

Análisis discriminante

MANADINE – Análisis de Datos 1

10 de marzo de 2025

Se trata de una técnica de clasificación supervisada, es decir, en la que el modelo se calcula a partir de muestras que contienen información de variables independientes para individuos cuya clase o grupo es conocida de antemano. Se trata de una técnica interpretable, que permite entender por qué cierto individuo se clasifica en determinado grupo, en contraposición a las técnicas de tipo *caja negra* (como las redes neuronales).

El análisis discriminante puede fundamentarse en dos criterios distintos, que son equivalentes en ciertas condiciones.

Preliminares

Supóngase que existen g grupos. Sea i el índice de grupo, con $i \in \{1, \dots, g\}$. Sea n_i el tamaño muestral del grupo i . Sea \mathbf{x}_{ij} el vector de valores de las variables independientes para cierto individuo j perteneciente al grupo i ; por tanto, $j \in \{1, \dots, n_i\}$.

Sea $\bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij}$ la media del grupo i .

Sea \mathbf{X}_i la matriz de observaciones del grupo i , con los individuos por filas y las variables independientes por columnas. La matriz de covarianzas del grupo i es

$$\mathbf{S}_i = \frac{1}{n_i} \mathbf{X}_i \mathbf{H} \mathbf{X}_i^t$$

donde $\mathbf{H} = \mathbf{I} - \mathbf{J} = \mathbf{I} - \frac{1}{n_i} \mathbf{1}\mathbf{1}^t$ es una matriz simétrica e idempotente conocida como *matriz de centrado*, porque al premultiplicar \mathbf{X}_i por \mathbf{H} el resultado es que las variables se centran (se les resta la media).

Criterio de menor distancia de Mahalanobis

La distancia de Mahalanobis entre el vector \mathbf{x} y el grupo i es:

$$D^2(\mathbf{x}, \bar{\mathbf{x}}_i) = (\mathbf{x} - \bar{\mathbf{x}}_i)^t \mathbf{S}_i^{-1} (\mathbf{x} - \bar{\mathbf{x}}_i)$$

donde \mathbf{S}_i es la matriz de covarianzas del grupo i y $\bar{\mathbf{x}}_i$ es su vector de medias.

Si se supone la igualdad de las matrices de varianzas-covarianzas de los g grupos:

$$D^2(\mathbf{x}, \bar{\mathbf{x}}_i) = (\mathbf{x} - \bar{\mathbf{x}}_i)^t \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}_i) = \mathbf{x}^t \mathbf{S}^{-1} \mathbf{x} - 2\bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \mathbf{x} + \bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \bar{\mathbf{x}}_i$$

Como $\mathbf{x}^t \mathbf{S}^{-1} \mathbf{x}$ es igual en todos los grupos se puede eliminar. Por otra parte, minimizar la anterior expresión equivale a maximizarla cambiando su signo, es decir, se busca el máximo en i de

$$2\bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \mathbf{x} - \bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \bar{\mathbf{x}}_i = \mathbf{a}_i^t \mathbf{x} - c_i$$

con $\mathbf{a}_i^t = 2\bar{\mathbf{x}}_i^t \mathbf{S}^{-1}$ y $c_i = \bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \bar{\mathbf{x}}_i$. Se trata de las llamadas *funciones de clasificación lineal*.

Dadas la matriz $\mathbf{A} = (\mathbf{a}_1, \dots, \mathbf{a}_g)$ y el vector $\mathbf{c}^t = (c_1, \dots, c_g)$, para obtener las expresiones anteriores para cada individuo hay que calcular la matriz \mathbf{V} , que tiene por filas las puntuaciones de cada individuo para los diferentes grupos.

$$\mathbf{V} = \mathbf{X}\mathbf{A} - \mathbf{1}\mathbf{c}^t$$

El siguiente paso es calcular el máximo de la matriz \mathbf{A} por filas y asignar el individuo correspondiente a la columna (grupo) donde se alcanza dicho máximo.

