

LDA

Linear Discriminant Analysis

Ronald Fisher, 1936, p/ classificações de algumas flores.

PEDRO PATRÍCIO
PEDRO@MATH.

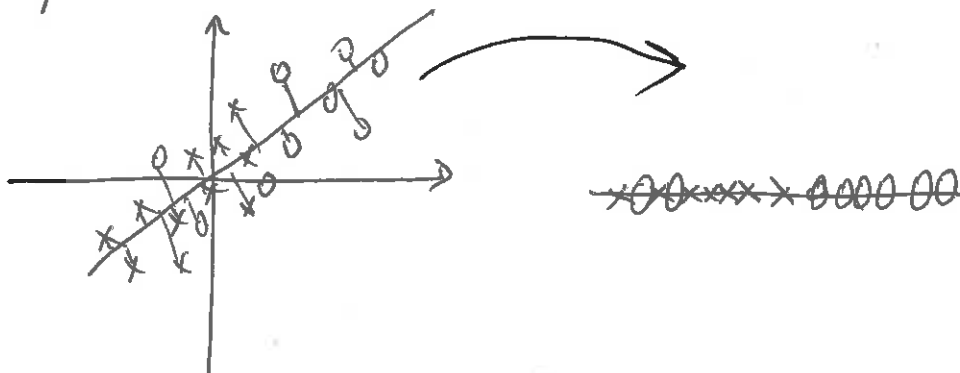
3.1

Tanto o PCA como o LDA são técnicas lineares para redução de dimensão. O PCA pode ser encarado como um método não supervisionado, já que ignora a classificação que à priori se tinha das amostras; apenas procura as direcções de maior variância.

① LDA é supervisionado já que calcula as direcções que maximizam a separação entre as classes.

No que toca ao reconhecimento por imagens, o PCA tende a ser superior ao LDA no caso do nº de amostras por classe ser reduzido.

Relembre que no PCA pretendemos a(s) direcção(ões) que maximiza(m) a variância, independentemente da classificação que tenhamos dos objectos (amostras)



Seja μ_i a média na classe i ,

B.3

$$\mu_i = \frac{1}{N_i} \sum_{x \in C_i} x, \text{ onde } N_i = \#C_i,$$

e μ a média amostral, $\mu = \frac{1}{N} \sum_{x \in G \cup C_2} x_i$

onde $N = \#C$

Seja S_B a matriz de dispersão inter-classes

$$S_B = \sum_{i=1}^{\#C} N_i (\mu_i - \mu)(\mu_i - \mu)^T$$

e S_W a matriz de dispersão intra-classe,

$$S_W = \sum_{i=1}^K \sum_{x_j \in C_i} (x_j - \mu_i)(x_j - \mu_i)^T$$

(resp. between scatter matrix e within scatter matrix)

Podemos escrever $S_W = \Sigma_1 + \Sigma_2 + \dots + \Sigma_K$,

onde Σ_i é a matriz de dispersão para a classe C_i .

O método de Fisher consiste em maximizar

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

Maximizar $J(w)$ é equivalente a maximizar o numerador mantendo o denominador constante.

$$\max_w w^T S_B w$$

sujeito a

$$w^T S_w w = k' \quad \dots$$

Que podemos resolver usando multiplicadores de Lagrange:

$$\begin{aligned} \text{Definindo } L(w, \lambda) &= w^T S_B w - \lambda (w^T S_w w - k') \\ &= w^T (S_B - \lambda S_w) w + \lambda k' \end{aligned}$$

Calculando a gradiente a ordem a w ,

$$\nabla_w L = 2(S_B - \lambda S_w)w = 0$$

$$\text{e portanto } S_B w = \lambda S_w w$$

que é um problema de valores próprios generalizado.

Se $S_w (= \Sigma_0 + \dots + \Sigma_k)$ for não-singular, então

$$\text{obtemos } S_B w = \lambda S_w w \Rightarrow S_w^{-1} S_B w = \lambda w$$

ou seja, w é vector prop. de $S_w^{-1} S_B$ assoc. valor próprio λ . Como $S_B w = \lambda S_w w$ então

$$w^T S_B w = \lambda (w^T S_w w) \quad \text{Maximizar}$$

$w^T S_B w$ é equivalente a maximizar λ , ou seja encontrar o maior valor próprio generalizado.

On seja, encontrar o máximo t_f . [3.5]

$S_B W = \lambda S_W W$. A ~~ndo~~ chamamos vector
próprio generalizado (no 2º sentido) e o valor próprio
generalizado (no 2º sentido).

Ora $S_B W = \lambda S_W W \Rightarrow (\lambda S_W - S_B) W = 0$, p/ algum $W \neq 0$
com $W \neq 0$

onde $\det(\lambda S_W - S_B) = 0$. Os vectores W serão
soluções não nulas de $(\lambda S_W - S_B) X = 0$.

On seja, $W \in \text{Ker}(\lambda S_W - S_B) \setminus \{0\}$.

Se S_B, S_W forem do tipo $n \times n$ e se tivermos
 w_1, \dots, w_n l.i. vectores próprios generalizados, para

$$P = \begin{bmatrix} | & & | \\ w_1 & \dots & w_n \\ | & & | \end{bmatrix}, \quad D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \text{ onde}$$

$$S_B = S_W P D P^{-1}$$

Prova-se que $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ e que

$$w_i^T S_W w_j = 0, \quad \text{se } w_i, w_j \text{ assoc. a } \lambda_i \neq \lambda_j$$

Isto é possível porque S_B é simétrica
e S_W é SPD (de facto S_B tb. é SPD)