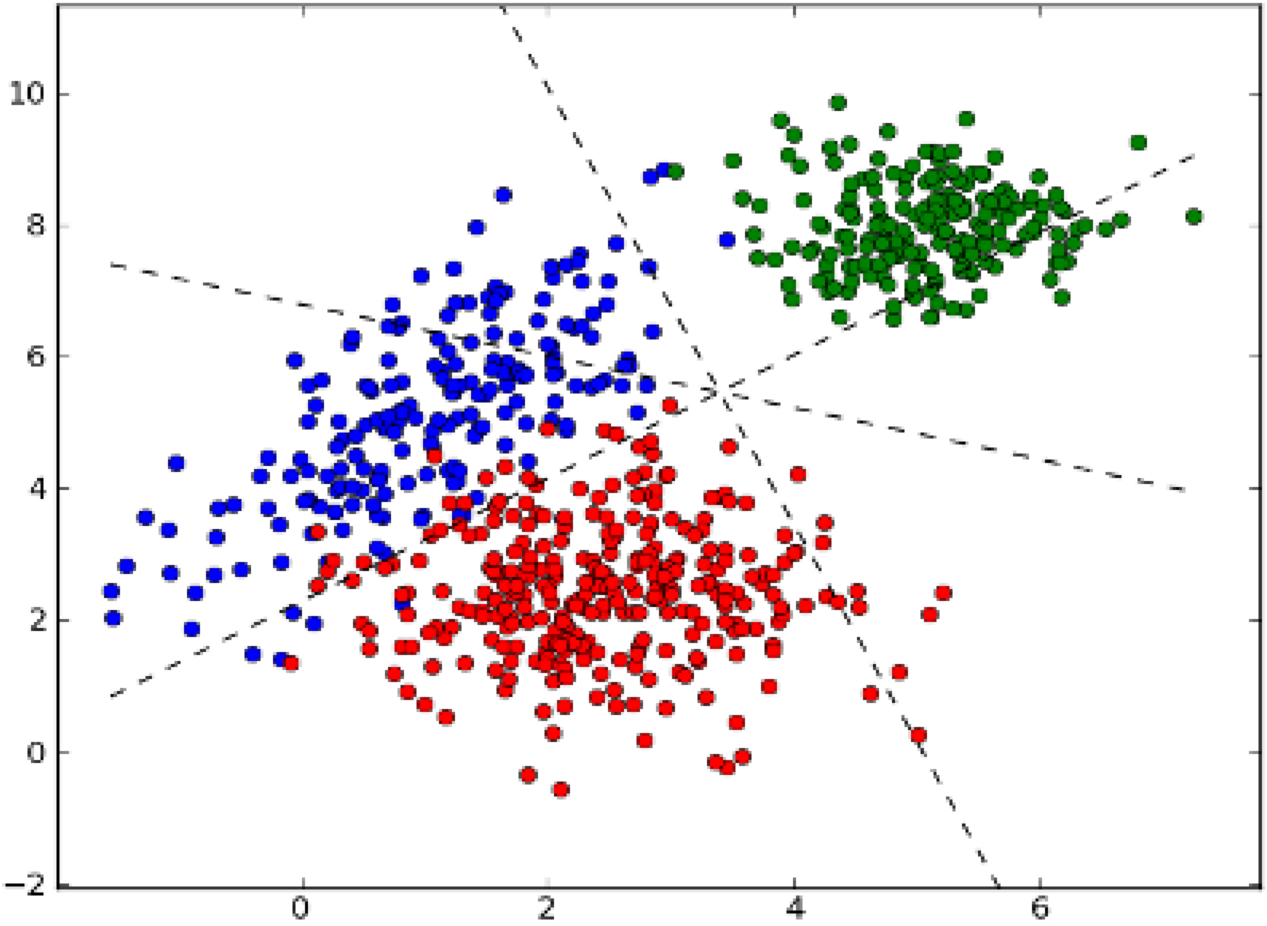
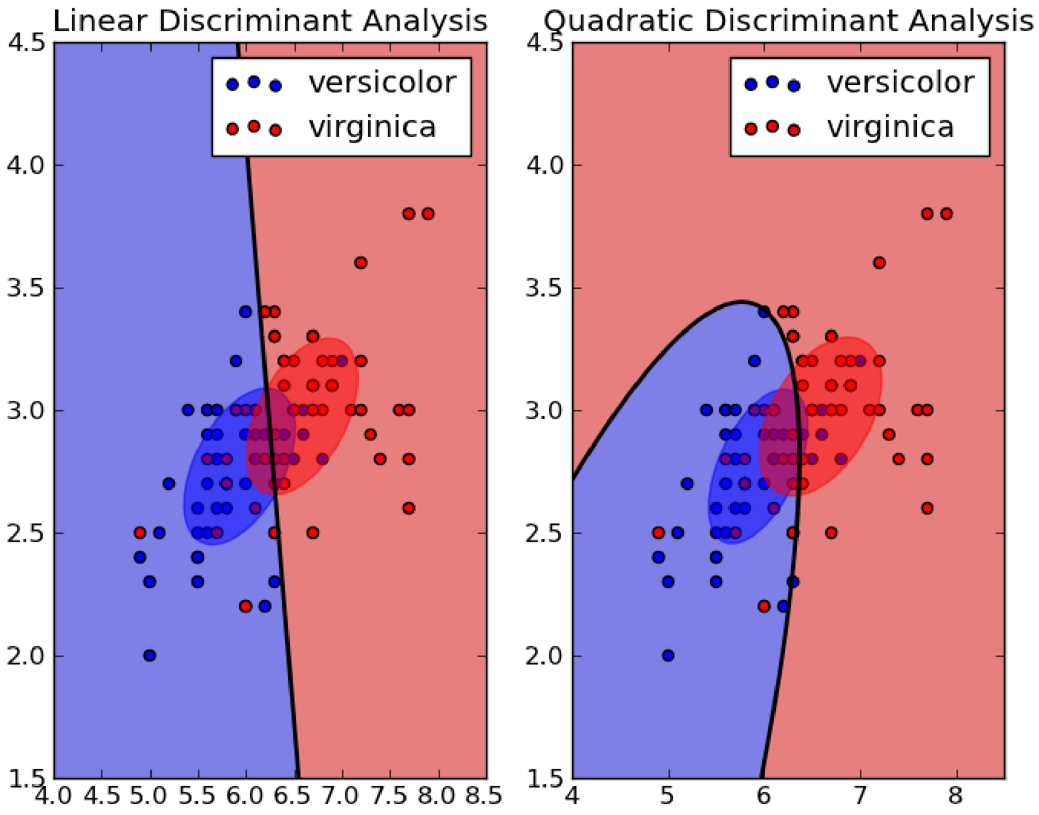
Análisis discriminante





Oscar Centeno Mora

Índice

Índice

Índice

Introducción

- El análisis discriminante es una técnica multivariante cuya finalidad es describir las diferencias significativas entre k grupos de objetos (k > 1) sobre los que se observan (variable de clasificación).

- En caso de que estas diferencias existan, intentará explicar en qué sentido se dan y proporcionar procedimientos de asignación sistemática de nuevas observaciones con grupo desconocido a uno de los grupos analizados, utilizando para ello sus valores en las k variables clasificadoras.

- Podemos ver este procedimiento como un modelo de predicción de una variable respuesta categórica (variable grupo) a partir de k variables explicativas generalmente continuas (variables clasificatorias).

Introducción

- Dicho de otra forma más simple, supongamos que un conjunto de observaciones están ya clasificado en una serie de grupos, es decir, se sabe previamente a qué grupos pertenecen. El Análisis Discriminante trabajo como si se tratara de un análisis de regresión, donde la variable dependiente es categórica y tiene como categorías la etiqueta de cada uno de los grupos, y las variables independientes son continuas y determinan a qué grupos pertenecen los objetos.

- Se pretende encontrar relaciones lineales o cuadráticas entre las variables continuas (independientes) que mejor discriminen en los grupos dados a los objetos (dependiente).

- El objetivo es asignar los casos a cierto grupo, a partir de una regla que trata de discriminar o separar a los sujetos en los k grupos. Un segundo objetivo es construir una regla de decisión que asigne uno o varios unidades de estudio nuevos, que no sabemos clasificar previamente, a uno de los grupos prefijados con un cierto grado de riesgo. Esto es, a partir de las variables independientes, y sin conocer el grupo de pertenencia, se trata de clasificar a una observación en uno de los k grupos.

Introducción

Algunos ejemplos en dónde se podría aplicar el análisis discriminante:

Medicina

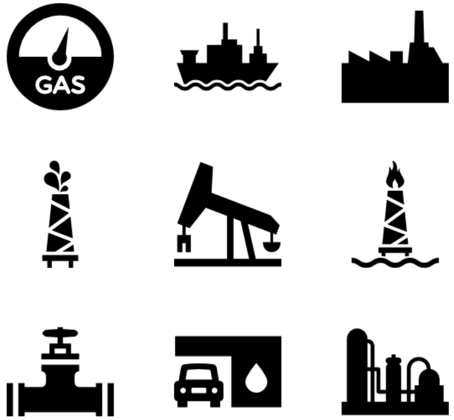
- A partir de los síntomas presentados por el paciente, ¿podemos hacer un diagnóstico de su enfermedad?

