Análisis Discriminante Lineal y Cuadrático

Dr. Oldemar Rodríguez R.

UCR

1 de octubre de 2022

Notación

- Supondremos que deseamos clasificar una observación (individuo) en una de K clases con $K \geq 2$. Es decir, si Y es la variable a predecir entonces Y tiene K modalidades o categorías.
- Denotamos por π_k la probabilidad a priori de obtener al azar un individuo de la clase k.
- Sea

$$f_k(X) := Pr(X = x | Y = k)$$

La Función de Densidad de la variable aleatoria X para una observación (individuo) que está en la clase k.

Notación

- Supondremos que deseamos clasificar una observación (individuo) en una de K clases con $K \geq 2$. Es decir, si Y es la variable a predecir entonces Y tiene K modalidades o categorías.
- Denotamos por π_k la probabilidad a priori de obtener al azar un individuo de la clase k.
- Sea

$$f_k(X) := Pr(X = x|Y = k)$$

La Función de Densidad de la variable aleatoria X para una observación (individuo) que está en la clase k.

Notación

- Supondremos que deseamos clasificar una observación (individuo) en una de K clases con $K \geq 2$. Es decir, si Y es la variable a predecir entonces Y tiene K modalidades o categorías.
- Denotamos por π_k la probabilidad a priori de obtener al azar un individuo de la clase k.
- Sea

$$f_k(X) := Pr(X = x|Y = k)$$

La Función de Densidad de la variable aleatoria X para una observación (individuo) que está en la clase k.

Teorema Bayes

Teorema (Naïve-Bayes)

Con la notación anterior se tiene que:

$$Pr(Y = k|X = x) = \frac{\pi_k f_k(x)}{\sum_{s=1}^{K} \pi_s f_s(x)} \tag{1}$$

Notación:

$$p_k(X) = Pr(Y = k|X = x)$$

Teorema Bayes

Teorema (Naïve-Bayes)

Con la notación anterior se tiene que:

$$Pr(Y = k|X = x) = \frac{\pi_k f_k(x)}{\sum_{s=1}^{K} \pi_s f_s(x)} \tag{1}$$

Notación:

$$p_k(X) = Pr(Y = k|X = x)$$

Análisis Discriminate Lineal para $p = 1$

Se asume que:

1. $p = 1$, es decir, que tenemos solamente una variable predictora.
2. $f_k(x)$ sigue la Normal, o sea, que su función de densidad tiene la forma:

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{\left(-\frac{1}{2\sigma_k^2}(x-\mu_k)^2\right)}, \tag{2}$$

   donde μ_k y σ_k son la media y la varianza de la k-ésima clase respectivamente.
3. $\sigma_1 = \sigma_2 = \cdots = \sigma_K = \sigma$, es decir, que la varianza es la misma en las K clases. Así la función de densidad es:

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}. \tag{3}$$

Análisis Discriminate Lineal para $p = 1$

Se asume que:

1. $p = 1$, es decir, que tenemos solamente una variable predictora.
2. $f_k(x)$ sigue la Normal, o sea, que su función de densidad tiene la forma:

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{\left(-\frac{1}{2\sigma_k^2}(x-\mu_k)^2\right)}, \tag{2}$$

   donde μ_k y σ_k son la media y la varianza de la k-ésima clase respectivamente.
3. $\sigma_1 = \sigma_2 = \cdots = \sigma_K = \sigma$, es decir, que la varianza es la misma en las K clases. Así la función de densidad es:

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}. \tag{3}$$

Análisis Discriminate Lineal para $p = 1$

Se asume que:

1. $p = 1$, es decir, que tenemos solamente una variable predictora.
2. $f_k(x)$ sigue la Normal, o sea, que su función de densidad tiene la forma:

