Análisis Discriminatorio Lineal

INTRODUCCIÓN

El análisis discriminatorio lineal (LDA) es conocido por sus siglas en inglés (LDA, linear discriminant analysis) se basa en el discriminatorio lineal de Fisher, un método estadístico desarrollado por Sir Ronald Fisher en la década de 1930 y simplificado posteriormente por C. R. Rao como versión de múltiples clases. El objetivo principal de reducción de dimensión es reducir las dimensiones de los datos eliminando la redundancia y datos dependientes entre sí. Existen dos formas de hacer esto, una supervisada y otra no supervisada.

Algunas técnicas no supervisadas son Independent Component Analysis (ICA), Non-negative Matrix Factorization (NMF), y Principal Component Analysis (PCA). Esta última es la más importante. Técnicas supervisadas son Mixture Discriminant Analysis (MDA), Neural Networks (NN), y Linear Discriminant Analysis (LDA). Esta última es la más importante.

El método de Fisher reduce las dimensiones separando clases de datos proyectados. Separación significa maximizar la distancia entre las medias proyectadas y minimizar la varianza proyectada dentro de las clases.

El Análisis Discriminante Lineal es un método de clasificación supervisado en el que dos o más grupos son conocidos *a priori* y nuevas observaciones se clasifican en uno de ellos en función de sus características. Haciendo uso del teorema de Bayes, *LDA* estima la probabilidad de que una observación, dado un determinado valor de los predictores, pertenezca a cada una de las clases, $P\{Y = k | X = x\}$. Se asigna la observación a la clase *k* para la que la probabilidad es mayor.

El *LDA* presenta una serie de ventajas:

- Si las clases están bien separadas, los parámetros estimados en el método de *LDA* son estables.
- Si el número de observaciones es bajo y la distribución de los predictores es aproximadamente normal en cada una de las clases, *LDA* es muy estable.

El proceso de un análisis discriminante puede resumirse en 6 pasos:

- Disponer de un conjunto de datos de entrenamiento (*training data*) en el que se conoce a que grupo pertenece cada observación.
- Calcular las probabilidades previas (*prior probabilities*): la proporción esperada de observaciones que pertenecen a cada grupo.

- Determinar si la varianza o matriz de covarianzas es homogénea en todos los grupos. De esto dependerá que se emplee *LDA*.
- Estimar los parámetros necesarios para las funciones de probabilidad condicional, verificando que se cumplen las condiciones para hacerlo.
- Calcular el resultado de la función discriminante. El resultado de esta determina a qué grupo se asigna cada observación.
- Utilizar validación cruzada (*cross-validation*) para estimar las probabilidades de clasificaciones erróneas.

Las estimaciones empleadas en *LDA* son tres: la media de las observaciones del grupo (clase) *k*, la media ponderada de las varianzas muestrales de las *K* clases y la proporción de observaciones de la clase *k* respecto al tamaño total de la muestra.

Se ilustrará un ejemplo simple de LDA, ya que muchos conceptos son intuitivos.