Factor discriminante de Fisher

En esta técnica se buscan nuevas variables \mathbf{y} llamadas *factores* y definidas como combinaciones de las variables originales centradas que maximicen la separación entre los grupos respecto a la variabilidad intra-grupos, es decir,

$$\mathbf{y} = \mathbf{X}\mathbf{a}$$

La variabilidad total de la nueva variable \mathbf{y} vale:

$$\begin{aligned} \sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{..})^2 &= \sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i.})^2 + \sum_{i=1}^g \sum_{j=1}^{n_i} (\bar{y}_{i.} - \bar{y}_{..})^2 \\ &= \text{variabilidad intra-grupos} + \text{variabilidad entre grupos} \end{aligned}$$

En términos matriciales, esto se puede expresar como

$$(\mathbf{y} - \bar{\mathbf{y}})^t (\mathbf{y} - \bar{\mathbf{y}}) = \mathbf{a}^t \mathbf{T} \mathbf{a} = \mathbf{a}^t \mathbf{W} \mathbf{a} + \mathbf{a}^t \mathbf{B} \mathbf{a}$$

donde $\mathbf{T} = \mathbf{X}^t \mathbf{H} \mathbf{X}$, $\mathbf{W} = \mathbf{X}^t (\mathbf{I} - \mathbf{D}) \mathbf{X}$, $\mathbf{B} = \mathbf{X}^t (\mathbf{D} - \mathbf{J}) \mathbf{X}$, con $\mathbf{J} = \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^t$, $n = \sum_{i=1}^g n_i$ y \mathbf{D} es una matriz diagonal por bloques, donde cada bloque contiene una $\mathbf{J}_i = \frac{1}{n_i} \mathbf{1}_{n_i} \mathbf{1}_{n_i}^t$, con $i = 1, \dots, g$.

Se busca el vector \mathbf{a} que verifique

$$\max_{\mathbf{a}} \frac{\mathbf{a}^t \mathbf{B} \mathbf{a}}{\mathbf{a}^t \mathbf{W} \mathbf{a}}$$

Se tiene que

$$\frac{\mathbf{a}^t \mathbf{B} \mathbf{a}}{\mathbf{a}^t \mathbf{W} \mathbf{a}} = \frac{\mathbf{a}^t \mathbf{W}^{1/2} \mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2} \mathbf{W}^{1/2} \mathbf{a}}{\mathbf{a}^t \mathbf{W}^{1/2} \mathbf{W}^{1/2} \mathbf{a}} = \frac{\mathbf{a}^t \mathbf{W}^{1/2}}{\|\mathbf{W}^{1/2} \mathbf{a}\|} \mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2} \frac{\mathbf{W}^{1/2} \mathbf{a}}{\|\mathbf{W}^{1/2} \mathbf{a}\|}$$

Con el cambio $\mathbf{b} = \mathbf{W}^{1/2} \mathbf{a}$ se tiene que

$$\max_{\mathbf{a}} \frac{\mathbf{a}^t \mathbf{B} \mathbf{a}}{\mathbf{a}^t \mathbf{W} \mathbf{a}} = \max_{\mathbf{b}} \frac{\mathbf{b}^t}{\|\mathbf{b}\|} \mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2} \frac{\mathbf{b}}{\|\mathbf{b}\|} = \max_{\|\mathbf{b}\|=1} \mathbf{b}^t \mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2} \mathbf{b}$$

El vector unitario \mathbf{b} que maximiza esa última expresión es un vector propio de $\mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2}$ asociado a su mayor valor propio (véase apéndice). Entonces el vector $\mathbf{a} = \mathbf{W}^{-1/2} \mathbf{b}$ es un vector propio de la matriz $\mathbf{W}^{-1} \mathbf{B}$ asociado su mayor valor propio, ya que sus autovalores coinciden con los de $\mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2}$, pues si $\mathbf{v} = \mathbf{W}^{-1/2} \mathbf{u}$ entonces

$$\mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2} \mathbf{u} = \lambda \mathbf{u} \implies \mathbf{W}^{-1/2} \mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2} \mathbf{u} = \lambda \mathbf{W}^{-1/2} \mathbf{u} \implies \mathbf{W}^{-1} \mathbf{B} \mathbf{v} = \lambda \mathbf{v}$$

Repetiendo el proceso se obtienen nuevas variables que son ortogonales y conservan de manera óptima la variabilidad entre grupos en un número reducido de dimensiones. Con estas nuevas variables se hace la clasificación calculando la distancia de un individuo a cada grupo y asignándolo al más próximo.