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma_k} e^{\left(-\frac{1}{2\sigma_k^2}(x-\mu_k)^2\right)}, \tag{2}$$

   donde μ_k y σ_k son la media y la varianza de la k-ésima clase respectivamente.
3. $\sigma_1 = \sigma_2 = \cdots = \sigma_K = \sigma$, es decir, que la varianza es la misma en las K clases. Así la función de densidad es:

$$f_k(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}. \tag{3}$$

Análisis Discriminate Lineal para $p = 1$

1. La idea del Análisis Discriminante es simple, consiste en usar el Teorema de Bayes para clasificar una observación (individuo) en una de las K clases, con la diferencia de que la probabilidad $p_k(X) = Pr(Y = k|X = x)$ será calculada usando la función de densidad de la ecuación (3).
2. Así sustituyendo (3) en (1) se tiene que:

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi\sigma}} e^{\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}}{\sum_{s=1}^{K} \pi_s \frac{1}{\sqrt{2\pi\sigma}} e^{\left(-\frac{1}{2\sigma^2}(x-\mu_s)^2\right)}}. \tag{4}$$

3. El Análisis Discriminante consiste en calcular todas las K probabilidades de la ecuación (4) y asignar la observación (individuo) $X = x$ a la clase con mayor probabilidad (igual que como se procede en el Método de Bayes).

Análisis Discriminate Lineal para $p = 1$

1. La idea del Análisis Discriminante es simple, consiste en usar el Teorema de Bayes para clasificar una observación (individuo) en una de las K clases, con la diferencia de que la probabilidad $p_k(X) = Pr(Y = k|X = x)$ será calculada usando la función de densidad de la ecuación (3).
2. Así sustituyendo (3) en (1) se tiene que:

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}}{\sum_{s=1}^{K} \pi_s \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2\sigma^2}(x-\mu_s)^2\right)}}. \tag{4}$$

3. El Análisis Discriminante consiste en calcular todas las K probabilidades de la ecuación (4) y asignar la observación (individuo) $X = x$ a la clase con mayor probabilidad (igual que como se procede en el Método de Bayes).

Análisis Discriminate Lineal para $p = 1$

1. La idea del Análisis Discriminante es simple, consiste en usar el Teorema de Bayes para clasificar una observación (individuo) en una de las K clases, con la diferencia de que la probabilidad $p_k(X) = Pr(Y = k|X = x)$ será calculada usando la función de densidad de la ecuación (3).
2. Así sustituyendo (3) en (1) se tiene que:

$$p_k(x) = \frac{\pi_k \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2\sigma^2}(x-\mu_k)^2\right)}}{\sum_{s=1}^{K} \pi_s \frac{1}{\sqrt{2\pi}\sigma} e^{\left(-\frac{1}{2\sigma^2}(x-\mu_s)^2\right)}}. \tag{4}$$

3. El Análisis Discriminante consiste en calcular todas las K probabilidades de la ecuación (4) y asignar la observación (individuo) $X = x$ a la clase con mayor probabilidad (igual que como se procede en el Método de Bayes).

Teorema Análisis Discriminante

Teorema (Regla de Asignación)

El Análisis Discriminante asigna $X = x$ a la clase con mayor $\delta_k(x)$ donde:

$$\delta_k(x) = x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + log(\pi_k). \tag{5}$$

Prueba:

Basta aplicar logaritmo en la ecuación (4).

Teorema Análisis Discriminante

Teorema (Regla de Asignación)

El Análisis Discriminante asigna $X = x$ a la clase con mayor $\delta_k(x)$ donde:

$$\delta_k(x) = x \cdot \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + log(\pi_k). \tag{5}$$

Prueba:

Basta aplicar logaritmo en la ecuación (4).