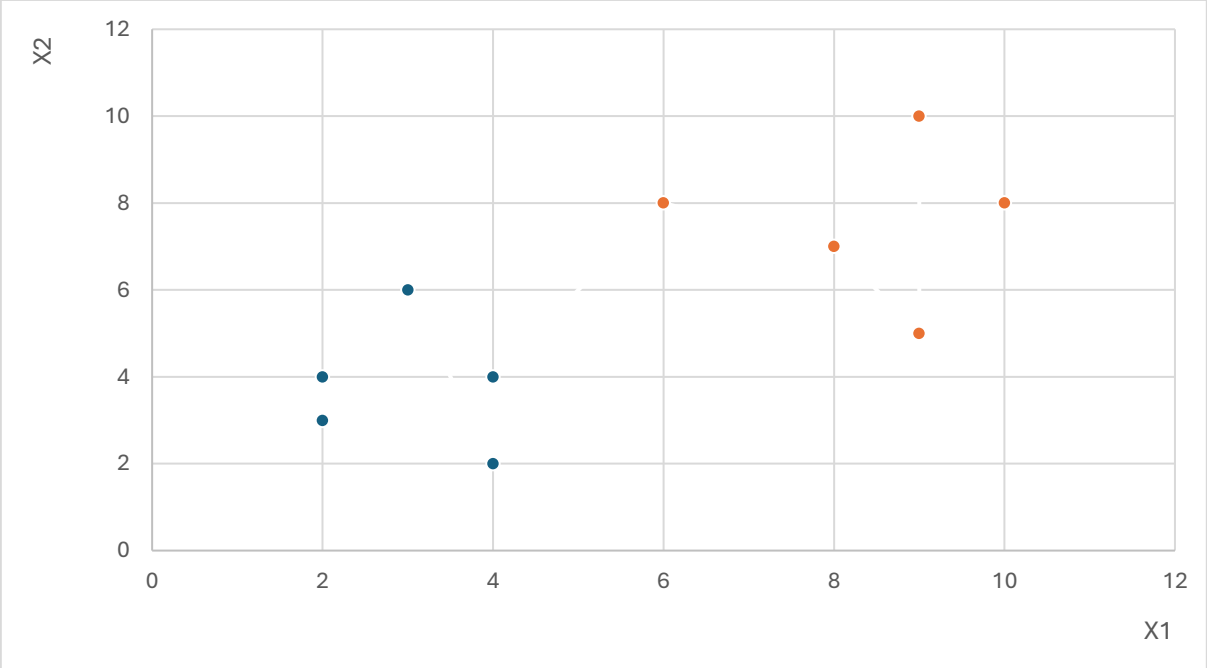
Ejemplo. Sean los conjuntos de vectores

Para clase w_1: $X_1 = (x_1, x_2) = \{ (4,2), (2,4), (2,3), (3,6), (4,4) \}$

Para clase w_2: $X_2 = (x_1, x_2) = \{ (9,10), (6,8), (9,5), (8,7), (10,8) \}$

Aplicar el método LDA para reducir a una dimensión, usando independencia de clases.

Nota: Antes de resolverlo, observe la gráfica:



Observe que para este ejemplo las dos clases son linealmente separables. El método LDA se puede aplicar para clases no linealmente separables o no separables. No es necesario que las clases tengan la misma cantidad de vectores.

Solución.

1) Calculamos las medias.

$$M_1 = \frac{1}{N_1}\sum_{x\,\varepsilon\,W_1} x = \frac{1}{5}\left[\binom{4}{2} + \binom{2}{4} + \binom{2}{3} + \binom{3}{6} + \binom{4}{4}\right] = \binom{3}{3.8}$$

$$M_2 = \frac{1}{N_2}\sum_{x\,\varepsilon\,W_2} x = \frac{1}{5}\left[\binom{9}{10} + \binom{6}{8} + \binom{9}{5} + \binom{8}{7} + \binom{10}{8}\right] = \binom{8.4}{7.6}$$

2) Se calcula las matrices de covariancias

$$S_1 = \frac{1}{N-1}\sum_{x\,\varepsilon\,W_1} [(x-\mu_1),(x-\mu_1)^T]$$

$$= \frac{1}{4}\left\{\left[\binom{4}{2} - \binom{3}{3.8}\right]\left[\binom{4}{2} - \binom{3}{3.8}\right]^T + \left[\binom{2}{4} - \binom{3}{3.8}\right]\left[\binom{2}{4} - \binom{3}{3.8}\right]^T + \left[\binom{2}{3} - \binom{3}{3.8}\right]\left[\binom{2}{3} - \binom{3}{3.8}\right]^T + \left[\binom{3}{6} - \binom{3}{3.8}\right]\left[\binom{3}{6} - \binom{3}{3.8}\right]^T + \left[\binom{4}{4} - \binom{3}{3.8}\right]\left[\binom{4}{4} - \binom{3}{3.8}\right]^T\right\} = \begin{bmatrix} 1 & -0.25 \\ -0.25 & 1 \end{bmatrix}$$

