TP Técnicas de reducción y visualización de datos

Jesica Charaf e Ignacio Spiousas

12 de septiembre de 2023

El problema (2.9)

Sea $\mathbf{x} \in \mathbb{R}^2$ un vector aleatorio y G una variable aleatoria discreta con rango $1, 2$ y probabilidades $P(G = 1) = \pi_1 = \frac{3}{4}$ y $P(G = 2) = \pi_2 = \frac{1}{4}$. Los vectores condicionados $\mathbf{x}|G = 1$ y $\mathbf{x}|G = 2$ tienen distribución Normal con medias $\boldsymbol{\mu}_1 = (1, \frac{1}{2})^t$ y $\boldsymbol{\mu}_2 = (-\frac{1}{2}, 1)^t$ respectivamente, y la misma varianza:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$$

a. Escribir la función de densidad de x. Calcular $\boldsymbol{\mu}_\mathbf{x}$ y $\boldsymbol{\Sigma}_\mathbf{x}$ e identificar las componentes de varianza dentro de grupos y entre grupos.

Cuando la variable $\mathbf{x}$ está condicionada a la pertenencia al grupo G, tiene una distribución normal multivariada de acuerdo a:

$$\mathbf{x}|G = 1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$$
$$\mathbf{x}|G = 2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$$

Entonces, la función de densidad va a ser una combinación lineal de la densidad de dos normales multivariadas $\mathcal{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$ pesadas por su correspondiente π_j con $j = 1, 2$.

$$\begin{aligned} f_\mathbf{x}(\boldsymbol{x}) &= f_{\mathbf{x}|G=1}(\boldsymbol{x})P(G = 1) + f_{\mathbf{x}|G=2}(\boldsymbol{x})P(G = 2) \\ &= \sum_{j=1}^{2} \frac{1}{2\pi[det(\boldsymbol{\Sigma})]^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)} \pi_j \\ &= \frac{[det(\boldsymbol{\Sigma})]^{-1/2}}{2\pi} \sum_{j=1}^{2} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)} \pi_j, \end{aligned}$$

donde el $[det(\boldsymbol{\Sigma})]^{-1/2} = [3/4]^{-1/2} = 2/\sqrt{3}$. De este modo, la función de densidad resulta:

$$\begin{aligned} f_\mathbf{x}(\boldsymbol{x}) &= \frac{1}{\pi\sqrt{3}} \sum_{j=1}^{2} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_j)^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_j)} \pi_j \\ &= \frac{1}{\pi\sqrt{3}} \left(\frac{3}{4} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_1)} + \frac{1}{4} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_2)^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_2)} \right), \end{aligned}$$

donde $\boldsymbol{\mu}_1 = (1, \frac{1}{2})^t$, $\boldsymbol{\mu}_2 = (-\frac{1}{2}, 1)^t$ y

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{pmatrix}.$$

Luego, calculamos $\boldsymbol{\mu}_{\mathbf{x}}$ a partir de un promedio pesado por π_j:

$$\begin{aligned}\boldsymbol{\mu}_x &= \sum_{i=1}^{2} \pi_i \boldsymbol{\mu}_i \\ &= \frac{3}{4}\left(1, \frac{1}{2}\right)^t + \frac{1}{4}\left(-\frac{1}{2}, 1\right)^t. \\ &= \left(\frac{5}{8}, \frac{5}{8}\right)^t\end{aligned}$$

La componente de varianza *within* ($\boldsymbol{\Sigma}_w$) se calcula también directamente como una combinación lineal de las varianzas de cada distribución:

$$\boldsymbol{\Sigma}_w = \sum_{i=1}^{2} \pi_i \boldsymbol{\Sigma}_i = \boldsymbol{\Sigma} \sum_{i=1}^{2} \pi_i = \boldsymbol{\Sigma} = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix}$$

Como en este caso la varianza de ambos grupos es igual, el promedio pesado es igual a ellas también.

