cos(\alpha), \sin(\alpha)]^T$ . Como exemplo, a Figura 1.2 ilustra os dados da Figura 1.1, vetores arbitrários  $\mathbf{u}$  com origem em  $\mathbf{x}_m$  e também um vetor arbitrário  $(\mathbf{x}_i - \mathbf{x}_m)$ , cuja projeção sobre a reta de direção  $\mathbf{u}$  é obtida por meio da Equação 1.4.

$$a_i = \mathbf{u}^T (\mathbf{x}_i - \mathbf{x}_m) \quad (1.4)$$

A Figura 1.3 mostra as projeções de todas as amostras do conjunto de dados sobre uma reta arbitrária que passa por  $\mathbf{x}_m$ . As magnitudes de  $a_i$ , obtidas conforme Equação 1.4 são proporcionais às variâncias destas projeções no novo sistema de coordenadas. Assim, os componentes principais de um conjunto de amostras, segundo a análise de componentes principais, é determinada pelas variâncias dos dados em cada eixo do novo sistema de coordenadas. Neste caso, o novo sistema de coordenadas possui origem em  $\mathbf{x}_m$ , porém, será rotacionado visando a maximizar as projeções  $a_i$ . Deseja-se, assim, encontrar um novo sistema de coordenadas que, após uma translação e uma rotação, maximize as projeções das amostras nos seus eixos.

A Figura 1.4 mostra as variâncias das projeções no novo sistema de coordenadas para retas arbitrárias, cujas direções variam de 0 a  $2\pi$  com passo de  $\frac{\pi}{10}$ . Pode-se observar que a variância máxima para este exemplo ocorre no terceiro valor do ângulo, ou seja, quando a rotação da reta é de  $\frac{\pi}{5}$ . Este valor máximo se

repete para a reta simétrica em  $\frac{13\pi}{10}$ , já que o vetor  $\mathbf{u}$  foi rotacionado em torno de  $\mathbf{x}_m$ .

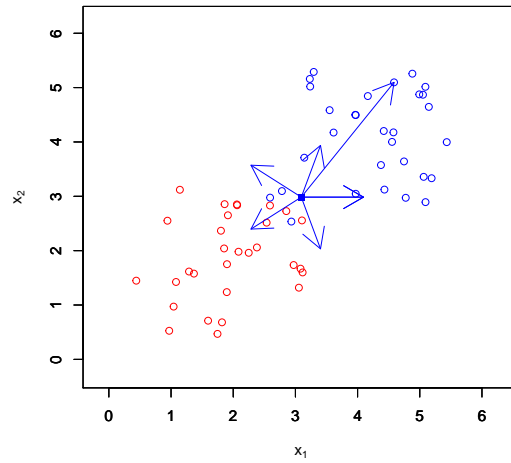


Figura 1.2: Vetores  $\mathbf{u}$  indicando direções arbitrárias para as possíveis retas que passam pelo ponto médio.

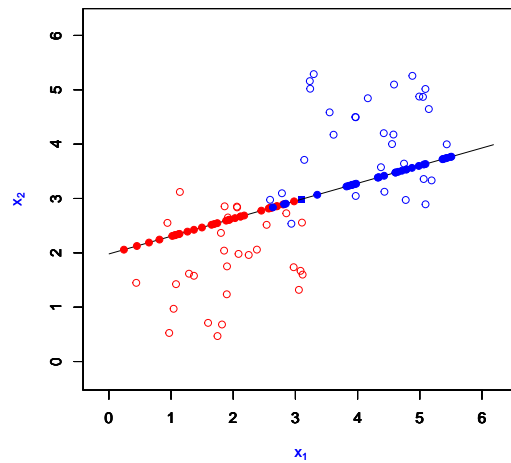


Figura 1.3: Reta de direção arbitrária  $\mathbf{u}$  que passa pelo ponto  $\mathbf{x}_m$ . A figura mostra também as projeções das amostras das duas classes sobre a mesma.

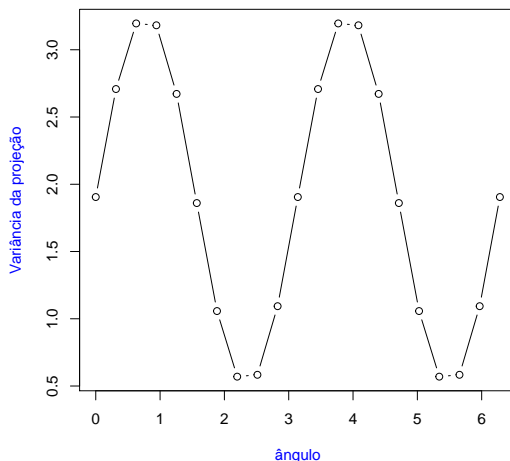


Figura 1.4: Variância das projeções  $a_i$  em função do ângulo de rotação do ângulo de rotação do vetor  $\mathbf{u}$ .