Análisis Discriminate Lineal para $p > 1$

1. La idea es extender el Análisis Discriminate Lineal (LDA) al caso de múltiples predictores, para esto se asume que tienen $X = (X_1, X_2, \ldots, X_p)$ p variables que siguen la Distribución Normal y que tienen asociados un vector de medias y una matriz de covarianzas.
2. Es decir, se tiene una variable aleatoria p dimensional X que tiene una distribución normal (Gaussiana) multivariada, esto se denota como $X \sim N(\mu, \Sigma)$, donde $E(X) = \mu$ es un vector con p componentes con la medias de X y $\Sigma = \mathrm{Cov}(X)$ es la matriz de covarianzas de X de tamaño $p \times p$.
3. Donde la Normal Multivariada se define como:

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{\left(-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right)}. \tag{6}$$

Análisis Discriminate Lineal para $p > 1$

1. La idea es extender el Análisis Discriminate Lineal (LDA) al caso de múltiples predictores, para esto se asume que tienen $X = (X_1, X_2, \ldots, X_p)$ p variables que siguen la Distribución Normal y que tienen asociados un vector de medias y una matriz de covarianzas.
2. Es decir, se tiene una variable aleatoria p dimensional X que tiene una distribución normal (Gaussiana) multivariada, esto se denota como $X \sim N(\mu, \Sigma)$, donde $E(X) = \mu$ es un vector con p componentes con la medias de X y $\Sigma = \mathrm{Cov}(X)$ es la matriz de covarianzas de X de tamaño $p \times p$.
3. Donde la Normal Multivariada se define como:

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{\left(-\frac{1}{2}(x-\mu)^t\Sigma^{-1}(x-\mu)\right)}. \tag{6}$$

Análisis Discriminate Lineal para $p > 1$

1. La idea es extender el Análisis Discriminate Lineal (LDA) al caso de múltiples predictores, para esto se asume que tienen $X = (X_1, X_2, \ldots, X_p)$ p variables que siguen la Distribución Normal y que tienen asociados un vector de medias y una matriz de covarianzas.
2. Es decir, se tiene una variable aleatoria p dimensional X que tiene una distribución normal (Gaussiana) multivariada, esto se denota como $X \sim N(\mu, \Sigma)$, donde $E(X) = \mu$ es un vector con p componentes con la medias de X y $\Sigma = \text{Cov}(X)$ es la matriz de covarianzas de X de tamaño $p \times p$.
3. Donde la Normal Multivariada se define como:

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} e^{\left(-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right)}. \tag{6}$$

Análisis Discriminante Lineal para $p > 1$

En el caso en que se tienen $p > 1$ predictores, el clasificador LDA asume que la observaciones (individuos) en la clase k siguen una distribución Gaussiana (Normal) Multivariante $N(\mu_k, \Sigma)$, donde μ_k es un vector específico con la medias de de clase k, y Σ es un matriz de covarianzas, que en LDA **se asumen iguales** para todas las K clases.

Teorema (Regla de Asignación)

El Análisis Discriminante asigna $X = x$ a la clase con mayor $\delta_k(x)$ donde:

$$\delta_k(x) = x^t\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^t\Sigma^{-1}\mu_k + log(\pi_k). \qquad (7)$$

Prueba:

Se debe sustituir (6) en (1) y aplicar logaritmo en la ecuación resultante, completar los detalles en la Tarea.

Análisis Discriminante Lineal para $p > 1$

En el caso en que se tienen $p > 1$ predictores, el clasificador LDA asume que la observaciones (individuos) en la clase k siguen una distribución Gaussiana (Normal) Multivariante $N(\mu_k, \Sigma)$, donde μ_k es un vector específico con la medias de de clase k, y Σ es un matriz de covarianzas, que en LDA **se asumen iguales** para todas las K clases.

Teorema (Regla de Asignación)

El Análisis Discriminante asigna $X = x$ a la clase con mayor $\delta_k(x)$ donde:

$$\delta_k(x) = x^t\Sigma^{-1}\mu_k - \frac{1}{2}\mu_k^t\Sigma^{-1}\mu_k + log(\pi_k). \tag{7}$$

Prueba:

Se debe sustituir (6) en (1) y aplicar logaritmo en la ecuación resultante, completar los detalles en la Tarea.

Análisis Discriminante Lineal para $p > 1$

En el caso en que se tienen $p > 1$ predictores, el clasificador LDA asume que la observaciones (individuos) en la clase k siguen una distribución Gaussiana (Normal) Multivariante $N(\mu_k, \Sigma)$, donde μ_k es un vector específico con la medias de de clase k, y Σ es un matriz de covarianzas, que en LDA **se asumen iguales** para todas las K clases.

