Juan Manuel Rodríguez Gómez
49 55 94 94 Z

ESTADÍSTICA MULTIVARIANTE

TAREA VOLUNTARIA 5

① Deducir la ecuación del clasificador discriminante lineal con $n > 1$ predictores.

Dada Y una variable aleatoria de respuesta categórica con $k \geq 2$ niveles y $X = (X_1, \ldots, X_n)$, $n \in \mathbb{N}$, un vector aleatorio continuo de variables explicativas ⊛, se pretende clasificar en los diferentes niveles de Y para valores específicos de X.

⊛ con distribución multivariante Gaussiana y varianza homogénea

Para el caso $n = 1$ ya se tiene la ecuación del clasificador discriminante lineal gracias al Teorema de Bayes: Teniendo en cuenta que $f_k(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu_k)^2}{2\sigma^2}}$ es una densidad Gaussiana (univariante) con media μ_k y varianza homogénea, σ^2, en los k niveles, aplicando logaritmo para linealizar se tiene que, el registro es asignado al nivel $i \in \{1, \ldots, k\}$ si, y solo si,

$$\log\left(\frac{f_i(x)}{f_j(x)}\right) > \log\left(\frac{\pi_j}{\pi_i}\right) \Leftrightarrow \frac{\mu_i - \mu_j}{\sigma^2} x - \frac{\mu_i^2 - \mu_j^2}{2\sigma^2} - \log\left(\frac{\pi_j}{\pi_i}\right) > 0$$

Desarrollemos el caso $n > 1$:

Antes $x \in \mathbb{R}$, ahora $x \in \mathbb{R}^n$

- Se necesita estimar, $\frac{P(Y=i / X=x)}{P(Y=j / X=x)} = \frac{P(Y=i, X=x)}{P(Y=j, X=x)}$, $i, j \in \{1, \ldots, k\}$.

- De acuerdo con la expresión general del Teorema de Bayes y la notación de la diapositiva 8 se tiene que, $\frac{P(Y=i / X=x)}{P(Y=j / X=x)} = \frac{\pi_i P(X=x / Y=i)}{\pi_j P(X=x / Y=j)} = \frac{\pi_i f_i(x)}{\pi_j f_j(x)}$.

- Regla de decisión: Si $\dfrac{\pi_i f_i(x)}{\pi_j f_j(x)} > 1$, o $\dfrac{f_i(x)}{f_j(x)} > \dfrac{\pi_j}{\pi_i}$, entonces, el registro se asigna a la clase o nivel de respuesta i.

- Teniendo en cuenta que $f_k(x) = \dfrac{1}{(2\pi)^{n/2}\,|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu_k)^t\,\Sigma^{-1}(x-\mu_k)\right\}$ es una densidad Gaussiana (multivariante) con vector de medias μ_k y matriz de covarianzas homogénea, Σ, en los k niveles, aplicando logaritmo para linealizar se tiene que, el registro es asignado al nivel $i \in \{1, \dots, k\}$ si, y solo si,

$$\log\left(\frac{f_i(x)}{f_j(x)}\right) > \log\left(\frac{\pi_j}{\pi_i}\right) \iff -\frac{1}{2}(x-\mu_i)^t\,\Sigma^{-1}(x-\mu_i) + \frac{1}{2}(x-\mu_j)^t\,\Sigma^{-1}(x-\mu_j) - \log\left(\frac{\pi_j}{\pi_i}\right) > 0$$

$$\iff \cancel{-\frac{1}{2}x^t\Sigma^{-1}x} + \frac{1}{2}x^t\Sigma^{-1}\mu_i + \frac{1}{2}\mu_i^t\Sigma^{-1}x - \frac{1}{2}\mu_i^t\Sigma^{-1}\mu_i + \cancel{\frac{1}{2}x^t\Sigma^{-1}x} - \frac{1}{2}x^t\Sigma^{-1}\mu_j - \frac{1}{2}\mu_j^t\Sigma^{-1}x + \frac{1}{2}\mu_j^t\Sigma^{-1}\mu_j - \log\left(\frac{\pi_j}{\pi_i}\right) > 0$$

$$\iff \frac{1}{2}x^t\Sigma^{-1}(\mu_i-\mu_j) + \underbrace{\frac{1}{2}(\mu_i-\mu_j)^t\Sigma^{-1}x} + \frac{1}{2}\left[\mu_j^t\Sigma^{-1}\mu_j - \mu_i^t\Sigma^{-1}\mu_i\right] - \log\left(\frac{\pi_j}{\pi_i}\right) > 0$$

$$= \frac{1}{2}x^t(\Sigma^{-1})^t(\mu_i-\mu_j)$$

$$= \frac{1}{2}x^t\Sigma^{-1}(\mu_i-\mu_j) \quad \leftarrow \Sigma \text{ es simétrica} \Rightarrow \Sigma^{-1} \text{ es simétrica} \Rightarrow (\Sigma^{-1})^t = \Sigma^{-1}$$

$$\iff \boxed{x^t\Sigma^{-1}(\mu_i-\mu_j) + \frac{1}{2}\left[\mu_j^t\Sigma^{-1}\mu_j - \mu_i^t\Sigma^{-1}\mu_i\right] - \log\left(\frac{\pi_j}{\pi_i}\right) > 0}$$