Finanzas

- A partir de los balances de una empresa, ¿es posible estimar o determinar el estado financiero de la empresa?
- En la solicitud de un préstamo, ¿se puede predecirse, en función de las características del cliente, el riesgo de morosidad (puntuación de crédito)?

Introducción

Petroleras



Dado el análisis de núcleos de una perforación, es posible suponer la existencia de una mancha de aceite?

Marketing

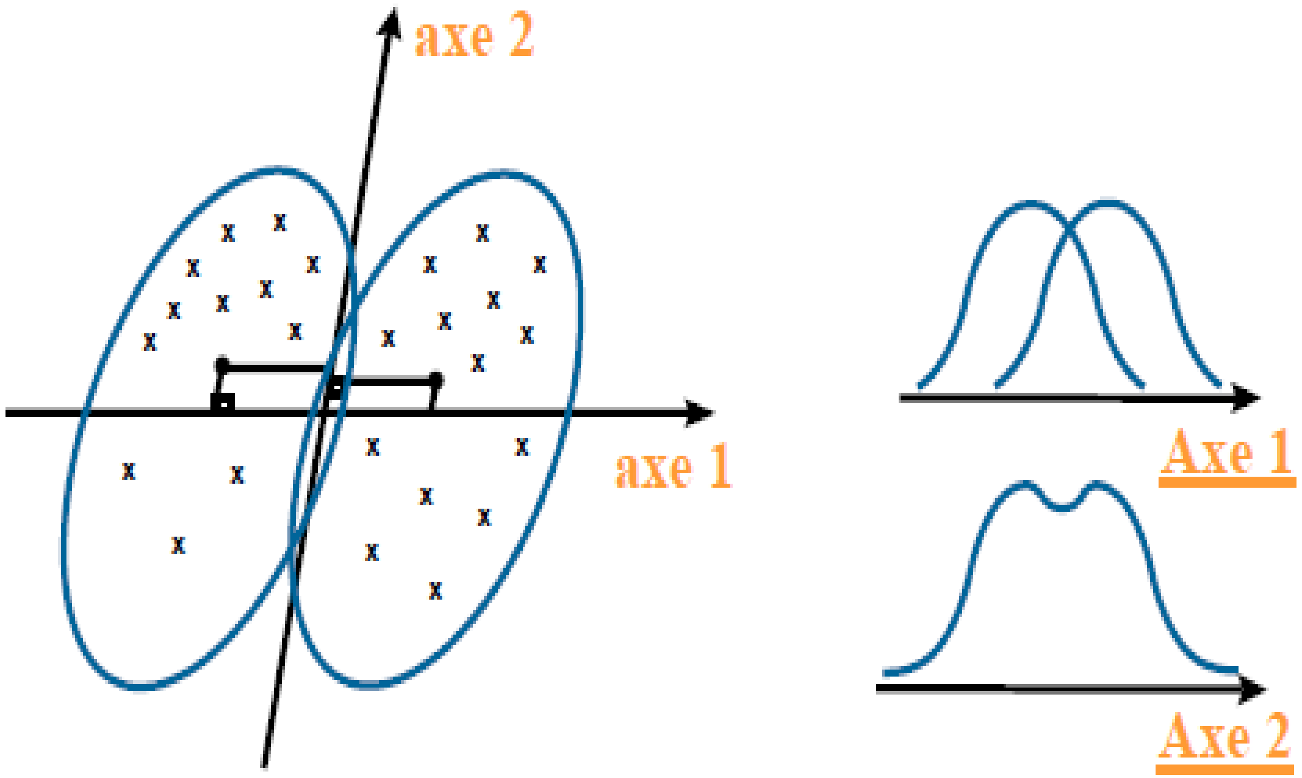
Es posible, a partir de las características de un cliente, predecir su respuesta a una oferta de productos por correo?

Educación

A partir de un test de aptitud, ¿se puede recomendar a cierta carrera?

Introducción

Por lo tanto, el Análisis Discriminante consiste en buscar nuevas variables (variables discriminantes), que separan lo mejor posible los grupos de proyección en las k variables, de las n observaciones, para poder predecir nuevas observaciones i.



El eje 1 posee un buen poder de discriminación

El eje 2 no permite separar según su proyección a los dos grupos

Índice

Índice

1

Introducción

2

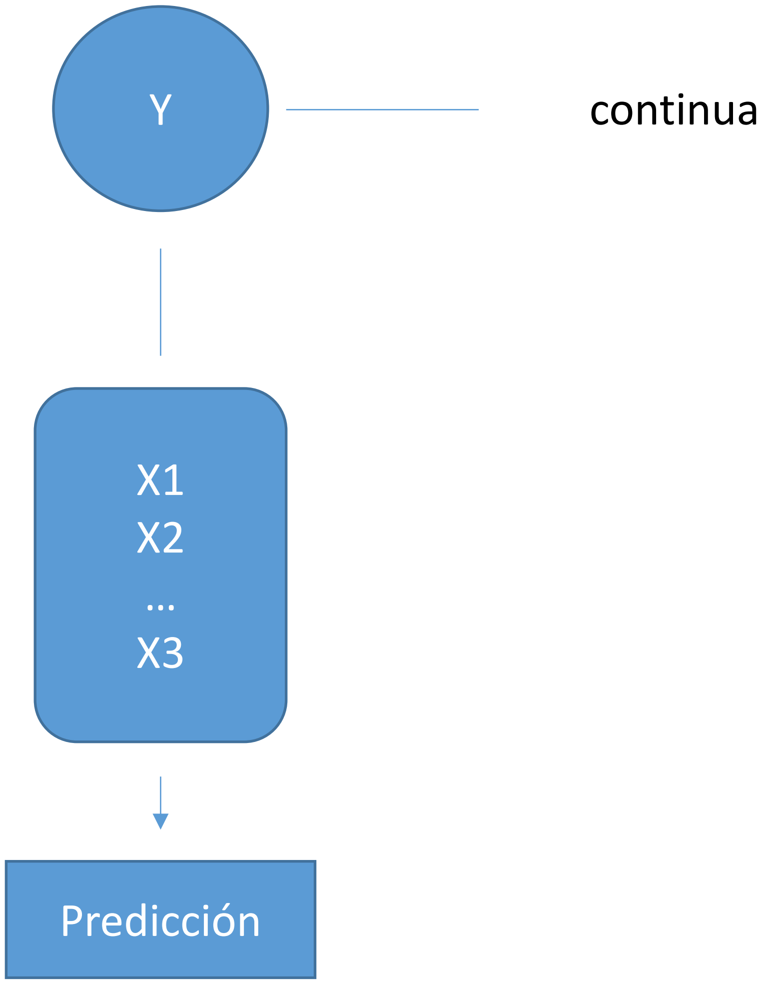
La clasificación

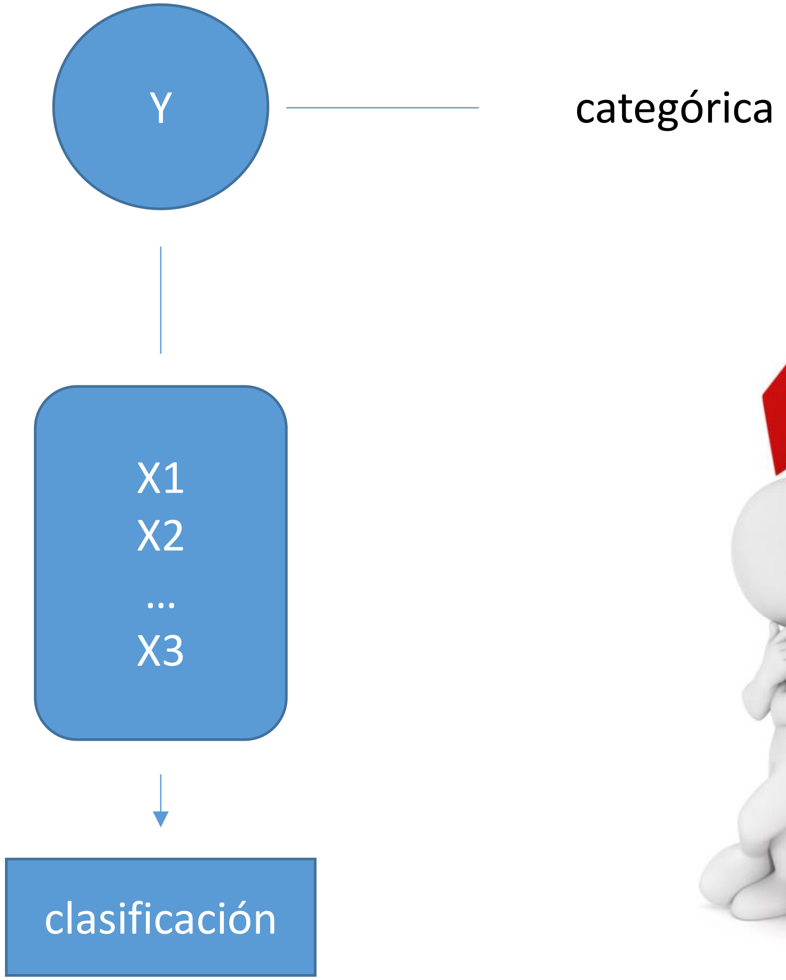
- Variables
- Objetivos
- Técnicas de clasificación

La clasificación

La clasificación: variables

- Hasta la fecha, en los análisis predictivos conocemos :



