Análisis discriminante

Daniel Czarnievicz

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 $\mathbf{x}_i' = (x_{i1}, x_{i2}, \ldots, x_{ip})$. 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 minimize 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:

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

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 pertenzca a al grupo g.

Principio de máxima verosimilitud

El pricipio 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:

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

Principio de probabilidad a posteriori

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

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

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

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

Normalidad

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

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

La densidad puede estimarse utilizando los estimadores MV de μ_g y $\mathbf{\Sigma}_g$, $\bar{\mathbf{x}}_g$ 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:

$$\Pr(i \in g|\mathbf{x} = \mathbf{x}_i) = \frac{\pi_g \; |\mathbf{\Sigma}_g|^{-1/2} \exp\left\{ (-1/2)\, D_{ig}^2 \right\}}{\sum\limits_{g'=1}^{k} \pi_{g'} \; |\mathbf{\Sigma}_{g'}|^{-1/2} \exp\left\{ (-1/2)\, D_{ig'}^2 \right\}}$$

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 $\mathbf{\Sigma}_g$ mencionadas anteriormente, obtenemos que:

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

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

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

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.