$$S_2 = \frac{1}{N-1}\sum_{x\,\varepsilon\,W_2} [(x-\mu_2),(x-\mu_2)^T]$$

$$= \frac{1}{4}\left\{\left[\binom{9}{10} - \binom{8.4}{7.6}\right]\left[\binom{9}{10} - \binom{8.4}{7.6}\right]^T + \left[\binom{6}{8} - \binom{8.4}{7.6}\right]\left[\binom{6}{8} - \binom{8.4}{7.6}\right]^T + \left[\binom{9}{5} - \binom{8.4}{7.6}\right]\left[\binom{9}{5} - \binom{8.4}{7.6}\right]^T + \left[\binom{8}{7} - \binom{8.4}{7.6}\right]\left[\binom{8}{7} - \binom{8.4}{7.6}\right]^T + \left[\binom{10}{8} - \binom{8.4}{7.6}\right]\left[\binom{10}{8} - \binom{8.4}{7.6}\right]^T\right\} = \begin{bmatrix} 2.3 & -0.05 \\ -0.05 & 3.3 \end{bmatrix}$$

3) Se calculan la matiz scatter (mezclada) entre clases (within-clases)

$$S_w = S_1 + S_2 = \begin{bmatrix} 1 & -0.25 \\ -0.25 & 1 \end{bmatrix} + \begin{bmatrix} 2.3 & -0.05 \\ -0.05 & 3.3 \end{bmatrix} = \begin{bmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{bmatrix}$$

Y la matriz mezclada de medias

$$S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

$$= \left[\binom{3}{3.8} - \binom{8.4}{7.6}\right]\left[\binom{3}{3.8} - \binom{8.4}{7.6}\right]^T = \binom{-5.4}{-3.8}\begin{pmatrix} -5.4 & -3.8 \end{pmatrix}$$

$$= \begin{pmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{pmatrix}$$

4. Se obtienen los valores propios (eigenvalores), de acuerdo a la siguiente expresión:

$$S_w^{-1} S_B w = \lambda w$$

Entonces:

$$|S_w^{-1} S_B - \lambda w| = 0$$

$$\left| \begin{pmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{pmatrix}^{-1} \begin{pmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right| = 0$$

$$\left| \begin{pmatrix} 0.3045 & 0.166 \\ 0.0166 & 0.1827 \end{pmatrix} \begin{pmatrix} 29.16 & 20.52 \\ 20.52 & 14.44 \end{pmatrix} - \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} \right| = 0$$

$$\left| \begin{pmatrix} 9.2213 - \lambda & 6.489 \\ 4.2339 & 2.9794 - \lambda \end{pmatrix} \right| = 0$$

$$(9.2213 - \lambda)(2.9794 - \lambda) - 6.489(4.2339) = 0$$

$$\lambda^2 - 12.2007\lambda = 0$$

$$\lambda_1 = 0, \qquad \lambda_2 = 12.2007$$

5. Se obtienen los vectores propios asociados a la expresión:

$$(S_w^{-1} S_B - \lambda I) \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 0$$

de donde:

$$w_1 = \begin{pmatrix} -0.5755 \\ 0.8178 \end{pmatrix} \quad y \quad w_2 = \begin{pmatrix} 0.9088 \\ 0.4173 \end{pmatrix}$$

Método alterno:

Después de (3)

4,5. Se usa w_2 porque es del valor propio mayor de la expresión:

$$w_2 = S_w^{-1}(\mu_1 - \mu_2)$$

$$w_2 = S_w^{-1}(\mu_1 - \mu_2) = \begin{pmatrix} 3.3 & -0.3 \\ -0.3 & 5.5 \end{pmatrix}^{-1} \left[\begin{pmatrix} 3 \\ 3.8 \end{pmatrix} - \begin{pmatrix} 8.4 \\ 7.6 \end{pmatrix} \right]$$

$$= \begin{pmatrix} 0.3045 & 0.0166 \\ 0.0166 & 0.1827 \end{pmatrix} \begin{pmatrix} -5.4 \\ -3.8 \end{pmatrix} = \begin{pmatrix} 0.9088 \\ 0.4173 \end{pmatrix}$$