Por otro lado, la componente de varianza *between* ($\boldsymbol{\Sigma}_b$) está relacionada con la distancia entre cada vector $\boldsymbol{\mu}_j$ y el promedio pesado $\boldsymbol{\mu}_{\mathbf{x}}$.

$$\begin{aligned}\boldsymbol{\Sigma}_b &= \sum_{i=1}^{2} \pi_i (\boldsymbol{\mu}_i - \boldsymbol{\mu}_{\mathbf{x}})(\boldsymbol{\mu}_i - \boldsymbol{\mu}_{\mathbf{x}})^t \\ &= \frac{3}{4}\left(\begin{pmatrix} 1 \\ 1/2 \end{pmatrix} - \begin{pmatrix} 5/8 \\ 5/8 \end{pmatrix}\right)\left(\left(1, \frac{1}{2}\right) - \left(\frac{5}{8}, \frac{5}{8}\right)\right) + \frac{1}{4}\left(\begin{pmatrix} -1/2 \\ 1 \end{pmatrix} - \begin{pmatrix} 5/8 \\ 5/8 \end{pmatrix}\right)\left(\left(-\frac{1}{2}, 1\right) - \left(\frac{5}{8}, \frac{5}{8}\right)\right) \\ &= \begin{pmatrix} 0.422 & -0.141 \\ -0.141 & 0.047 \end{pmatrix}.\end{aligned}$$

Y $\boldsymbol{\Sigma}_{\mathbf{x}}$ es la suma de ambas componentes:

$$\boldsymbol{\Sigma}_{\mathbf{x}} = \boldsymbol{\Sigma}_w + \boldsymbol{\Sigma}_b = \begin{pmatrix} 1 & 1/2 \\ 1/2 & 1 \end{pmatrix} + \begin{pmatrix} 0.422 & -0.141 \\ -0.141 & 0.047 \end{pmatrix} = \begin{pmatrix} 1.422 & 0.359 \\ 0.359 & 1.047 \end{pmatrix}$$

b. Sea $\mathbf{z} = (Z_1, Z_2)$ el vector de coordenadas discriminantes. Hallar la transformación de la forma $\mathbf{z} = \mathbf{A}^t(\mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}})$ necesaria para obtenerlo.

Vamos a buscar la matriz de transformación $\mathbf{A}$ a partir del cálculo de los autovectores y autovalores de la matriz $\mathbf{B}$, donde $\mathbf{B} = (\mathbf{C}^{-1})^t \boldsymbol{\Sigma}_b \mathbf{C}^{-1}$ y $\mathbf{C}$ es tal que $\boldsymbol{\Sigma}_w = \mathbf{C}^t\mathbf{C}$. Es decir, $\mathbf{C}$ es la "raiz cuadrada" de $\boldsymbol{\Sigma}_w$. A partir de la descomoposición espectral de $\boldsymbol{\Sigma}_w = \mathbf{U}\Lambda\mathbf{U}^t$ obtenemos $\mathbf{C} = \mathbf{U}\Lambda^{1/2}\mathbf{U}^t$:

$$\mathbf{C} = \begin{pmatrix} 0.966 & 0.259 \\ 0.259 & 0.966 \end{pmatrix}.$$

Ahora, calculamos $\mathbf{B}$:

$$\mathbf{B} = \begin{pmatrix} 0.623 & -0.344 \\ -0.344 & 0.190 \end{pmatrix}.$$

Luego, calculamos los autovectores $\boldsymbol{\beta}_j$ y sus respectivos autovalores λ_j:

$$\begin{aligned}\boldsymbol{\beta}_1 &= (-0.875, 0.483)^t \\ \boldsymbol{\beta}_2 &= (-0.483, -0.875)^t \\ \lambda_1 &= 0.813 \\ \lambda_2 &= 0\end{aligned}$$