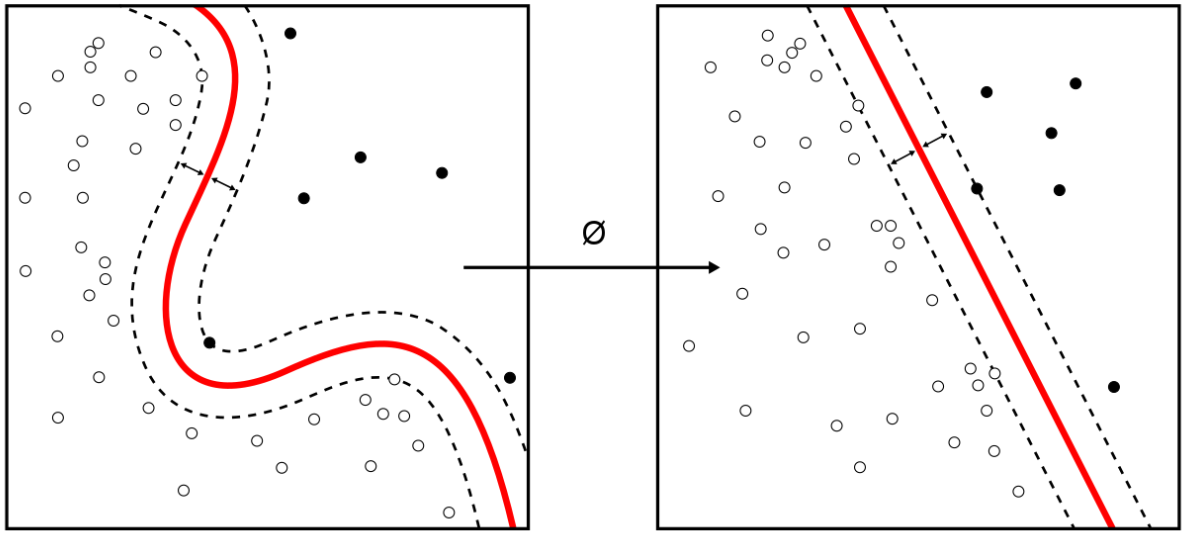


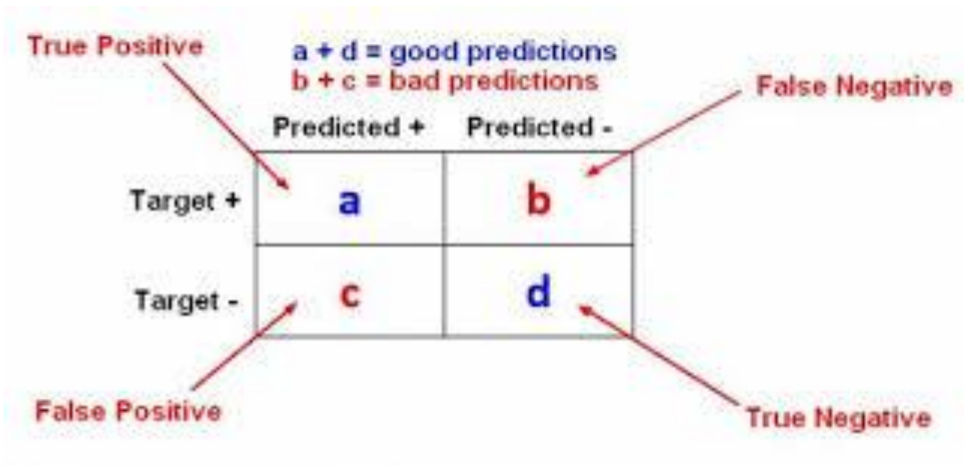
La clasificación: objetivo

En la clasificación se llevan a cado dos procesos fundamentales: construir la regla o ecuación de clasificación, y verificar la calidad del proceso, para luego introducir nuevos casos..

1

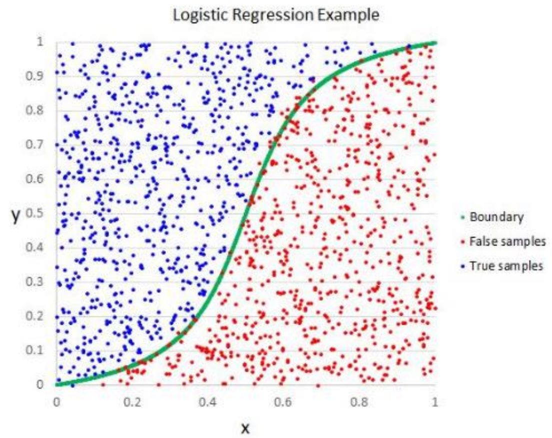
2





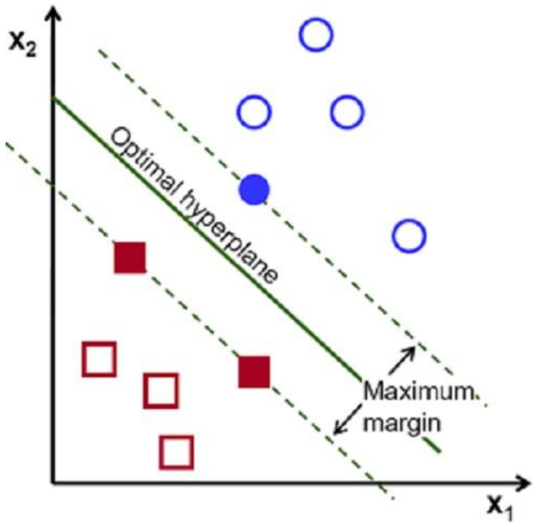
La clasificación: técnicas de clasificación

- ¿Cuál son las técnicas de predicción que hasta ahora conocemos?



- El presente capítulo introduce y explica el método de análisis discriminante como uno de los diversos métodos de clasificación presentes en la ciencia de datos.

- Algunas otras técnicas de clasificación:



1. Regresión dicotómica (logística).
2. Árboles de decisión.
3. Redes neuronales.
4. Support Vectors Machines (SMV).
5. Etc.

Índice

1. Introducción
2. La clasificación
3. Análisis discriminante lineal

Índice

Análisis Discriminante lineal: fundamento

El Análisis Discriminante Lineal o *Linear Discrimiant Analysis (LDA)* es un método de clasificación supervisado de variables cualitativas 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, el *LDA* estima la probabilidad de que una observación, dado un determinado valor de los predictores, pertenezca a cada una de las clases de la variable cualitativa, $P(Y = k|X = x)$. Finalmente se asigna la observación a la clase *k* para la que la probabilidad predicha es mayor.

Es una alternativa a la regresión logística cuando la variable cualitativa tiene más de dos niveles. Si bien existen extensiones de la regresión logística para múltiples clases, el *LDA* presenta una serie de ventajas:

1. Si las clases están bien separadas, los parámetros estimados en el modelo de regresión logística son inestables. El método de *LDA* no sufre este problema.
2. Si el número de observaciones es bajo y la distribución de los predictores es aproximadamente normal en cada una de las clases, *LDA* es más estable que la regresión logística.

Análisis Discriminante lineal: fundamento

Cuando se trata de un problema de clasificación con solo dos niveles, ambos métodos suelen llegar a resultados similares.

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

1. Disponer de un conjunto de datos de entrenamiento (*training data*) en el que se conoce a que grupo pertenece cada observación.
2. Calcular las probabilidades previas (*prior probabilities*): la proporción esperada de observaciones que pertenecen a cada grupo.
3. Determinar si la varianza o matriz de covarianzas es homogénea en todos los grupos. De esto dependerá que se emplee *LDA* o *QDA*.
4. Estimar los parámetros necesarios para las funciones de probabilidad condicional, verificando que se cumplen las condiciones para hacerlo.
5. Calcular el resultado de la función discriminante. El resultado de esta determina a qué grupo se asigna cada observación.
6. **Utilizar validación cruzada (*cross-validation*) para estimar las probabilidades de clasificaciones erróneas.

Análisis Discriminante lineal: teorema de Bayes

Considérense dos eventos A y B, el teorema de Bayes establece que la probabilidad de que B ocurra habiendo ocurrido A (denotado como $B|A$) es igual a la probabilidad de que A y B ocurran al mismo tiempo (AB), dividida entre la probabilidad de que ocurra A.

$$P(B|A) = \frac{P(AB)}{P(A)}$$

Supóngase que se desea clasificar una nueva observación en una de las K clases de una variable cualitativa Y, siendo $K \geq 2$, a partir de un solo predictor X. Se dispone de las siguientes definiciones:

- Se define como *overall*, *prior probability* o probabilidad previa (π_k) la probabilidad de que una observación aleatoria pertenezca a la clase K.
- Se define $f_k(X) \equiv P(X = x|Y = k)$ como la función de densidad de probabilidad condicional de *X* para una observación que pertenece a la clase *k*. Cuanto mayor sea $f_k(X)$ mayor la probabilidad de que una observación de la clase k adquiera un valor de $X \approx x$.
- Se define como *posterior probability* o probabilidad posterior $P(Y = k|X = x)$ la probabilidad de que una observación pertenezca a la clase k siendo x el valor del predictor.

