

Análisis discriminante

Daniel Czarniewicz

Descripción general

El análisis discriminante es una técnica con finalidades de descripción (analizar la existencia de diferencias entre grupos), predicción (clasificar nuevas observaciones) y re-clasificación. El problema consiste en construir un modelo que permita discriminar las observaciones según el grupo poblacional al que pertenecen. A la i -ésima observación se le miden p características, las cuales componen el vector $\mathbf{x}'_i = (x_{i1}, x_{i2}, \dots, x_{ip})$. Se asume que existen k grupos en la población.

Reglas de decisión

Existen distintas reglas de decisión para la asignación de observaciones a grupos.

Minimizar la probabilidad de error

La regla de decisión será aquella que minimice la probabilidad total de error. Supongamos que una población P está sub-dividida en k grupos excluyentes. Llamaremos $f_k(x)$ a la densidad de x , si x pertenece al k -ésimo grupo. El objetivo es encontrar una partición del espacio muestral R , tal que asigne x al grupo $k \Leftrightarrow x \in r_x$.

Llamaremos $\Pr(g'|g)$ al error de clasificar en el grupo g' una observación perteneciente al grupo g . Entonces:

$$\Pr(g'|g) = \int_{R_{g'}} f_g(x) dx$$

Por lo tanto, la probabilidad de clasificar erróneamente a todas las observaciones provenientes del grupo g está dada por:

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

De esta forma entonces, la probabilidad total de clasificación errónea está dada por:

$$\Pr(R, f) = \sum_{g=1}^k \pi_g \Pr(g)$$

donde π_g es la probabilidad a priori de que i pertenezca a al grupo g .

Principio de máxima verosimilitud

El principio de clasificación por máxima verosimilitud consiste en asignar la observación i a la población donde el vector observado \mathbf{x}'_i tenga mayor verosimilitud de ocurrir. Es decir, se asigna i al grupo g , sí y solo si:

$$f(\mathbf{x}_i|g) > f(\mathbf{x}_i|g') \quad \forall g' \neq g \Leftrightarrow \Pr(\mathbf{x}_i|g) > \Pr(\mathbf{x}_i|g') \quad \forall g' \neq g \Leftrightarrow \frac{f(\mathbf{x}_i|g)}{f(\mathbf{x}_i|g')} > 1$$

Principio de probabilidad a posteriori

La regla consiste en asignar la observación i a la población con mayor probabilidad a posteriori (la probabilidad de que i pertenezca a g , dado \mathbf{x}_i). Utilizando el Teorema de Bayes, tenemos que la probabilidad a posteriori está dada por:

$$\Pr(i \in g | \mathbf{x} = \mathbf{x}_i) = \frac{\pi_g \Pr(\mathbf{x}_i|g)}{\Pr(\mathbf{x}_i)} = \frac{\pi_g \Pr(\mathbf{x}_i|g)}{\sum_{g'=1}^k \pi_{g'} \Pr(\mathbf{x}_i|g')} = \frac{\pi_g f(\mathbf{x}_i|g)}{\sum_{g'=1}^k \pi_{g'} f(\mathbf{x}_i|g')}$$

De esta forma, la observación i se asignará al grupo g , sí y solo si:

$$\Pr(i \in g | \mathbf{x} = \mathbf{x}_i) > \Pr(i \in g' | \mathbf{x} = \mathbf{x}_i) \quad \forall g' \neq g$$

Normalidad

Si $\mathbf{x}_i \sim N_p(\mu, \Sigma)$ su función de densidad viene dada por:

$$f(\mathbf{x}_i|g) = \frac{1}{(2\pi)^{p/2} |\Sigma_g|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \mu_g)' \Sigma_g^{-1} (\mathbf{x}_i - \mu_g) \right\}$$

La densidad puede estimarse utilizando los estimadores MV de μ_g y Σ_g , $\bar{\mathbf{x}}_g$ y \mathbf{S}_g respectivamente, para obtener:

$$\hat{f}(\mathbf{x}_i|g) = \frac{1}{(2\pi)^{p/2} |\mathbf{S}_g|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x}_i - \bar{\mathbf{x}}_g)' \mathbf{S}_g^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_g) \right\}$$