Una vez que tenemos los vectores $\boldsymbol{\beta}_j$ podemos calcular los $\boldsymbol{\alpha}_j$ con la transformación $\boldsymbol{\alpha}_j = \mathbf{C}^{-1}\boldsymbol{\beta}_j$:

$$\begin{aligned}\boldsymbol{\alpha}_1 &= (-1.121, 0.801)^t \\ \boldsymbol{\alpha}_2 &= (-0.277, -0.832)^t.\end{aligned}$$

Entonces, la matriz de transformación $\mathbf{A}$ resulta:

$$\mathbf{A} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2) = \begin{pmatrix} -1.121 & -0.277 \\ 0.801 & -0.832 \end{pmatrix}$$

Finamlente, esta es la matriz $\mathbf{A}$ necesaria para la transformación $\mathbf{z} = \mathbf{A}^t(\mathbf{x} - \boldsymbol{\mu}_x)$.

c. Calcular las coordenadas de los centroides $\boldsymbol{\nu}_1$ y $\boldsymbol{\nu}_2$. Mostrar que la distancia euclídea entre ellos corresponde a la distancia de Mahalanobis entre $\boldsymbol{\mu}_1$ y $\boldsymbol{\mu}_2$ en el espacio original.

Los $\boldsymbol{\nu}_j$ son las medias de las nuevas variables $\mathbf{z}$, por lo tanto, $\boldsymbol{\nu}_j = \mathbf{A}^t\boldsymbol{\mu}_j$:

$$\begin{aligned}\boldsymbol{\nu}_1 &= (-0.721, -0.693)^t \\ \boldsymbol{\nu}_2 &= (1.361, -0.693)^t\end{aligned}$$

Ahora calculemos la distancia euclídea de los $\boldsymbol{\nu}_j$ y la distancia de Mahalanobis de los $\boldsymbol{\mu}_j$:

$$\begin{aligned}d_{euclidea} &= (\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2)^t(\boldsymbol{\nu}_1 - \boldsymbol{\nu}_2) = 4.333 \\ d_{Mahalanobis} &= (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^t\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) = 4.333\end{aligned}$$

Podemos ver que efectivamente dan lo mismo.

d. Calcular $\boldsymbol{\Sigma}_\mathbf{z}$ e identificar las componentes de varianza dentro de grupos y entre grupos. ¿Qué observa?

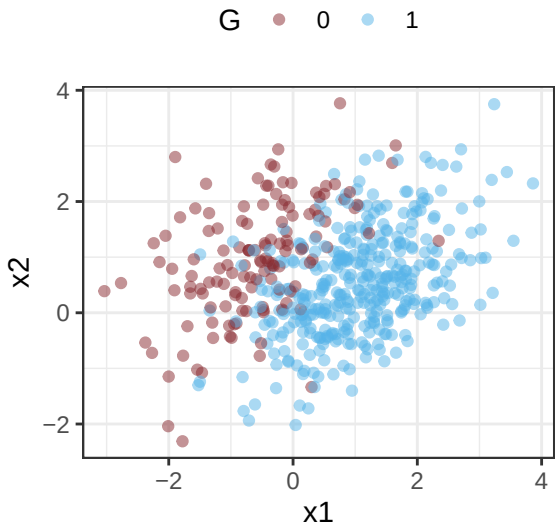
Para obtener la varianza y sus componentes de la variable transformada haremos uso de la propiedad que relaciona la varianza sin transformar con su versión transformada:

$$\begin{aligned}\boldsymbol{\Sigma}_\mathbf{z} &= \mathbf{A}^t\boldsymbol{\Sigma}\mathbf{A} = \begin{pmatrix} 1.813 & 0 \\ 0 & 1 \end{pmatrix} \\ \boldsymbol{\Sigma}_{\mathbf{z},w} &= \mathbf{A}^t\boldsymbol{\Sigma}_w\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ \boldsymbol{\Sigma}_{\mathbf{z},b} &= \mathbf{A}^t\boldsymbol{\Sigma}_b\mathbf{A} = \begin{pmatrix} 0.813 & 0 \\ 0 & 0 \end{pmatrix}\end{aligned}$$

Se observa que $\boldsymbol{\Sigma}_{\mathbf{z},w}$ es la identidad y $\boldsymbol{\Sigma}_{\mathbf{z},b}$ es la matriz de autovalores.