6)Se obtienen los LDA como:

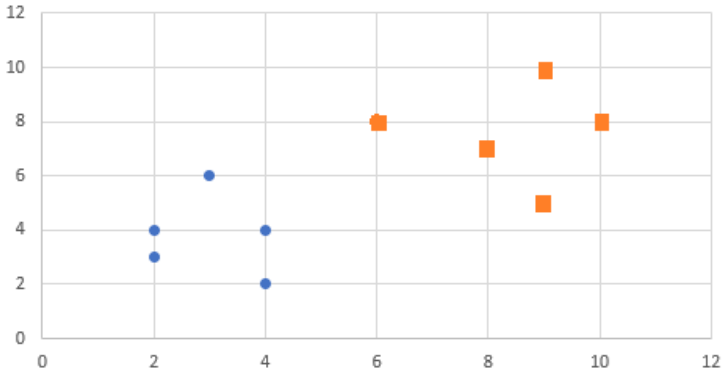
$$x_i w_2$$

Así, LDA termina como:

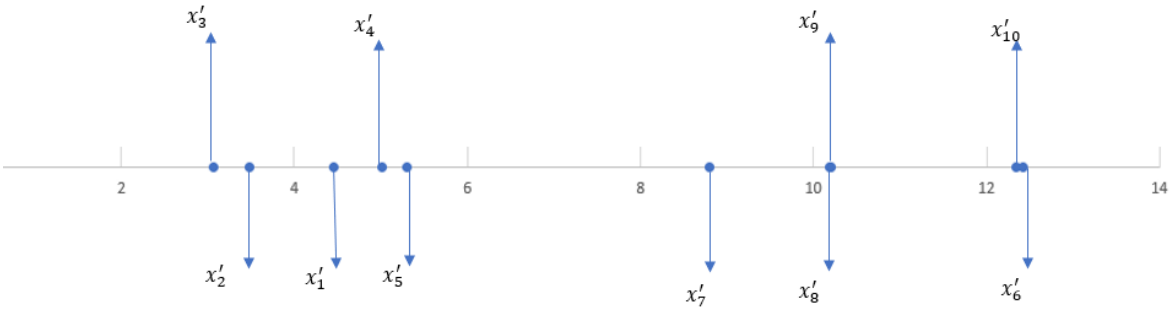
(4,2)	(2,4)	(2,3)	(3,6)	(4,4)	(9.10)
4.46	3.8	3.06	5.02	5.3	12.35

(6.8)	(9,5)	(8,7)	(10,8)
8.8	10.2	10.19	12.42

Para ilustrarlo, de la gráfica



Se mapean a los puntos en una dimensión.



Termina.

Existen dos tipos de técnicas LDA para manejar las clases: dependiente de la clase o independiente de la clase.

En la LDA dependiente de la clase un espacio por separado de menor dimensión se calcula para cada clase, proyectando sus datos en este. En la LDA independiente de la clase cada clase debe ser considerada como una clase separada contra las demás; así, hay una solo espacio de dimensión menoral que todas las clases proyectan sus datos.

A pesar de que la técnica LDA es la más utilizada para reducción de dimensiones, afronta varios problemas. Uno es que se le dificulta obtener el espacio dimensional menor cuando las dimensiones son mucho mayores que el número de vectores de datos. Lo que va a suceder es que la matriz dentro de una clase se vuelve singular, lo que se conoce como *small sample problem* (SSS). Existen varias propuestas para resolverlo. Una es remover el espacio nulo dentro de la matriz de una clase.

La segunda propuesta usa un espacio intermedio (por ejemplo PCA) para convertir la matriz de una clase en una matriz de

rango completo.

La tercera propuesta es muy conocida, usa el método de regularización para resolver un sistema singular.

Un segundo problema, llamado problema de linealidad, se tiene cuando las clases no son linealmente separables, entonces LDA no puede discriminar entre las clases. Una solución es usar kernels.