Ecuación del clasificador discriminante lineal con $n>1$ predictores

② Deducir la ecuación de un clasificador discriminante si la varianza es heterogénea (clasificador discriminante cuadrático).

Deduzcamos la ecuación del clasificador discriminante cuadrático con $n=1$ predictor (para $n>1$ predictores se procedería de igual forma que en ①, solo que ahora se tendría matriz de covarianzas heterogénea, Σ_k, en los k niveles). ⊛

Dada una variable aleatoria de respuesta categórica con $k \geq 2$ niveles y X una variable aleatoria continua explicativa con distribución univariante Gaussiana y varianza heterogénea (antes, cuando la varianza era homogénea, se tenía que $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma^2$, es decir, todas las clases tenían la misma varianza, σ^2. Ahora, con la varianza heterogénea, no todas las clases tienen que tener la misma varianza), se pretende clasificar en los diferentes niveles de Y para valores específicos de X.

- Se necesita estimar, $\dfrac{P(Y=i/X=x)}{P(Y=j/X=x)} = \dfrac{P(Y=i, X=x)}{P(Y=j, X=x)}$, $i,j \in \{1, \dots, k\}$.

- De acuerdo con el Teorema de Bayes y la notación de la diapositiva 8 se tiene que, $\dfrac{P(Y=i/X=x)}{P(Y=j/X=x)} = \dfrac{\pi_i P(X=x/Y=i)}{\pi_j P(X=x/Y=j)} = \dfrac{\pi_i f_i(x)}{\pi_j f_j(x)}$.

- Regla de decisión: Si $\dfrac{\pi_i f_i(x)}{\pi_j f_j(x)} > 1$, o $\dfrac{f_i(x)}{f_j(x)} > \dfrac{\pi_j}{\pi_i}$, entonces, el registro se asigna a la clase o nivel de respuesta i.

⊛ La ecuación resultante en ese caso sería: $\frac{1}{2}\log\left(\frac{|\Sigma_j|}{|\Sigma_i|}\right) + \frac{1}{2}\left[(x-\mu_j)^t\Sigma_j^{-1}(x-\mu_j) - (x-\mu_i)^t\Sigma_i^{-1}(x-\mu_i)\right] - \log\left(\frac{\pi_j}{\pi_i}\right) > 0$

- Teniendo en cuenta que $f_k(x) = \frac{1}{\sigma_k\sqrt{2\pi}} e^{-\frac{(x-\mu_k)^2}{2\sigma_k^2}}$ es una densidad Gaussiana (univariante) con media μ_k y varianza heterogénea, σ_k^2, en los k niveles, aplicando logaritmo para linealizar se tiene que, el registro es asignado al nivel $i \in \{1, \dots, k\}$ si, y solo si,

$$\log\left(\frac{f_i(x)}{f_j(x)}\right) > \log\left(\frac{\pi_j}{\pi_i}\right) \iff \log\left(\frac{\sigma_j}{\sigma_i}\right) - \frac{(x-\mu_i)^2}{2\sigma_i^2} + \frac{(x-\mu_j)^2}{2\sigma_j^2} - \log\left(\frac{\pi_j}{\pi_i}\right) > 0$$

$$\iff \log\left(\frac{\sigma_j}{\sigma_i}\right) + \frac{(-x^2-\mu_i^2+2x\mu_i)\sigma_j^2 + (x^2+\mu_j^2-2x\mu_j)\sigma_i^2}{2\sigma_i^2\sigma_j^2} - \log\left(\frac{\pi_j}{\pi_i}\right) > 0$$

$$\iff \boxed{\frac{\sigma_i^2-\sigma_j^2}{2\sigma_i^2\sigma_j^2}x^2 + \frac{\mu_i\sigma_j^2-\mu_j\sigma_i^2}{\sigma_i^2\sigma_j^2}x - \frac{\mu_i^2\sigma_j^2-\mu_j^2\sigma_i^2}{2\sigma_i^2\sigma_j^2} + \log\left(\frac{\sigma_j}{\sigma_i}\right) - \log\left(\frac{\pi_j}{\pi_i}\right) > 0}$$

Ecuación del clasificador discriminante cuadrático con $n=1$ predictor