Clasificación de varias poblaciones

Wednesday, May 11, 2022 2:18 PM

Clasificación de varias poblaciones

Buscamos generalizar $g \geq 2$

Sea $f_i(x)$ la densidad de π_i

! P_i : Proba previa

! $C(k|i)$: Clasificar como π_k cuando es π_i

$C(i|i) = 0$

$P(k|i) = P(\text{clasif. como } \pi_k \mid \text{siendo } \pi_i) = \int_{R_k} f_i(x)\,dx.$

$$P(i|i) = 1 - \sum_{\substack{k=1\\k\neq i}}^{g} P(k|i)$$

Entonces, el TPM de un x de π_i es:

- $TPM(i) = \sum_{\substack{k=1\\k\neq i}}^{g} P(k|i)\,C(k|i) \longrightarrow$ Minimizar
- $TPM = p_1\,TPM(1) + \cdots + p_g\,TPM(g)$
  $= \sum_{i=1}^{g} p_i \sum_{\substack{k=1\\k\neq i}}^{g} P(k|i)\,C(k|i)$

Para determ. $R_1, \dots, R_g$ mutuam. excluyentes y exhaustivas, buscamos minimizar TPM.

Teo: TPM se minimiza si la regla de clasificaº es:

Clasificar x como π_k $k = 1, \dots, g$ t.q.

$\sum_{\substack{i=1\\k\neq i}}^{g} p_i f_i(x)\,C(k|i)$ es mínima. $\Rightarrow$ si quiero ver en la poblaº i, pruebo en todas las demás. Dónde sea menor la equivocaº, ahí es!

Obs: Si los costos de clasificaº son =:

$\sum_{\substack{i=1\\i\neq k}}^{g} p_i f_i(x) \longrightarrow$ se clasifica x_0 como π_k si:

$p_k f_k(x) \geq p_i f_i(x) \quad \forall i \neq k$

- Proba de asignar a k una nueva obs (x_0):

$P(\pi_k|x_0) = \dfrac{p_k f_k(x_0)}{\sum_{i=1}^{g} p_i f_i(x_0)}$ $\begin{matrix}\longrightarrow \text{Casos favorables} \\ \longrightarrow \text{Casos totales}\end{matrix}$ $\Big\}$ si los costos son =. Si no? TAREA.

- Con costos $\neq$:

Obs: Si las poblaciones son normales con media μ_i y cov Σ_i, y hay costos =.

Clasificar x como π_k si:

$$\ln(p_k f_k(x)) = \ln(p_k) - \frac{p}{2}\ln(2\pi) - \frac{1}{2}\ln|\Sigma_k| - \frac{1}{2}(x-\mu_k)'\,\Sigma_k^{-1}(x-\mu_k)$$

$\longrightarrow$ fº discrimin. $d_k^Q(x)$

Lo clasificamos en la poblaº i, dónde la fº discriminante sea máxima.

$\max_i (\ln(p_i f_i(x)))$

Obs: Si μ_i y Σ_i desconocida, pero tenemos un conjunto de entrenamiento.

- $\bar{x}_i$: media muestral de π_i
- n_i = Tamaño de la muestra π_i
- S_i : Cov muestral de π_i

$$\hat{d}_i^Q(x) = -\frac{1}{2}\ln|S_i| - \frac{1}{2}(x-\bar{x}_i)'S_i^{-1}(x-\bar{x}_i) + \ln(p_i) - \frac{p}{2}\ln(2\pi)$$

Se puede clasificar x_0 como π_k cuando $\hat{d}_k^Q(x) = \max_i \hat{d}_i^Q(x)$.

Si todas las Σ_i son =,

$$d_i^Q(x) = \mu_i'\Sigma^{-1}x - \frac{1}{2}\mu_i'\Sigma^{-1}\mu_i + \ln(p_i)$$

Lo podemos estimar con:

$$\hat{d}_i^Q(x) = \bar{x}_i'\,S_{pooled}^{-1}x - \frac{1}{2}\bar{x}_i'\,S_{pooled}^{-1}\bar{x}_i + \ln(p_i)$$

$$S_{pooled} = \frac{1}{n_1 + \cdots + n_g - g}\left((n_1 - 1)S_1 + \cdots + (n_g - 1)S_2\right)$$

Método discriminante de Fisher

→ No necesita normalidad
→ Necesita varianzas $=$: $\Sigma_1 = \Sigma_2 = \ldots = \Sigma$ invertible

Usar pocas comb. lin $a_1' X_1, \ldots, a_k' X_k$ para lograr clasificar en varias poblaciones (y sirve xa reducir dimensionalidad).