### 1.2.3 Cálculo da projeção

Conforme seção anterior, considere que  $a_i$  seja o módulo da projeção do vetor  $\mathbf{x}_i$  sobre a reta de direção  $\mathbf{u}$  que passa pelo vetor  $\mathbf{x}_m$ . A projeção  $\mathbf{x}'_i$  de  $\mathbf{x}_i$  em relação à origem do sistema de coordenadas corresponde então à Equação 1.5.

$$\mathbf{x}'_i = \mathbf{x}_m + a_i \mathbf{u} \quad (1.5)$$

Substituindo-se a Equação 1.5 na Equação 1.1, chega-se à Equação 1.6 com argumento  $a_i$ , a qual é expandida na Equação 1.7.

$$J(a_i) = \frac{1}{2} \sum_{i=1}^{N_u} ((\mathbf{x}_m + a_i \mathbf{u}) - \mathbf{x}_i)^2 \quad (1.6)$$

$$J(a_i) = \frac{1}{2} \sum_{i=1}^{N_u} (\mathbf{x}_m^2 + 2\mathbf{x}_m a_i \mathbf{u} + a_i^2 \mathbf{u}^2 - 2\mathbf{x}_m \mathbf{x}_i - 2a_i \mathbf{u} \mathbf{x}_i + \mathbf{x}_i^2) \quad (1.7)$$

Derivando-se a Equação 1.7 em relação a  $a_i$  e considerando-se que  $\mathbf{u}^2 = 1$ , chega-se à Equação 1.8.

$$\frac{\partial J(a_i)}{\partial a_i} = \frac{1}{2} \sum_{i=1}^{N_u} (2\mathbf{x}_m \mathbf{u} + 2a_i - 2\mathbf{u} \mathbf{x}_i) \quad (1.8)$$

Igualando-se a Equação 1.8 a zero, obtém-se o valor de  $a_i$  que minimiza a função de custo da Equação 1.1, que é apresentada na Equação 1.9.

$$a_i^* = \mathbf{u}^T (\mathbf{x}_i - \mathbf{x}_m) \quad (1.9)$$



A interpretação da Equação 1.9 é que a projeção  $a_i$  que minimiza a função de custo da Equação 1.1 é aquela que corresponde à projeção do vetor  $(\mathbf{x} - \mathbf{x}_m)$  sobre a reta de direção  $\mathbf{u}$  com origem em  $\mathbf{x}_m$ . Considerando-se que  $\mathbf{u}$  pode ser um vetor unitário qualquer com origem em  $\mathbf{x}_m$  deseja-se saber agora qual o vetor  $\mathbf{u}$  maximiza a projeção.

### 1.2.4 Maximizando a projeção

Substituindo-se agora a projeção ótima da Equação 1.9 na Equação 1.1 chega-se às Equações 1.10 e 1.11.

$$J(\mathbf{u}) = \frac{1}{2} \sum_{i=1}^{N_u} ((\mathbf{x}_m + \mathbf{u}^T(\mathbf{x}_i - \mathbf{x}_m)\mathbf{u}) - \mathbf{x}_i)^2 \quad (1.10)$$

$$J(\mathbf{u}) = -\mathbf{u} \sum_{i=1}^{N_u} \underbrace{(\mathbf{x}_i - \mathbf{x}_m)(\mathbf{x}_i - \mathbf{x}_m)^T}_{(N-1) \times \text{Matriz de Covariância}} \mathbf{u} + \sum_{i=1}^N (\mathbf{x}_i - \mathbf{x}_m)^2 \quad (1.11)$$

Assim, de acordo com a Equação 1.11, o vetor  $\mathbf{u}$  que minimiza a função de custo da Equação 1.1 é aquele que maximiza o produto  $-\mathbf{u}^T \mathbf{S} \mathbf{u}$ , em que  $\mathbf{S}$  equivale a  $(N - 1)$  vezes a matriz de covariância. A Figura 1.5 a variação de  $J\mathbf{u}$  da Equação 1.11 em função do ângulo de rotação para o exemplo das seções anteriores, a qual pode ser comparada com gráfico análogo apresentado na Figura 1.4. Observa-se, assim, que o vetor  $\mathbf{u}$  que minimiza  $J\mathbf{u}$  é o mesmo que maximiza a variância da projeção.

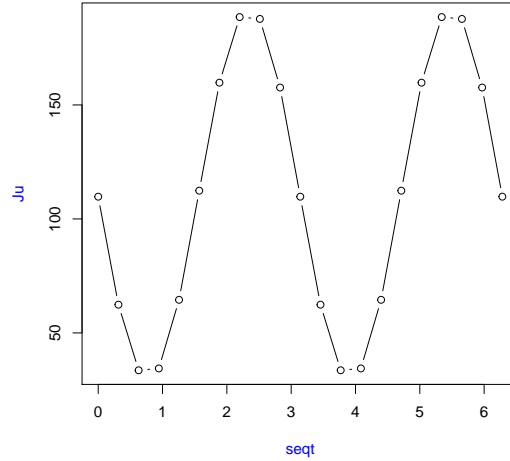


Figura 1.5: Função de custo determinada pela Equação 1.11 em função do ângulo de rotação do vetor  $\mathbf{u}$ .

