Clasificación 2

Sunday, May 8, 2022 6:39 PM

Clasificación 2

Clasificac° con 2 poblac°es multivariadas

Bajo la suposic° de normalidad, la clasificac°n es + sencilla y resulta + práctica (incluso si son aprox normales)

Suponemos $f_1(x)$, $f_2(x)$ PDF's normal multivar con medias μ_1 y μ_2, y cov. Σ_1, Σ_2.

Caso 1: $\Sigma = \Sigma_1 = \Sigma_2$

$$f_i(x) = \frac{1}{(2\pi)^{p/2}\,|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu_i)'\Sigma^{-1}(x-\mu_i)\right)$$

Sup. q' μ_1, μ_2 y Σ conocidos

Teo:

$$R_1: \exp\left(-\frac{1}{2}(x-\mu_1)'\Sigma^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)'\Sigma^{-1}(x-\mu_2)\right) \geq \frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1} \quad \hookrightarrow \frac{f(x_1)}{f(x_2)}$$

$$R_2: \frac{f(x_1)}{f(x_2)} < \frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}$$

Teo: Sean π_1 y π_2 poblac°es con PDF normal multivar. entonces la regla de clasificac° q' minimiza el error es:

x_0 se clasifica como π_1 si:

$$(\mu_1-\mu_2)'\Sigma^{-1}x_0 - \frac{1}{2}(\mu_1-\mu_2)'\Sigma^{-1}(\mu_1+\mu_2) \geq \ln\left(\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}\right)$$

Obs: Esto sirve si conocemos μ_1, μ_2 y Σ, pero en muchos casos no lo conocemos.

Sup. q' hay n_1 obs de $x = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ de π_1

n_2 obs. de $x = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ de π_2 con $n_1+n_2-2 \geq p$

$$x_1 = \begin{bmatrix} x'_{11} \\ \vdots \\ x'_{1n_1} \end{bmatrix} \qquad x_2 = \begin{bmatrix} x'_{21} \\ \vdots \\ x'_{2n_2} \end{bmatrix}$$

con medias muestrales $\bar{x}_1$, $\bar{x}_2$ y cov. muestrales:

$$S_1 = \frac{1}{n_1-1}\sum_{i=1}^{n_1}(x_{1i}-\bar{x}_1)(x_{1i}-\bar{x}_1)' \qquad S_2 = \frac{1}{n_2-1}\sum_{j=1}^{n_2}(x_{2j}-\bar{x}_2)(x_{2j}-\bar{x}_2)'$$

$$\text{Spooled} = \left(\frac{n_1-1}{n_1+n_2-2}\right)S_1 + \left(\frac{n_2-1}{n_1+n_2-2}\right)S_2$$

Est. insesgado de Σ

(Seguimos asumiendo q' es = en π_1 y π_2)

▸ Clasificar x_0 como π_1 si:

$$(\bar{x}_1-\bar{x}_2)'\,\text{Spooled}^{-1}x_0 - \frac{1}{2}(\bar{x}_1-\bar{x}_2)'\,\text{Spooled}^{-1}(\bar{x}_1+\bar{x}_2) \geq \ln\left(\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}\right)$$

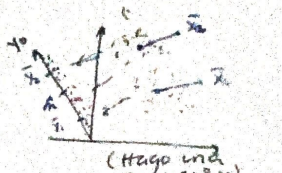
Clasificar x_0 como π_1 si no.

Obs: si $\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1} = 1$ $\quad \ln(1) = 0$.

Sea $\hat{y}_0 = (\bar{x}_1-\bar{x}_2)'\,\text{Spooled}^{-1}x_0 = \hat{a}'x_0$

$$\hat{m} = \frac{1}{2}(\bar{x}_1-\bar{x}_2)'\,\text{Spooled}^{-1}(\bar{x}_1+\bar{x}_2) = \frac{1}{2}(\bar{y}_1+\bar{y}_2)$$

donde $\bar{y}_1 = \hat{a}'\bar{x}_1$, $\bar{y}_2 = \hat{a}'\bar{x}_2$



(Hago una proyec° q' me max la ⊕)

Podemos clasificar comparando $\hat{y}_0$ con $\hat{m}$ (si es mayor o menor del promedio de medias $\bar{y}_1$, $\bar{y}_2$)

si es $\geq \hat{m}$, es π_2

Regla de clasificac° usando f° discriminante de Fisher

Clasifica x_0 como π_1 si.

$$\hat{y}_0 = (\bar{x}_1-\bar{x}_2)'\,\text{Spooled}^{-1}x_0 \geq \hat{m} \quad \text{ó} \quad \hat{y}_0 - \hat{m} \geq 0.$$

si $\hat{y}_0 - \hat{m} < 0$, se clasifica como π_2

Obs: $\hat{y}$ (min error cuando los costos y la prob previas =) es la f° lín de Fisher q' maximiza la separac°.

Caso 2: $\Sigma_1 \neq \Sigma_2$

$$R_1: -\frac{1}{2}x'(\Sigma_1^{-1}-\Sigma_2^{-1})x + (\mu_1'\Sigma_1^{-1}-\mu_2'\Sigma_2^{-1})x - k \geq \ln\left(\frac{c(1|2)}{c(2|1)}\cdot\frac{p_2}{p_1}\right)$$

... es menor, π_2.

$$K = \frac{1}{2} \ln\left(\frac{\det(\Sigma_1)}{\det(\Sigma_2)}\right) + \frac{1}{2}(\mu_1' \Sigma_1^{-1} \mu_1 - \mu_2' \Sigma_2^{-1} \mu_2)$$

Como en la practica no se conocen, estimo μ_1, μ_2, Σ_1 y Σ_2.

Se clasifica x_0 como π_1 (π_2) si:

$$-\frac{1}{2} x_0' (S_1^{-1} - S_2^{-1}) x_0 + (\bar{x}_1' S_1^{-1} - \bar{x}_2' S_2^{-1}) x_0 - K \geq \ln\left(\frac{c(1|2)}{c(2|1)} \frac{p_2}{p_1}\right)$$

S_1 < , π_2.

Evaluación de clasificación

Podemos eval la calidad de la clasificación con tasas de error (proba de clasif. incorrecta).

i) Si se conoce la ~ de la pobla°, es directo.

ii) Si no, usar regla de clasificación muestral y se evalua con muest

Para i) TPM= $p_1 \int_{R_2} f_1(x)\,dx + p_2 \int_{R_1} f_2(x)\,dx$ (Total Proba Mistake).

↳ es el OER Optimar Error Rate

Para ii) Algunos params. desconocidos.

Caso 1: se conocen f_1, f_2

$$\underset{\text{(Actual error rate)}}{AER} = p_1 \int_{\hat{R}_2} f_1(x)\,dx + p_2 \int_{\hat{R}_1} f_2(x)\,dx$$

$\hat{R}_2$ → det. con n_1 y n_2.

Caso 2: f_1, f_2 desconocidas.

APER (apparent error rate) ⟶ matriz de confusión: pob. real

	Pobl 1	Pobl 2
	π_1	π_2
π_1	n_{1c}	$n_{1m} = n_1 - n_{1c}$
π_2	$n_{2m} = n_2 - n_{2c}$	n_{2c}.

$$APER = \frac{n_{1m} + n_{2m}}{n_1 + n_2}$$: (entre menor el APER, mejor ☺).

Se necesitan n_1 y n_2 MUY grandes.

2 grupos ⟶ entrenamiento.

⟶ validación.