Análisis Discriminante lineal: teorema de Bayes

- Aplicando del teorema de Bayes se pueden conocer la *probabilidad posterior* para cada clase:

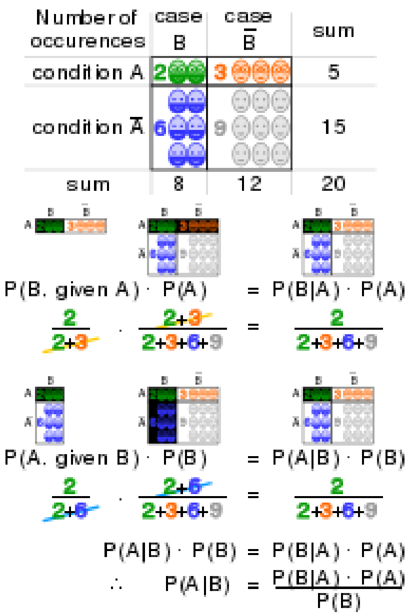
$$P(pertenece\ clase\ k|valor\ x\ obserdado) = \frac{P(pertene\ clase\ k\ y\ observar\ x)}{P(observar\ x)}$$

Si se introducen los términos, definidos anteriormente, dentro la ecuación se obtiene:

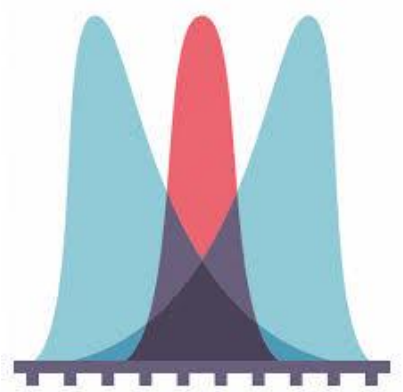
$$P(Y=k|X=x) = \frac{\pi_k P(X=x|Y=k)}{\sum_{j=1}^{K} \pi_j P(X=x|Y=j)} = \frac{\pi_k f_k(x)}{\sum_{j=1}^{K} \pi_j f_k(x)}$$

Análisis Discriminante lineal: teorema de Bayes

La clasificación con menor error (clasificación de Bayes) se consigue asignando la observación a aquel grupo que maximice la probabilidad posterior. Dado que el denominador $\sum_{j=1}^{K} \pi_j f_k(x)$ es igual para todas las clases, la norma de clasificación es equivalente a decir que se asignará cada observación a aquel grupo para el que $\pi_k f_k(x)$ sea mayor.



Para que la clasificación basada en Bayes sea posible, se necesita conocer la probabilidad poblacional de que una observación cualquiera pertenezca a cada clase (π_k) y la probabilidad poblacional de que una observación que pertenece a la clase k adquiera el valor x en el predictor ($f_k(X) \equiv P(X = x|Y = k)$).



En la práctica, raramente se dispone de esta información, por lo que los parámetros tienen que ser estimados a partir de la muestra. Como consecuencia, el clasificador *LDA* obtenido se aproxima al clasificador de Bayes pero no es igual.

Análisis Discriminante lineal: estimación de π_k y $f_k(X)$

La capacidad del *LDA* para clasificar correctamente las observaciones depende de cómo de buenas sean las estimaciones de π_k y de $f_k(X)$. Cuanto más cercanas al valor real, más se aproximará el clasificador *LDA* al clasificador de Bayes. En el caso de la probabilidad a priori π_k la estimación suele ser sencilla, la probabilidad de que una observación cualquiera pertenezca a la clase *k* es igual al número de observaciones de esa clase entre el número total de observaciones $\hat{\pi}_k = \frac{n_k}{N}$.

La estimación de $f_k(X)$ no es tan directa y para conseguirla se requiere de ciertas suposiciones. Si se considera que $f_k(X)$ se distribuye de forma normal en las *k clases,* entonces se puede estimar su valor a partir de la ecuación:

$$f_k(X) = P(Y = k|X = x) = \frac{1}{\sqrt{2\pi\sigma_k}} \exp(-\frac{1}{2\sigma_k^2}(x - \mu_k)^2)$$

Donde μ_k y σ_k^2 son la media y varianza para la clase k.

Análisis Discriminante lineal: estimación de π_k y $f_k(X)$

Si además se asume que la varianza es constante en todos los grupos $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2 = \sigma^2$ entonces el sumatorio $\sum_{j=1}^{K} \pi_j f_k(x)$ se simplifica en gran medida permitiendo calcular la probabilidad posterior según la ecuación:

$$P(Y = k|X = x) = \frac{\pi_k \frac{1}{\sqrt{2\pi\sigma}} \, exp(-\frac{1}{2\sigma^2}(x-\mu_k)^2)}{\sum_{j=1}^{K} \pi_j \frac{1}{\sqrt{2\pi\sigma}} \, exp(-\frac{1}{2\sigma^2}(x-\mu_j)^2)}$$

Esta ecuación se simplifica aun más mediante una transformación logarítmica de sus dos términos:

$$\hat{\delta}_k(x) = log(P(Y = k|X = x)) = x \, \frac{\mu_k}{\sigma^2} - \frac{\mu_k^2}{2\sigma^2} + log(\pi_k)$$

El término **lineal** en el nombre *Análisis discriminante lineal* se debe al hecho de que la función discriminatoria es lineal respecto de ”*X*”.

Análisis Discriminante lineal: estimación de π_k y $f_k(X)$

En la práctica, a pesar de tener una certeza considerable de que *X* se distribuye de forma normal dentro de cada clase, los valores $\mu_1 \ldots \mu_k, \pi_1 \ldots \pi_k$ y σ^2 se desconocen, por lo que tienen que ser estimados a partir de las observaciones. Las estimaciones empleadas en *LDA* son:

$$\hat{\mu}_k = \frac{1}{n_k} \sum_{i:y_1} x_i$$

$$\hat{\sigma}_k = \frac{1}{N-K} \sum_{k=1}^{K} \sum_{i:y_1} (x_i - \hat{\mu}_k)^2$$

$$\hat{\pi}_k = \frac{n_k}{N}$$

$\hat{\mu}_k$ es la media de las observaciones del grupo k, $\hat{\sigma}_k$ es la media ponderada de las varianzas muestrales de las K clases y $\hat{\pi}_k$ la proporción de observaciones de la clase k respecto al tamaño total de la muestra.

Análisis Discriminante lineal: estimación de π_k y $f_k(X)$

La clasificación de Bayes consiste en asignar cada observación $X = x$ a aquella clase para la que $P(Y = k|X = x)$ sea mayor. En el caso particular de una variable cualitativa Y con solo dos niveles, se puede expresar la regla de clasificación como un ratio entre las dos probabilidades posteriores. Se asignará la observación a la clase 1 si:

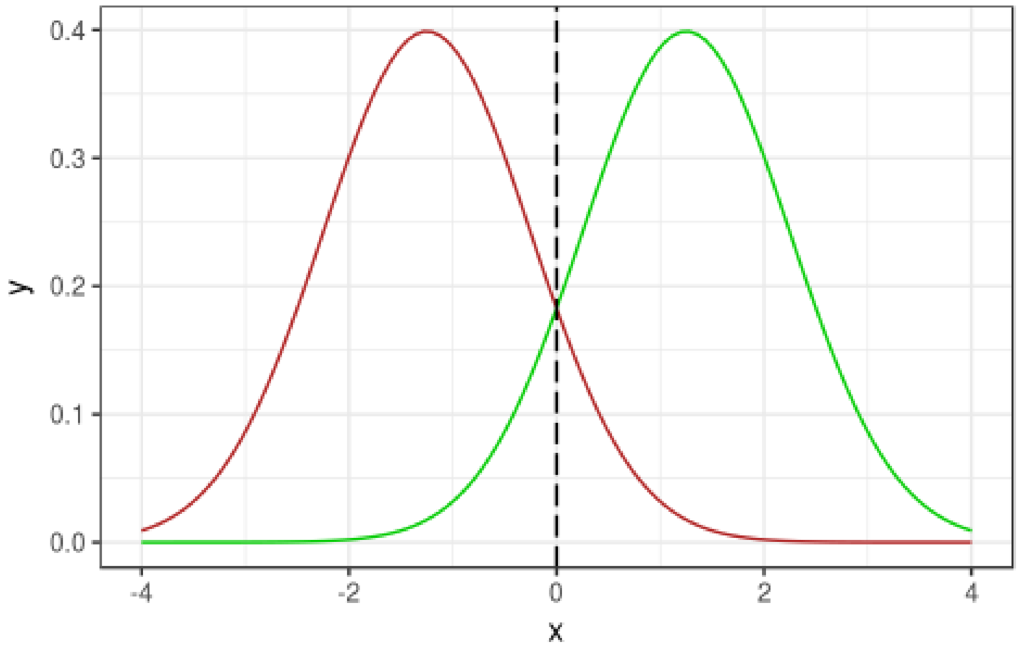
$$\frac{P(Y = 1|X = x)}{P(Y = 2|X = x)} > 1,$$

y a la clase 2 si es menor. En este caso particular el límite de decisión de Bayes viene dado por $x = \frac{\mu_1+\mu_2}{2}$.

La siguiente imagen muestra dos grupos distribuidos de forma normal con medias $\mu_1 = -1.25$ y $\mu_2 = 1.25$ y varianzas $\sigma_1^2 = \sigma_2^2 =$1. Dado que se conoce el valor real de las medias y varianzas poblacionales (esto en la realidad no suele ocurrir), se puede calcular el límite de decisión de Bayes $x = \frac{-1.25+1.12}{2} = 0$ → línea discontinua.

