LDA

Linear Discriminant Analysis

Pedro Patrício
pedro@math.

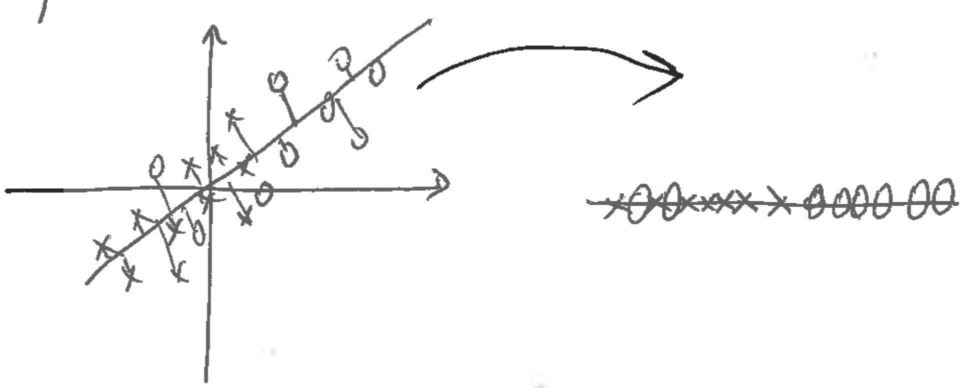
Ronald Fisher, 1936, p/ classificação de algumas flores.

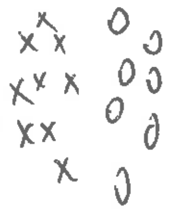
Tanto o PCA como o LDA são técnicas lineares para redução de dimensão. O PCA pode ser encarado como um método não supervisionado, já que ignora a classificação que à priori se tinha das amostras; apenas procura as direcções de maior variância.

O 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.

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





PCA

perde a classificação original

LDA

quando pretendemos algo como

Pretendemos efectuar uma redução da dimensão, de um esp. de dimensão d, para um de dimensão K, com $K < d$.

O valor de K e a redução da dimensão, tal como no PCA, está associado à teoria espectral de matrizes de dispersão.

Vemos dois tipos de matrizes de dispersão: as inter-classes e as intra-classe.

Se os valores próprios forem todos da mesma ordem de grandeza, é um indicador que temos já um "bom" espaço.

Classificação:

Suponhamos que temos amostas "classificadas" como estando em K classes disjuntas, $C_1, \dots, C_K$, cada uma não vazia, e com $C_1 \cup C_2 \dots = C$, o espaço das amostras.

Seja μ_i a média na classe i,

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

e μ a média amostral, $\mu = \frac{1}{N} \sum_{x \in C_1 \cup C_2} x_i$

onde $N = \# C$

Seja S_B a matriz de dispersão inter-classes

$$S_B = \sum_{i=1}^{k} N_i (\mu_i - \mu)(\mu_i - \mu)^T , \quad \text{rank } S_B \leq \cancel{k}-1$$

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 $S_W = \Sigma_1 + \Sigma_2 + \cdots + \Sigma_k$,
onde Σ_i é a matriz de dispersão para a classe C_i.

O método de Fisher consiste em <u>maximizar</u>

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

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:

Definindo $L(w,\lambda) = w^T S_B w - \lambda(w^T S_W w - k')$

$$= w^T (S_B - \lambda S_W) w + \lambda k'$$

Calculando o gradiente em ordem a w,

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

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

obtemos $\quad S_B w = \lambda S_W w \Rightarrow S_W^{-1} S_B w = \lambda w$

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

$$w^T S_B w = \lambda \underbrace{(w^T S_W w)}_{= k' \text{ constante}}$$

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

Ou seja, encontrar λ máximo t.q.

$S_B w = \lambda S_w w$. A $w \neq 0$ chamamos vector próprio generalizado (no 2º sentido) e λ valor próprio generalizado (no 2º sentido).

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

donde $\det(\lambda S_w - S_B) = 0$. Os vectores w serão soluções não nulas de $(\lambda S_w - S_B) x = 0$.

Ou seja, $w \in \mathrm{Ker}(\lambda S_w - S_B) \setminus \{0\}$.

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

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

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

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

$w_i^T S_w w_j = 0$, se w_i, w_j 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)