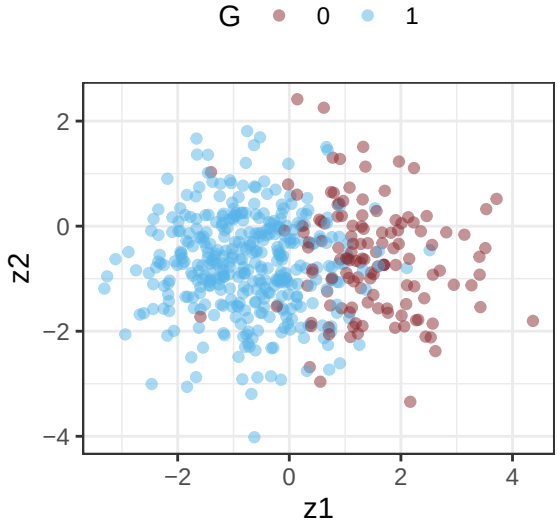
e. Simular $n = 500$ observaciones del vector x. Visualizarlas en un scatterplot con colores distintos según el valor de G.

A continuación se muestra el *scatterplot* obtenido a partir de las 500 observaciones del vector $\mathbf{x}$:



f. Aplicar la transformación hallada para obtener las coordenadas discriminantes y visualizarlas en otro gráfico. ¿Qué observa?

Se observa que en las nuevas coordenadas los grupos están "mejor" diferenciados y que hay un efecto de estandarización multivariada, es decir, los datos dentro de cada grupo están distribuidos en forma de "pelota". Algo esperable ya que la matriz $\mathbf{\Sigma}_{\mathbf{z},w}$ es la identidad.



BONUS

En base a los datos simulados vamos a demostrar que la transformación de coordenadas discriminantes que obtuvimos es equivalente a la transformación de correlación canónica si hubiéramos considerado a un vector $\mathbf{y}$ que contenga al grupo de pertenencia G.

Empecemos obteniendo la matriz $\mathbf{A}$ estimada a partir de los datos. El procedimiento es el mismo que el descrito anteriormente pero estimando $\boldsymbol{\mu}_j$ por $\hat{\boldsymbol{\mu}}_j$ y $\boldsymbol{\Sigma}$ (y sus componentes) por $\hat{\boldsymbol{\Sigma}}$. De esta forma llegamos a una matriz de transformación $\hat{\mathbf{A}}_{cd}$ (el cd por coordenadas discriminantes, para diferenciarlo del de correlación canónica) estimado:

$$\hat{\mathbf{A}}_{cd} = \begin{pmatrix} -1.174 & -0.298 \\ 0.802 & -0.893 \end{pmatrix}$$

Ahora vamos a crear un vector $\mathbf{y}$ que codifique el grupo de pertenencia G, a construir la matriz $\hat{\boldsymbol{\Upsilon}} = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{xy}}\hat{\boldsymbol{\Sigma}}_{\mathbf{y}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{xy}}^{t}$ y a calcular la matriz de proyección de correlación canónica $\hat{\mathbf{A}}_{cc}$ a partir de los autovectores de la matriz $\hat{\boldsymbol{\Upsilon}}$:

$$\hat{\mathbf{A}}_{cc} = \begin{pmatrix} 0.844 & 0.234 \\ -0.519 & 0.907 \end{pmatrix}$$

La matriz no parece ser la misma, pero si graficamos las proyecciones obtenidas con los dos métodos vemos que, si bien hay un cambio de escala (incluido un cambio de signo en u_1), las coordenadas representan lo mismo.