Análisis Discriminante lineal: estimación de π_k y $f_k(X)$

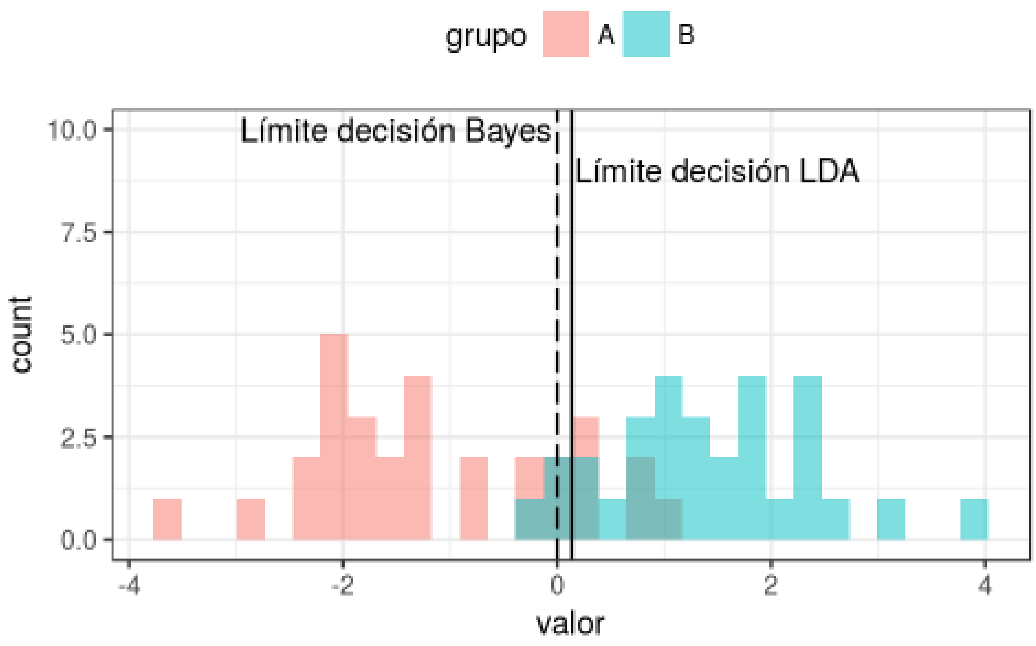
```
library(ggplot2)
ggplot(data.frame(x = c(-4, 4)), aes(x)) +
stat_function(fun = dnorm, args = list(mean = -1.25, sd = 1),
              color = "firebrick") +
stat_function(fun = dnorm, args = list(mean = 1.25, sd = 1), color = "green3") +
geom_vline(xintercept = 0, linetype = "longdash") +
theme_bw()
```



Análisis Discriminante lineal: estimación de π_k y $f_k(X)$

Si en lugar de conocer la verdadera distribución poblacional de cada grupo solo se dispone de muestras, escenario que suele ocurrir en los casos reales, el límite de decisión *LDA* se aproxima al verdadero límite de decisión de Bayes pero no es exacto. Cuanto más representativas sean las muestras mejor la aproximación.

```
set.seed(6911)
library(ggplot2)
grupo_a <- rnorm(n = 30, mean = -1.25, sd = 1)
grupo_b <- rnorm(n = 30, mean = 1.25, sd = 1)
datos <- data.frame(valor = c(grupo_a, grupo_b),
                    grupo = rep(c("A","B"), each = 30))

ggplot(data = datos, aes(x = valor, fill = grupo)) +
geom_histogram(alpha = 0.5, position = "identity") +
geom_vline(xintercept = 0, linetype = "longdash") +
geom_vline(xintercept = (mean(grupo_a) + mean(grupo_b))/2)  +
annotate(geom = "text", x = 1.5, y = 9, label = "Límite decisión LDA") +
annotate(geom = "text", x = -1.5, y = 10, label = "Límite decisión Bayes") +
theme_bw() +
theme(legend.position = "top")
```



Análisis Discriminante lineal: extensión LDA múltiples predictores

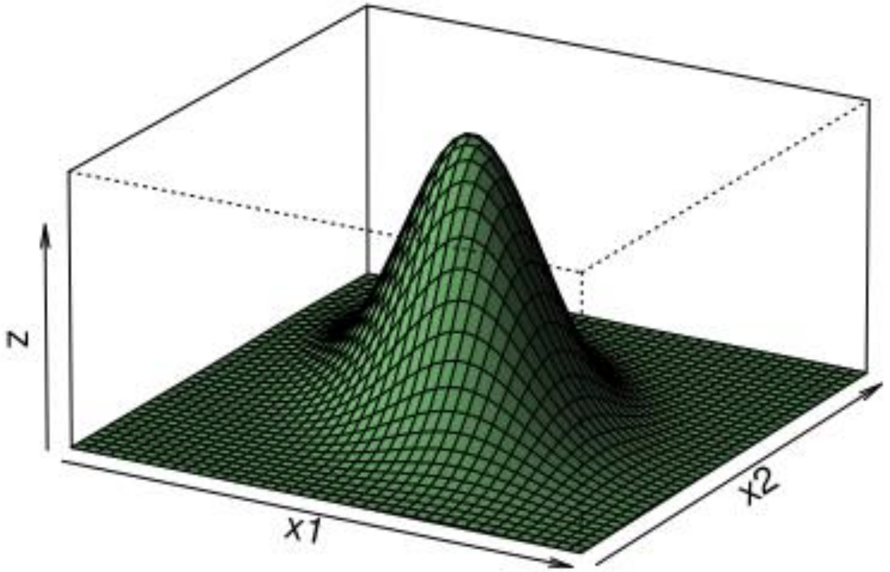
Los conceptos anteriormente descritos empleando un único predictor pueden generalizarse para introducir múltiples predictores en el modelo. La diferencia reside en que X, en lugar de ser un único valor, es un vector formado por el valor de *p* predictores $X = (X_1, X_2, \dots, X_n)$ y que, en lugar de proceder de una distribución normal, procede de una distribución normal multivariante.

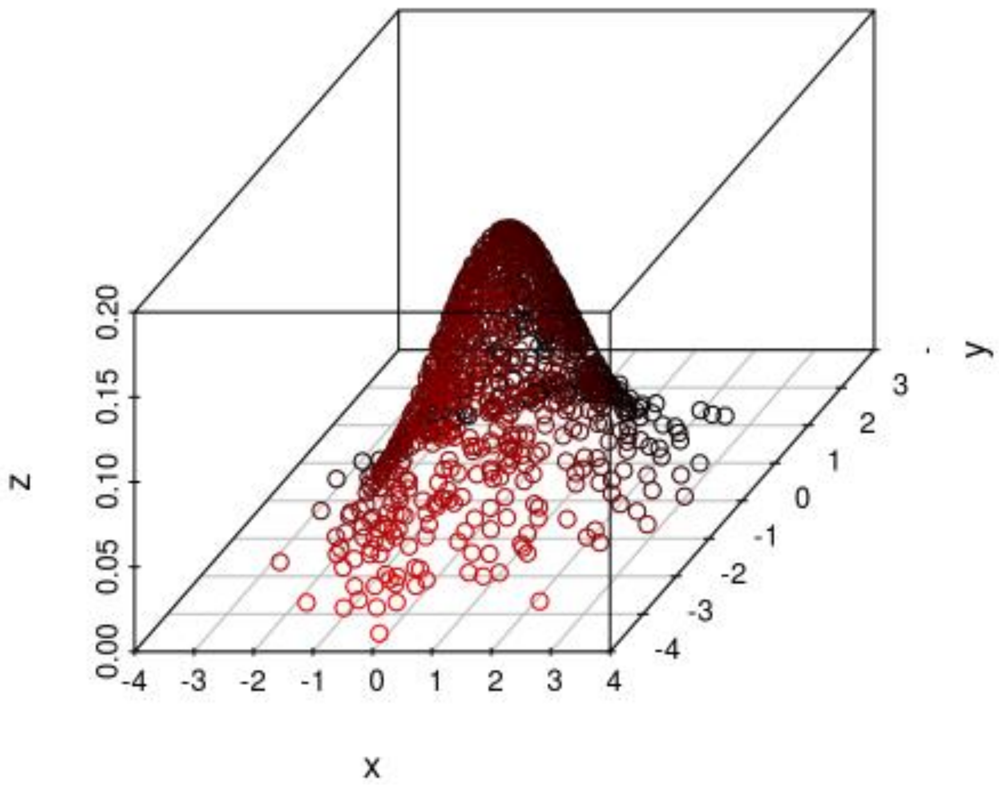
Un vector sigue una distribución k-normal multivariante si cada uno de los elementos individuales que lo forman sigue una distribución normal y lo mismo para toda combinación lineal de sus k elementos.

Las siguientes imágenes muestran representaciones gráficas de distribuciones normales multivariante de 2 elementos (distribución normal bivariante).