Si aplicamos el supuesto de normalidad a la probabilidad posteriori, obtenemos que:

$$\Pr(i \in g | \mathbf{x} = \mathbf{x}_i) = \frac{\pi_g |\Sigma_g|^{-1/2} \exp \left\{ (-1/2) D_{ig}^2 \right\}}{\sum_{g'=1}^k \pi_{g'} |\Sigma_{g'}|^{-1/2} \exp \left\{ (-1/2) D_{ig'}^2 \right\}}$$

donde D_{ig}^2 y $D_{ig'}^2$ son la distancia de Mahalanobis entre la observación i y los grupos g y g' respectivamente. Utilizando los estimadores de μ_g y Σ_g mencionadas anteriormente, obtenemos que:

$$\hat{\Pr}(i \in g | \mathbf{x} = \mathbf{x}_i) = \frac{\hat{\pi}_g |\mathbf{S}_g|^{-1/2} \exp \left\{ (-1/2) \hat{D}_{ig}^2 \right\}}{\sum_{g'=1}^k \hat{\pi}_{g'} |\mathbf{S}_{g'}|^{-1/2} \exp \left\{ (-1/2) \hat{D}_{ig'}^2 \right\}}$$

y la observación i se asignará al grupo g , sí, y solo si se cumple que:

$$\hat{\pi}_g |\mathbf{S}_g|^{-1/2} \exp \left\{ (-1/2) \hat{D}_{ig}^2 \right\} > \hat{\pi}_{g'} |\mathbf{S}_{g'}|^{-1/2} \exp \left\{ (-1/2) \hat{D}_{ig'}^2 \right\} \quad \forall g' \neq g$$

Costos

Existen situaciones en las que el error de clasificación es más costo para algunos grupos que para otros. La regla de decisión puede modificarse de forma tal de contemplar estas situaciones de la siguiente forma. Se define un costo para cada error de clasificación, $c(g|g')$. Luego, se asigna i al grupo g sí, y solo si, se cumple que:

$$\frac{f(\mathbf{x}_i|g)}{f(\mathbf{x}_i|g')} > \frac{\pi_{g'} c(g|g')}{\pi_g c(g'|g)} \quad \forall g' \neq g$$

Errores de clasificación

Tasa de error aparente

Luego de elegida una regla de clasificación, se utilizan los n datos para construir la función discriminante y clasificar las observaciones. Una vez clasificadas, se calcula la *tasa de error aparente*,

$$e_{i,app} = \frac{m_i}{n_i}$$

donde m_i es la cantidad de observaciones clasificadas errónemente, de las n_i observaciones asignadas al grupo g_i .

LOOCV

Leave-one-out cross-validation consiste en:

- apartar una observación de la muestra.
- construir la función discriminante con las $n - 1$ observaciones restantes.
- clasificar la observación apartada y registrar si dicha observación fue correcta o incorrectamente clasificada.

- repetir para cada una de las n observaciones

La proporción de observaciones mal clasificadas dentro de cada grupo se define como:

$$e_{i,c} = \frac{a_i}{n_i}$$

Funciones discriminantes

AD factorial

AD probabilístico (distribución normal)

AD probabilístico (distribución desconocida)

Referencias

Beygelzimer, Alina, Sham Kakadet, John Langford, Sunil Arya, David Mount, and Shengqiao Li. 2018. *FNN: Fast Nearest Neighbor Search Algorithms and Applications*. <https://CRAN.R-project.org/package=FNN>.

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An Introduction to Statistical Learning*. Vol. 112. Springer.

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.

Rencher, Alvin C. 1998. *Multivariate Statistical Inference and Applications*. Wiley New York.

Wasserman, Larry. 2007. *All of Nonparametric Statistics*. Springer, New York.

Wickham, Hadley. 2017. *Tidyverse: Easily Install and Load the 'Tidyverse'*. <https://CRAN.R-project.org/package=tidyverse>.