Teorema (Regla de Asignación)

El Análisis Discriminante asigna $X = x$ a la clase con mayor $\delta_k(x)$ donde:

$$\delta_k(x) = x^t \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^t \Sigma^{-1} \mu_k + log(\pi_k). \tag{7}$$

Prueba:

Se debe sustituir (6) en (1) y aplicar logaritmo en la ecuación resultante, completar los detalles en la Tarea.

Un ejemplo con 3 clases

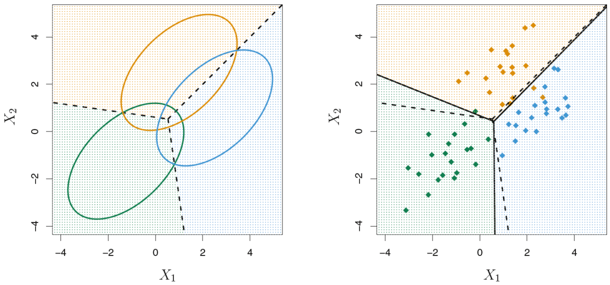

FIGURE 4.6. *An example with three classes. The observations from each class are drawn from a multivariate Gaussian distribution with $p = 2$, with a class-specific mean vector and a common covariance matrix.* Left: *Ellipses that contain* 95 % *of the probability for each of the three classes are shown. The dashed lines are the Bayes decision boundaries.* Right: 20 *observations were generated from each class, and the corresponding LDA decision boundaries are indicated using solid black lines. The Bayes decision boundaries are once again shown as dashed lines.*

Figura: 2. La línea vertical discontinua representa la frontera de decisión de Bayes y la línea sólida al Análisis Discriminante).

Nótese que en (7) $\delta_k(x)$ es una función lineal de x; esto es, la regla de decisión en LDA depende de solo de x a través de una combinación lineal de sus elementos, esta es la razón por la cual se conoce como Análisis Discriminante **Lineal**.

Análisis Discriminante Cuadrático - QDA

- Como ya hemos comentado, en LDA se asume que las observaciones (individuos) dentro de cada clase siguen una distribución Gaussiana (Normal) Multivariante con un vector medias específico para cada clase pero con una matriz de covarianzas que es común para todas las K clases.
- En el Análisis Discriminante Cuadrático (QDA) se propone una alternativa cuadrática al enfoque del Análisis Discriminante. Al igual LDA, en QDA se parte del supuesto de que las observaciones (individuos) de cada clase siguen una distribución de Gaussiana (Normal).
- Sin embargo, a diferencia de LDA, QDA asume que cada clase tiene su propia matriz de covarianzas Σ_k. Es decir, $X \sim N(\mu_k, \Sigma_k)$ donde Σ_k es la matriz de covarianzas de la k-ésima clase.

Análisis Discriminante Cuadrático - QDA

- Como ya hemos comentado, en LDA se asume que las observaciones (individuos) dentro de cada clase siguen una distribución Gaussiana (Normal) Multivariante con un vector medias específico para cada clase pero con una matriz de covarianzas que es común para todas las K clases.
- En el Análisis Discriminante Cuadrático (QDA) se propone una alternativa cuadrática al enfoque del Análisis Discriminante. Al igual LDA, en QDA se parte del supuesto de que las observaciones (individuos) de cada clase siguen una distribución de Gaussiana (Normal).
- Sin embargo, a diferencia de LDA, QDA asume que cada clase tiene su propia matriz de covarianzas Σ_k. Es decir, $X \sim N(\mu_k, \Sigma_k)$ donde Σ_k es la matriz de covarianzas de la k-ésima clase.