Análisis Discriminante lineal: extensión LDA múltiples predictores

Distribución multivariante con dos predictores





Análisis Discriminante lineal: extensión LDA múltiples predictores

Para indicar que una variable aleatoria *p*-dimensional *X* sigue una distribución normal multivariante se emplea la terminología $X \sim N(\mu, \Sigma)$. Donde μ es el vector promedio de X y Σ es la covarianza de *X*, que al ser un vector con *p* elementos, es una matriz *p x p* con la covarianza de cada par de predictores. La ecuación que define la función de densidad de una distribución normal multivariante es:

$$f(x) = \frac{1}{(2\pi)^{p/2} |\sum|^{1/2}} \, exp \, (-\frac{1}{2}(x-\mu)^T \sum{}^{-1} (x-\mu))$$

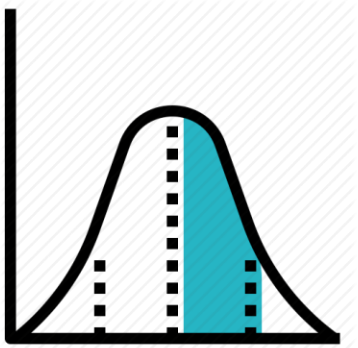
Si se sigue el mismo procedimiento que el mostrado para *LDA* con un solo predictor, pero esta vez con la ecuación de multivariante normal, y se asume que la matriz de covarianzas Σ es igual para las *K* clases, se obtiene que el clasificador de Bayes es:

$$\hat{\delta}_k(x) = log(P(Y=k|X=x)) = x^T \sum{}^{-1} \mu_k - \frac{1}{2}\mu_k^T \sum{}^{-1} \mu_k + log(\pi_k)$$

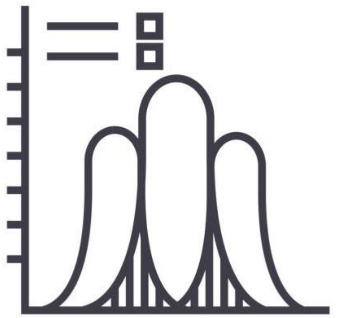
Cuando los parámetros poblacionales se desconocen, no se puede calcular el límite de decisión de Bayes exacto, por lo que se recurre a la estimación de $\mu_1 \ldots \mu_k, \pi_1 \ldots \pi_k$ y Σ para obtener los límites de decisión de *LDA*.

Análisis Discriminante lineal: condición de LDA

- Las condiciones que se deben cumplir para que un Análisis Discriminante Lineal sea válido son:



1. Cada predictor que forma parte del modelo se distribuye de forma normal en cada una de las clases de la variable respuesta. En el caso de múltiples predictores, las observaciones siguen una distribución normal multivariante en todas las clases.
2. La varianza del predictor es igual en todas las clases de la variable respuesta. En el caso de múltiples predictores, la matriz de covarianza es igual en todas las clases. Si esto no se cumple se recurre a Análisis Discriminante Cuadrático (*QDA*).

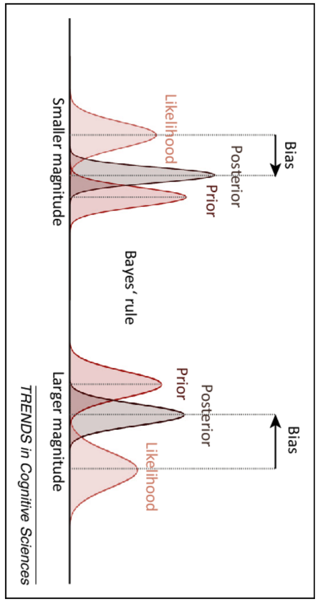


- Cuando la condición de normalidad no se cumple, el LDA pierde precisión pero aun así puede llegar a clasificaciones relativamente buenas. *Using discriminant analysis for multi-class classification: an experimental investigation (Tao Li, Shenghuo Zhu, Mitsunori Ogihara)*.

Análisis Discriminante lineal: LDA Bayes + Fischer

Existen varios enfoques posibles para realizar un *LDA.* La aproximación descrita anteriormente está basada en el clasificador de Bayes, y utiliza todas las variables originales para calcular las probabilidades posteriores de que una observación pertenezca a cada grupo.

Antes de que el clasificador de Bayes fuese introducido en el *LDA*, Fisher propuso una aproximación en la que el espacio *p*-dimensional (donde *p* es el número de predictores originales) se reduce a un subespacio de menos dimensiones formado por las combinaciones lineales de los predictores que mejor explican la separación de las clases. Una vez encontradas dichas combinaciones se realiza la clasificación en este subespacio. Fisher definió como subespacio óptimo a aquel que maximiza la distancia entre grupos en términos de varianza. Los términos de *discriminante lineal de Fisher* y *LDA*son a menudo usados para expresar la misma idea, sin embargo, el artículo original de Fisher describe un discriminante ligeramente diferente, que no hace algunas de las suposiciones del *LDA* como la de una distribución normal de las clases o covarianzas iguales entre clases.

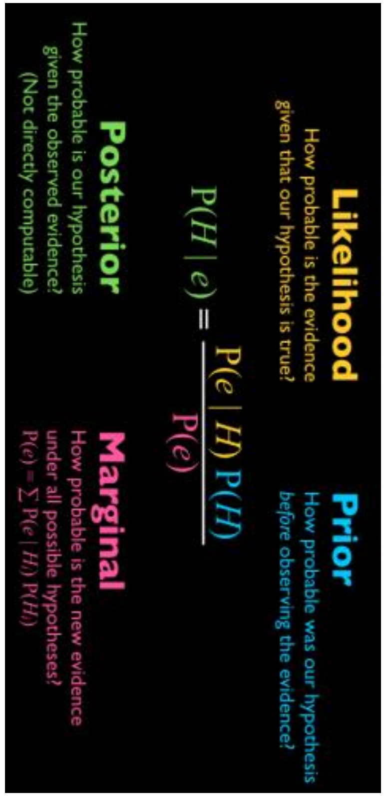


Análisis Discriminante lineal: LDA Bayes + Fischer

La aproximación de Fisher se puede ver como un proceso con dos partes:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\sum_{i=1}^{n}[P(B|A_i)P(A_i)]}$$

- Reducción de dimensionalidad: Se pasa de p variables predictoras originales a k combinaciones lineales de dichos predictores (variables discriminantes) que permiten explicar la separación de los grupos pero con menos dimensiones ($k < p$).
- Clasificación de las observaciones empleando las variables discriminantes.



Los resultados de clasificación obtenidos mediante el método de Fisher son iguales a los obtenidos por el método de Bayes cuando:

- En el método de Bayes se asume que la matriz de covarianzas en igual en todos los grupos y se emplea como estimación la *pooled within-class covariance matrix*.
- En el método de Fisher, todos los discriminantes lineales se utilizan para la clasificación. El número máximo de discriminantes obtenido tras la reducción de dimensionalidad es *número grupos*-1.

Análisis Discriminante lineal: LDA Bayes + Fischer

Para mayor referencia:

Bayes Optimality in Linear Discriminant Analysis Onur C. Hamsici and Aleix M. Martinez

Generalizing Fisher's linear discriminant analysis via the SIR approach, Chapter 14

http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema1dm.pdf

http://www2.stat.unibo.it/montanari/Didattica/Multivariate/Discriminant_analysis.pdf

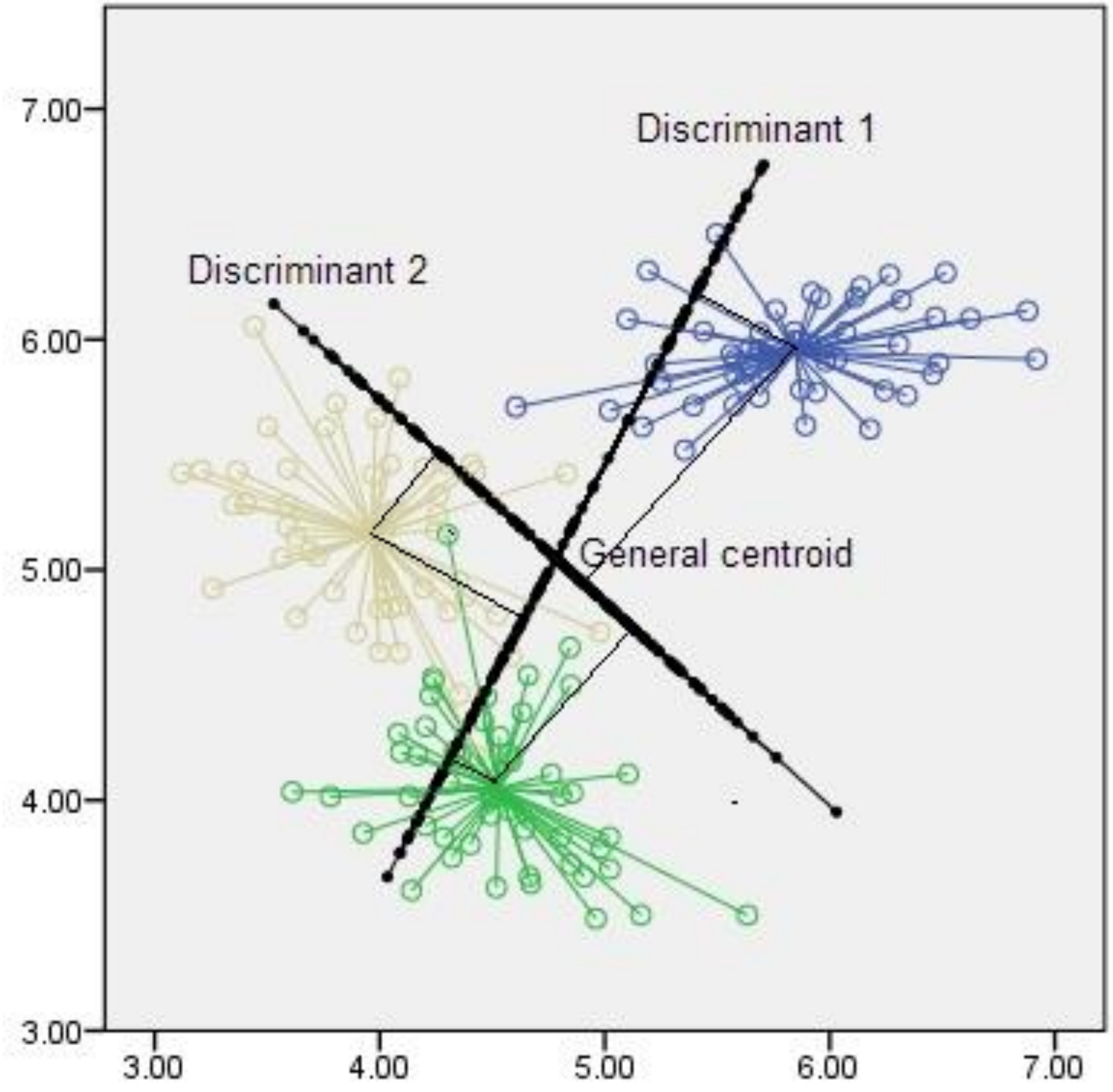
Análisis Discriminante lineal: precisión del LDA