Sea $\bar{\mu}$ la media de la media de las poblaºes combinadas.

$$\bar{\mu} = \frac{1}{g}\sum_{i=1}^{g} \mu_i$$

Sea $$\mathbb{B}_\mu = \sum_{i=1}^{g} (\mu_i - \bar{\mu})(\mu_i - \bar{\mu})'$$

Sea $Y = a' X$ comb lineal

- $E[Y|\pi_i] = a' E[X|\pi_i] = a'\mu$
- $var(Y) = a'\Sigma a$ $\forall$ las poblaºes.

$\mu_{iy} = a'\mu_i$ depende de la poblaº a la q' pertenece.

$$\bar{\mu}_y = \frac{1}{g}\sum_{i=1}^{g} \mu_{iy} = \frac{1}{g}\sum_{i=1}^{g} a'\mu_i = a'\bar{\mu}$$

Definimos $$\frac{\sum_{i=1}^{g}(\mu_{iy} - \bar{\mu}_y)^2}{\sigma_y^2} = \frac{\sum_{i=1}^{g}(a'\mu_i - a'\bar{\mu})}{a'\Sigma a} = \frac{a'\mathbb{B}_\mu a}{a'\Sigma a} \odot$$

var entre grupos en relaº con la var dentro de los grupos.

$\odot$ Busco un vec a que maximice

Si μ_i y Σ desconocidos, y el conjunto de entrenam. tiene n_i obs de π_i.

Los datos de π_i son: $X_i = \begin{bmatrix} x_{i1}' \\ \vdots \\ x_{in_i}' \end{bmatrix}$ con $\bullet\ \bar{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$

- $\bar{X} = \frac{1}{g}\sum_{i=1}^{g} \bar{X}_i$

- $\mathbb{B}$ (estimador $\mathbb{B}_\mu$) $= \sum_{i=1}^{g} (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})'$

Definimos: • $W = \sum_{i=1}^{g}(n_i - 1)S_i = \sum_{i=1}^{g}\sum_{j=1}^{n_i}(X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)'$

- $\frac{W}{n_1 + \cdots + n_g - g} = S_{pooled}$ (estimador de Σ)

Ley discrim. de Fisher: → varianzas $=$.

Sean $\hat{\lambda}_1, \ldots, \hat{\lambda}_s > 0$ los $s \leq \min(g-1, p)$ vals propios de $W^{-1}\mathbb{B}$ asoc. a $\hat{e}_1, \ldots, \hat{e}_s$ vecs propios normalizados t.q' $\hat{e}_i' S_{pooled}\, \hat{e}_i = 1$

Entonces, el a q' maximiza $\frac{\hat{a}'\mathbb{B}\hat{a}}{\hat{a}' W \hat{a}}$ es $\hat{a}_1 = \hat{e}_1$

- $\hat{a}_1' X$ es el 1er discriminante muestral
- $\hat{a}_k' X$ produce el k-ésimo discrim. muestral $(k \leq s)$

Clasificaº discrim. Fisher

Los discriminantes buscan otorgar representaº de baja dimensionalidad q' separe poblaºes lo mejor posible.

$Y_k = a_k' X$ el k-ésimo discrim $(k \leq s)$

$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_s \end{bmatrix}$ tiene media μ_{iy} para π_i

- $Cov(Y) = \mathbb{I}$ TAREA

Luego, $(Y - \mu_{iy})'(Y - \mu_{iy}) = \sum_{j=1}^{g}(y_i - \mu_{iy_i})^2 \odot$ mide las distancias.

Podemos clasificar Y a π_k si $\odot$ es mínimo.

↳ Clasificar X como π_k si:

- $\sum_{j=1}^{r}(y_j - \mu_{ky_j})^2 = \sum_{j=1}^{r}(a_j'(X - \mu_k))^2$

$$= \sum_{j=1} (a_j'(x - \mu_i))^2 \quad \forall i \neq k$$

usando r discriminantes

Obs: si $p_1 \dots p_g = \frac{1}{g}$