Análisis Discriminante Cuadrático - QDA

- Como ya hemos comentado, en LDA se asume que las observaciones (individuos) dentro de cada clase siguen una distribución Gaussiana (Normal) Multivariante con un vector medias específico para cada clase pero con una matriz de covarianzas que es común para todas las K clases.
- En el Análisis Discriminante Cuadrático (QDA) se propone una alternativa cuadrática al enfoque del Análisis Discriminante. Al igual LDA, en QDA se parte del supuesto de que las observaciones (individuos) de cada clase siguen una distribución de Gaussiana (Normal).
- Sin embargo, a diferencia de LDA, QDA asume que cada clase tiene su propia matriz de covarianzas Σ_k. Es decir, $X \sim N(\mu_k, \Sigma_k)$ donde Σ_k es la matriz de covarianzas de la k-ésima clase.

Análisis Discriminante Cuadrático para $p > 1$

Teorema (Regla de Asignación)

El Análisis Discriminante Cuadrático asigna $X = x$ a la clase con mayor $\delta_k(x)$ donde:

$$\begin{aligned}\delta_k(x) &= -\frac{1}{2}(x-\mu_k)^t\Sigma_k^{-1}(x-\mu_k) + log(\pi_k) \\ &= -\frac{1}{2}x^t\Sigma_k^{-1}x + x^t\Sigma_k^{-1}\mu_k - \frac{1}{2}\mu_k^t\Sigma_k^{-1}\mu_k + log(\pi_k)\end{aligned} \quad (8)$$

Prueba:
En este caso la función de densidad es:

$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} e^{\left(-\frac{1}{2}(x-\mu_k)^t\Sigma_k^{-1}(x-\mu_k)\right)}. \quad (9)$$

Se debe sustituir (9) en (1) y aplicar logaritmo en la ecuación resultante.

Análisis Discriminante Cuadrático para $p > 1$

Teorema (Regla de Asignación)

El Análisis Discriminante Cuadrático asigna $X = x$ a la clase con mayor $\delta_k(x)$ donde:

$$\begin{aligned} \delta_k(x) &= -\frac{1}{2}(x-\mu_k)^t\Sigma_k^{-1}(x-\mu_k) + log(\pi_k) \\ &= -\frac{1}{2}x^t\Sigma_k^{-1}x + x^t\Sigma_k^{-1}\mu_k - \frac{1}{2}\mu_k^t\Sigma_k^{-1}\mu_k + log(\pi_k) \end{aligned} \tag{8}$$

Prueba:

En este caso la función de densidad es:

$$f_k(x) = \frac{1}{(2\pi)^{p/2}|\Sigma_k|^{1/2}} e^{\left(-\frac{1}{2}(x-\mu_k)^t\Sigma_k^{-1}(x-\mu_k)\right)}. \tag{9}$$

Se debe sustituir (9) en (1) y aplicar logaritmo en la ecuación resultante.

Gráficamente

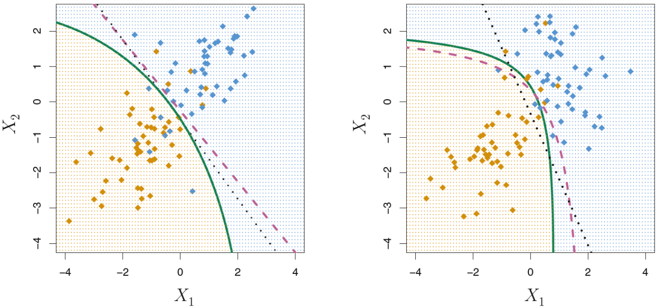

FIGURE 4.9. Left: *The Bayes (purple dashed), LDA (black dotted), and QDA (green solid) decision boundaries for a two-class problem with* $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2$. *The shading indicates the QDA decision rule. Since the Bayes decision boundary is linear, it is more accurately approximated by LDA than by QDA.* Right: *Details are as given in the left-hand panel, except that* $\mathbf{\Sigma}_1 \neq \mathbf{\Sigma}_2$. *Since the Bayes decision boundary is non-linear, it is more accurately approximated by QDA than by LDA.*

Figura: 3. Comparación entre Bayes, LDA y QDA.