- Una vez que las normas de clasificación se han establecido, se tiene que evaluar como de buena es la clasificación resultante. En otras palabras, evaluar el porcentaje de aciertos en las clasificaciones.

- Las matrices de confusión son una de las mejores formas de evaluar la capacidad de acierto que tiene un modelo *LDA*. Muestran el número de verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos. El método *LDA* busca los límites de decisión que más se aproximan al clasificador de Bayes, que por definición, tiene el menor ratio de error total de entre todos los clasificadores (si se cumple la condición de normalidad). Por lo tanto, el *LDA* intenta conseguir el menor número de clasificaciones erróneas posibles, pero no diferencia entre falsos positivos o falsos negativos. Si se quiere intentar reducir el número de errores de clasificación en una dirección determinada (por ejemplo, menos falsos negativos) se puede modificar el límite de decisión, aunque como consecuencia aumentará el número de falsos positivos.

- Cuando para evaluar el error de clasificación se emplean las mismas observaciones con las que se ha creado el modelo, se obtiene lo que se denomina el *training error*. Si bien esta es una forma sencilla de estimar la precisión en la clasificación, tiende a ser excesivamente optimista. Es más adecuado evaluar el modelo empleando observaciones nuevas que el modelo no ha visto, obteniendo así el *test error*.

Análisis Discriminante lineal: la regla de decisión



Restricciones del LDA

- El número máximo de funciones discriminantes es igual al mínimo entre el número de variables y el número de grupos menos 1 (con q grupos, (q − 1) funciones discriminantes).

- Las matrices de covarianzas dentro de cada grupo deben ser aproximadamente iguales. Las variables continuas deben seguir una distribución normal multivariante.

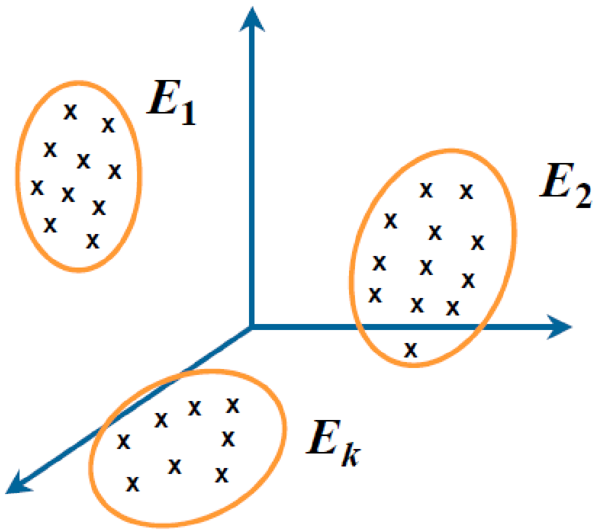
- A partir de *q* grupos donde se asignan a una serie de objetos y de k variables medidas sobre ellos (x1, . . . , xp), se trata de obtener para cada objeto una serie de puntuaciones que indican el grupo al que pertenecen (y1, . . . , ym), de modo que sean funciones lineales de x1, . . . , xp.

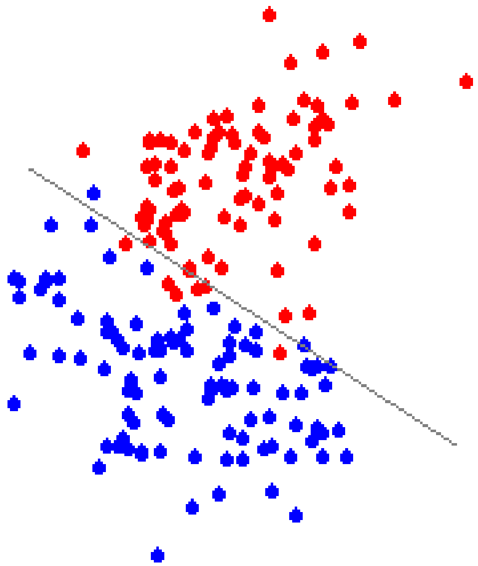
$$
\begin{aligned}
y_1 &= a_{11}x_1 + \cdots + a_{1p}x_p + a_{10} \\
&\cdots\cdots\cdots \\
y_m &= a_{m1}x_1 + \cdots + a_{mp}x_p + a_{m0}
\end{aligned}
$$

Modelo matemático

- Se busca que las funciones y_i discriminen o separen lo mejor posible a los grupos.

- Estas combinaciones lineales de las ***k*** variables deben maximizar la varianza entre los grupos y minimizar la varianza dentro de los grupos.





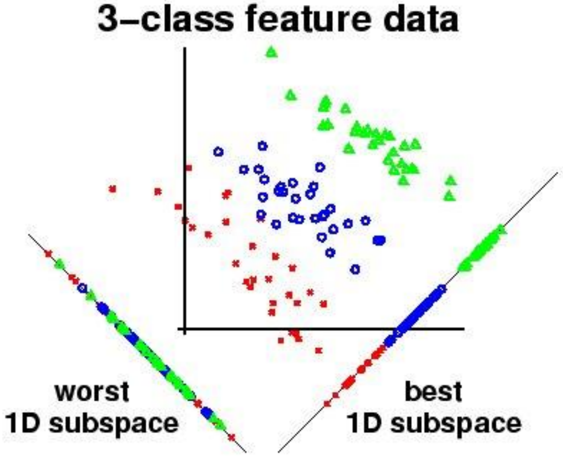
Modelo matemático

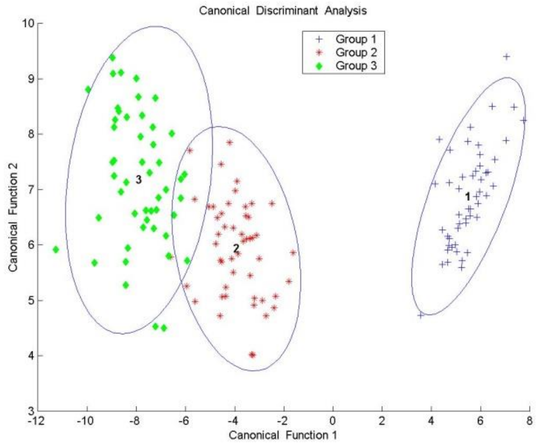
- Se puede descomponer la variabilidad total de la muestra en variabilidad dentro de los grupos y entre los grupos.

- Se busca que la covarianza total, que es igual a la covarianza dentro de grupos más la covarianza entre grupos, sea lo más pequeña entre los grupos, y lo más grande dentro de los grupos.

- Esto es, que todo se pueda explicar dentro de los grupos, y no entre los grupos.

- Esto produce una separación correcta entre los grupos.





Modelo matemático

- En notación matricial esto es equivalente a

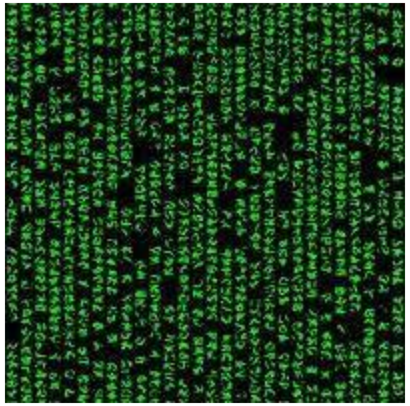
T = E + D

donde:

T = matriz de covarianzas total

E = matriz de covarianzas entre grupos

D = matriz de covarianzas dentro de grupos.



Extracción de las funciones discriminantes

- La idea básica del Análisis Discriminante consiste en extraer a partir de x1, . . . , xk variables observadas en k grupos, m funciones y1, . . . , ym de forma:

$$y_i = a_{i1}x_1 + \cdots + a_{ip}x_p + a_{i0}$$

- Las funciones y1, . . . , ym se extraen de modo que:

1. y1 sea la combinación lineal de x1, . . . , xp que proporciona la mayor discriminaciónposible entre los grupos.
2. y2 sea la combinación lineal de x1, . . . , xp que proporciona la mayor discriminación posible entre los grupos, después de y1, tal que corr(yi, y2) = 0. Se supone independencia entre las reglas.