Assim, a Equação 1.12 apresenta a função de custo  $Q(\mathbf{u})$  para a minimização de  $-\mathbf{u}^T \mathbf{S} \mathbf{u}$  utilizando-se Multiplicadores de Lagrange [NW06].

$$Q(\mathbf{u}) = -\mathbf{u}^T \mathbf{S} \mathbf{u} + \lambda(\mathbf{u}^T \mathbf{u} - 1) \quad (1.12)$$

onde  $\lambda$  é um Multiplicador de Lagrange.

Igualando-se a derivada parcial de  $Q(\mathbf{u})$  e igualando-a a zero, chega-se à Equação 1.13.

$$\mathbf{S} \mathbf{u} = \lambda \mathbf{u} \quad (1.13)$$

A primeira conclusão importante da Equação 1.13 é que  $\mathbf{u}$  é um autovetor de  $\mathbf{S}$ . Outra conclusão importante é que a magnitude de  $\mathbf{S} \mathbf{u}$  que minimiza  $Q(\mathbf{u})$  é proporcional a  $\lambda$ , ou seja, a importância de cada eixo do novo sistema de coordenadas formado pelos autovetores de  $\mathbf{S}$  na maximização da projeção é proporcional ao seu autovalor.

### 1.2.5 Exemplo

```
> rm(list=ls())
> data(iris)
> X<-as.matrix(iris[, (1:4)])
> meanx<-colMeans(X)
> X<- X - t(replicate(dim(X)[1],meanx))
> pcaX<-prcomp(X)
> us<-pcaX$rotation
> projX<-X %*% us
> barplot(pcaX$sdev)
> # Reduzindo para os dois primeiros eixos
> plot(projX[,1],projX[,2],type='p',xlim=c(-4,4),ylim=c(-2,2),xlab='PCA1',ylab='PCA2')
> par(new=TRUE)
> plot(projX[(1:50),1],projX[(1:50),2],type='p',xlim=c(-4,4),ylim=c(-2,2),col='red',xlab='PCA1',ylab='PCA2')
> par(new=TRUE)
> plot(projX[(51:100),1],projX[(51:100),2],type='p',xlim=c(-4,4),ylim=c(-2,2),col='blue',xlab='PCA1',ylab='PCA2')
```

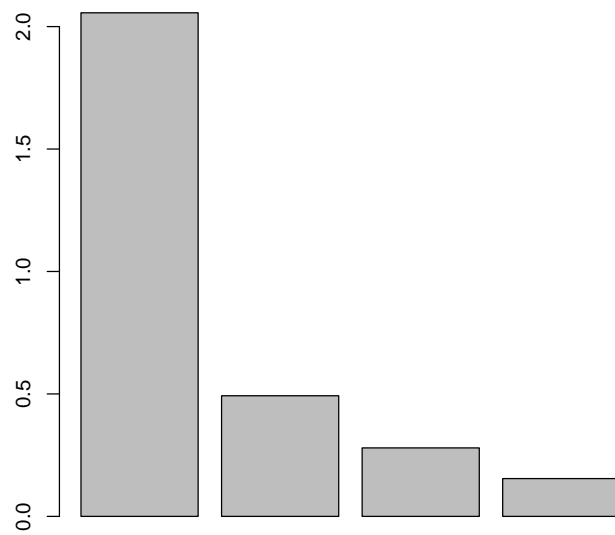


Figura 1.6: Autovalores para a base de dados Iris. As duas primeiras coordenadas concentram boa parte da variância da projeção.

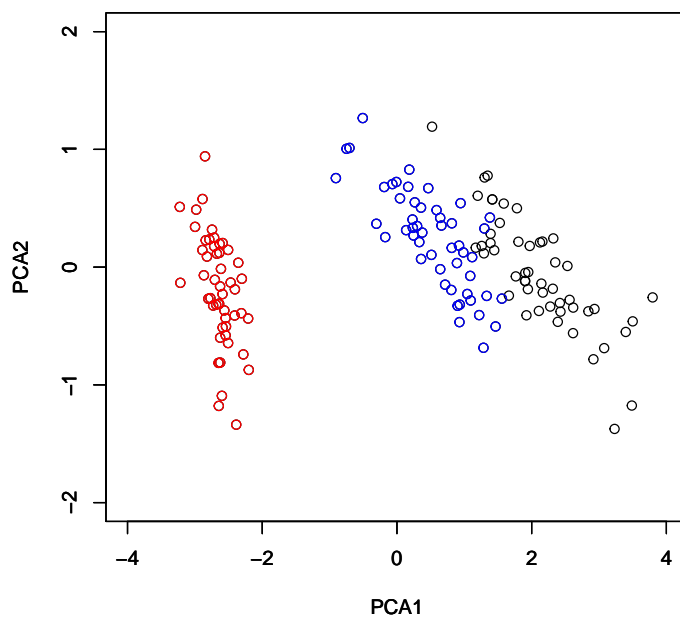


Figura 1.7: Projeção das três classes da base de dados Iris nas duas primeiras coordenadas PCA.

# Referências Bibliográficas

- [NW06] J. Nocedal and S. J. Wright. *Numerical Optimization*. Springer, New York, 2nd edition, 2006.