

Análisis Discriminante

N. Corral, C. Carleos

Departamento de Estadística — Universidad de Oviedo

26 de octubre de 2018

Planteamiento del problema

- ▶ g poblaciones caracterizadas por p variables $X_1, \dots, X_p$
- ▶ objetivo: establecer criterios para determinar la población a la que pertenece un individuo genérico con valores $x_1, \dots, x_p$

Criterio de máxima verosimilitud

- ▶ Sea $L(x_1, \dots, x_p \mid j)$ la verosimilitud de un individuo con valores $x_1, \dots, x_p$ en la población j
- ▶ El criterio asigna a la población J donde se alcanza el máximo de la verosimilitud:

$$L(x_1, \dots, x_p \mid J) = \max_{j=1, \dots, g} L(x_1, \dots, x_p \mid j)$$

Criterio de máxima verosimilitud

- Para dos poblaciones, el criterio se basa en la razón de verosimilitudes:

$$LRV(\vec{x}) = \ln \frac{L(x_1, \dots, x_p \mid 1)}{L(x_1, \dots, x_p \mid 2)}$$

- Se define el clasificador por

$$h(\vec{x}) = \begin{cases} 1 & \text{si } LRV(\vec{x}) > 0 \\ 2 & \text{si } LRV(\vec{x}) < 0 \end{cases}$$

Criterio de máxima verosimilitud: caso gaussiano

- Sean

$$\vec{x} \mid 1 \equiv \mathcal{N}(\vec{\mu}_1, \Sigma_1)$$

$$\vec{x} \mid 2 \equiv \mathcal{N}(\vec{\mu}_2, \Sigma_2)$$

- Caso 1: $\Sigma_1 = \Sigma_2$

$$LRV = (\vec{\mu}_1 - \vec{\mu}_2)' \Sigma^{-1} (\vec{x} - \vec{\mu})$$

donde $\vec{\mu} = \frac{\vec{\mu}_1 + \vec{\mu}_2}{2}$.

Si los parámetros de las distribuciones son desconocidos, se sustituyen por sus estimaciones máximo-verosímiles:

$$\hat{\mu}_1 = \bar{x}_1 \quad \hat{\mu}_2 = \bar{x}_2 \quad S = \frac{n_1}{n} S_1 + \frac{n_2}{n} S_2 \quad (n = n_1 + n_2)$$

Criterio de máxima verosimilitud: caso gaussiano

- Sean

$$\vec{x} \mid 1 \equiv \mathcal{N}(\vec{\mu}_1, \Sigma_1)$$

$$\vec{x} \mid 2 \equiv \mathcal{N}(\vec{\mu}_2, \Sigma_2)$$

- Caso 2: $\Sigma_1 \neq \Sigma_2$

$$LRV = -\frac{1}{2} (\vec{x} - \vec{\mu}_1)' \Sigma_1^{-1} (\vec{x} - \vec{\mu}_1) - \left(-\frac{1}{2}\right) (\vec{x} - \vec{\mu}_2)' \Sigma_2^{-1} (\vec{x} - \vec{\mu}_2)$$

Equivale a asignar el individuo $\vec{x}$ a la población J tal que

$$D^2(\vec{x}, \vec{\mu}_J) = \min_{j=1, \dots, g} D^2(\vec{x}, \vec{\mu}_j)$$

donde D^2 es la distancia de Mahalanobis.

Criterio de Mahalanobis

- La distancia de Mahalanobis es

$$D^2(\vec{x}, \bar{x}_j) = (\vec{x} - \bar{x}_j)' S^{-1} (\vec{x} - \bar{x}_j) \quad (1)$$

$$= \vec{x}' S^{-1} \vec{x} - 2\bar{x}_j' S^{-1} \vec{x} + \bar{x}_j S^{-1} \bar{x}_j \quad (2)$$

- Se asigna el individuo $\vec{x}$ a la población J tal que

$$D^2(\vec{x}, \bar{x}_J) = \min_{j=1, \dots, g} D^2(\vec{x}, \bar{x}_j)$$

- Equivale a maximizar

$$2\bar{x}_j' S^{-1} \vec{x} - \bar{x}_j' S^{-1} \bar{x}_j = \vec{a}_j' \vec{x} - c_j$$

Criterio del factor discriminante de Fisher

- ▶ Buscar combinaciones lineales de las variables originales que maximicen la separación entre los grupos.
- ▶ Se basa en la siguiente descomposición de la variabilidad total:
 - ▶ Sea $\vec{y} = H X \vec{a}$
 - ▶ $\vec{y}' H \vec{y} = \vec{a}' X' H X \vec{a} = \vec{a}' T \vec{a}$
 - ▶ $T = W + B$
- ▶ Se busca el vector $\vec{a}$ que maximice $\frac{\vec{a}' B \vec{a}}{\vec{a}' W \vec{a}}$

Criterio del factor discriminante de Fisher (detalle)

$$\vec{a}' H \vec{a} = \vec{a}' \left(I - \frac{1}{n} J \right) \vec{a} \quad (3)$$

$$= \vec{a}' \left(I - D + D - \frac{1}{n} J \right) \vec{a} \quad (4)$$

$$= \vec{a}' (I - D) \vec{a} + \vec{a}' \left(D - \frac{1}{n} J \right) \vec{a} \quad (5)$$

$$= \text{SCB} - \text{SCW} \quad (6)$$

$$\text{donde } D = \begin{bmatrix} \frac{1}{n_1} J_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \frac{1}{n_g} J_g \end{bmatrix}$$

Criterio bayesiano

- ▶ Sea $L(x_1, \dots, x_p \mid j)$ la verosimilitud de un individuo con valores $x_1, \dots, x_p$ en la población j
- ▶ Sea $\Pr(j)$ la probabilidad *a priori* de la población j
- ▶ El criterio Bayes asigna a la población J donde se alcanza el máximo de la probabilidad *a posteriori*:

$$\Pr(J \mid x_1, \dots, x_p) = \max_{j=1, \dots, g} \Pr(j \mid x_1, \dots, x_p)$$

- ▶ Probabilidad a posteriori:

$$\Pr(j \mid x_1, \dots, x_p) = \frac{L(x_1, \dots, x_p \mid j) \cdot \Pr(j)}{L(x_1, \dots, x_p)}$$

$$\Pr(j \mid x_1, \dots, x_p) \propto L(x_1, \dots, x_p \mid j) \cdot \Pr(j)$$

Criterio bayesiano

- ▶ Permite incluir probabilidades a priori de las poblaciones.
- ▶ Equivale al criterio de MV si equiprobables a priori.
- ▶ No hace falta conocer el denominador $L(x_1, \dots, x_p)$:

$$\Pr(J \mid x_1, \dots, x_p) = \frac{L(x_1, \dots, x_p \mid J) \cdot \Pr(J)}{\sum_{j=1, \dots, g} L(x_1, \dots, x_p \mid j) \cdot \Pr(j)}$$

Criterio bayesiano ingenuo

- ▶ Supone independencia entre las variables:

$$L(x_1, \dots, x_p \mid j) = L(x_1 \mid j) \cdots L(x_p \mid j)$$

- ▶ Se usa si no se dispone de distribución conjunta.
- ▶ Casi siempre poco realista.