Para obtener estos factores es necesario hacer los siguientes pasos:

1. Se calculan los valores propios y vectores propios de la matriz $\mathbf{W}^{-1/2} \mathbf{B} \mathbf{W}^{-1/2}$.
2. Se obtiene los cocientes de variación explicada $\frac{\sum_{i=1, \dots, q} \lambda_i}{\sum_{i=1, \dots, g} \lambda_i}$ y se elige el número de factores.

3. Los vectores solución son $\mathbf{U} = \mathbf{W}^{-1/2}\mathbf{V}$, donde \mathbf{V} son los vectores propios calculados previamente.
4. Se obtiene las puntuaciones factoriales $\mathbf{F} = \mathbf{X}\mathbf{U}$
5. Se calculan las medias de cada grupo en las nuevas variables \mathbf{F}

Para asignar los individuos a los grupos es necesario calcular sus distancias a las medias de los grupos y ver dónde se alcanza el mínimo.

Probabilidades a posteriori

Por la regla de Bayes,

$$\begin{aligned} \Pr(\text{grupo } i \mid \mathbf{x}) &= \frac{\Pr(\text{grupo } i) \cdot \mathcal{L}(\mathbf{x} \mid \text{grupo } i)}{\sum_{k=1}^g \Pr(\text{grupo } k) \cdot \mathcal{L}(\mathbf{x} \mid \text{grupo } k)} \\ &\propto \Pr(\text{grupo } i) \cdot \mathcal{L}(\mathbf{x} \mid \text{grupo } i) \end{aligned}$$

donde $\Pr(\text{grupo } i)$ son las probabilidades a priori y \mathcal{L} es la densidad o verosimilitud supuesta la pertenencia a cierto grupo.

En caso de gaussianidad homoscedástica, $\mathbf{x}_i \hookrightarrow N(\mu_i, \Sigma)$ y

$$\begin{aligned} \mathcal{L}(\mathbf{x} \mid \text{grupo } i) &= \text{cte.} \cdot e^{-\frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_i)^t \mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_i)} \\ \ln \mathcal{L}(\mathbf{x} \mid \text{grupo } i) &= \text{cte.} - \frac{1}{2}(\mathbf{x} - \bar{\mathbf{x}}_i)^t \mathbf{S}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_i) \\ &= \text{cte.} - \frac{1}{2}[\mathbf{x}^t \mathbf{S}^{-1} \mathbf{x} - \mathbf{x}^t \mathbf{S}^{-1} \bar{\mathbf{x}}_i - \bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \mathbf{x} + \bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \bar{\mathbf{x}}_i] \\ &= \text{cte.} - \frac{1}{2}[\mathbf{x}^t \mathbf{S}^{-1} \mathbf{x} - 2\bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \mathbf{x} + \bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \bar{\mathbf{x}}_i] \\ &= \text{cte.} - \frac{1}{2}[-2\bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \mathbf{x} + \bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \bar{\mathbf{x}}_i] \\ &= \text{cte.} + \bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \mathbf{x} - \frac{1}{2} \bar{\mathbf{x}}_i^t \mathbf{S}^{-1} \bar{\mathbf{x}}_i \\ \ln \Pr(\text{grupo } i \mid \mathbf{x}) &= \ln \Pr(\text{grupo } i) + \ln \mathcal{L}(\mathbf{x} \mid \text{grupo } i) + \text{cte.} \end{aligned}$$

Las probabilidades a priori pueden seguir diversos criterios:

- grupos equiprobables: $\Pr(\text{grupo } i) = 1/g$
- probabilidad proporcional al tamaño muestral (por omisión en R): $\Pr(\text{grupo } i) = n_i/n$

Apéndice

Si \mathbf{M} es una matriz simétrica entonces $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^t$ con $\mathbf{\Lambda}$ matriz diagonal de autovalores ordenados de mayor a menor y \mathbf{U} matriz con autovectores por columnas; entonces

$$\max_{\|\mathbf{a}\|=1} \mathbf{a}^t \mathbf{M} \mathbf{a} = \max_{\|\mathbf{a}\|=1} \mathbf{a}^t \mathbf{U} \mathbf{\Lambda} \mathbf{U}^t \mathbf{a} = \max_{\|\mathbf{b}\|=1} \mathbf{a}^t \mathbf{\Lambda} \mathbf{a} = \lambda_{\text{máx}} \implies \mathbf{b} = (1, 0, \dots, 0)^t \implies \mathbf{a} = \mathbf{U} \mathbf{b} = \mathbf{u}_{\text{máx}}$$