Extracción de las funciones discriminantes

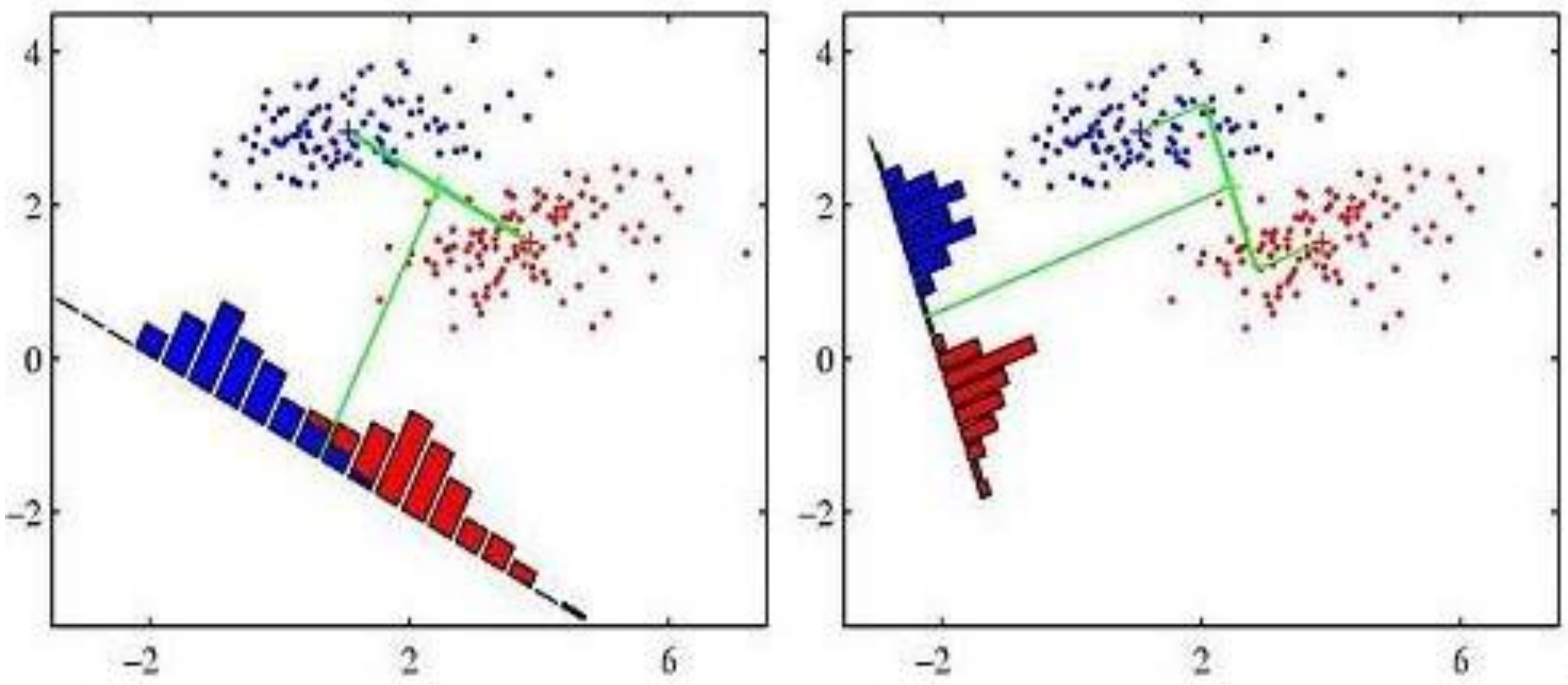
- En general, y_i es la combinación lineal de x1, . . . , xp que proporciona la mayor discriminación posible entre los grupos después de yi−1 y tal que corr(yi, yj) = 0 para j = 1, . . . , (i − 1). Lo que rige la cantidad de reglas o ecuaciones yi, es el total de variancia explicada por cada uno de las reglas.

- Una cantidad excesivas de reglas podría sesgar el modelo, y muy pocas reglas podría no explicar el total de casos. Según las ecuaciones siguientes:

$$\begin{aligned} y_1 &= a_{11}x_1 + \cdots + a_{1p}x_p + a_{10} \\ & \ldots\ldots\ldots \\ y_m &= a_{m1}x_1 + \cdots + a_{mp}x_p + a_{m0} \end{aligned}$$

- Las anteriores suponen una relación lineal para poder discriminar los grupos. Sin embargo, existen relaciones cuadráticas, y de otros tipos para tratar de mejorar la variancia explicada en los grupos.

Extracción de las funciones discriminantes



Índice

Índice

4

Análisis discriminante cuadrático

- Idea del QDA
- QDA de 2 predictores

Análisis Discriminante Cuadrático: idea

El clasificador cuadrático o *Quadratic Discriminat Analysis QDA* se asemeja en gran medida al *LDA*, con la única diferencia de que el *QDA* considera que cada clase *k* tiene su propia matriz de covarianza (Σ_k) y omo consecuencia, la función discriminante toma forma cuadrática:

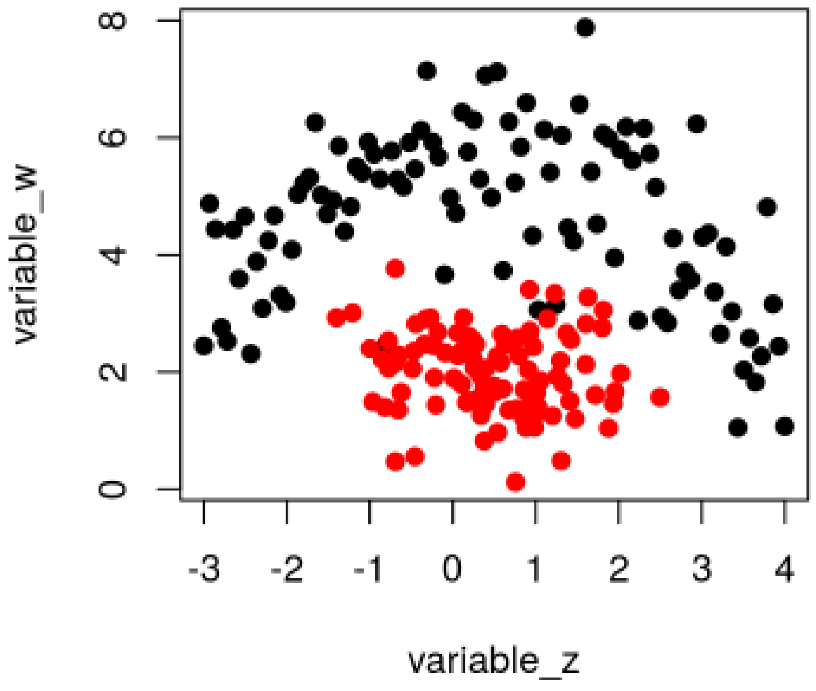
$$log(P(Y=k|X=x)) = -\frac{1}{2}\log|\Sigma_k| - \frac{1}{2}(x-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) + \log(\pi_k)$$

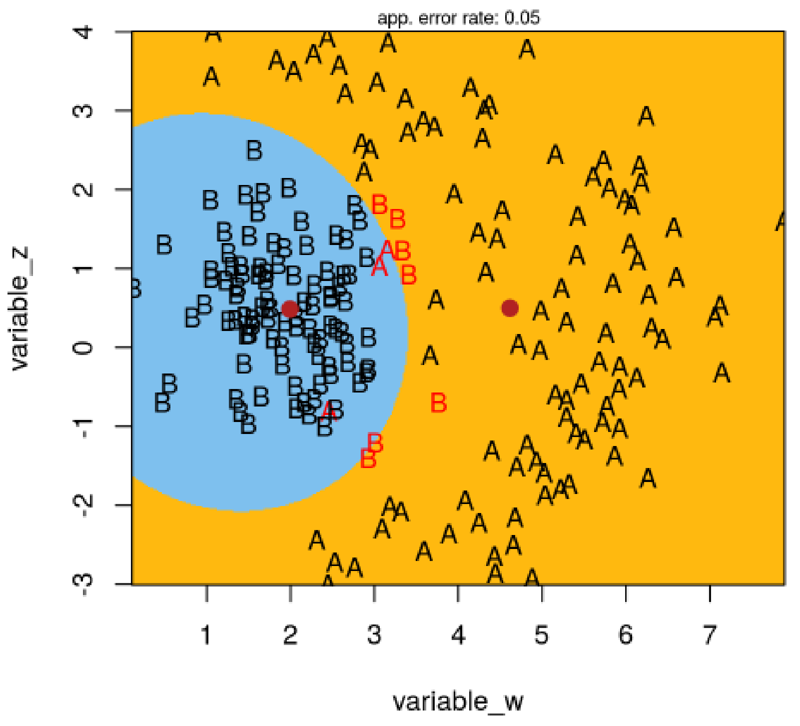
Para poder calcular la *posterior probability* a partir de esta ecuación discriminante es necesario estimar, para cada clase, (Σ_k), μ_k y π_k, a partir de la muestra. Cada nueva observación se clasifica en aquella clase para la que el valor de la probabilidad posterior sea mayor.

QDA genera límites de decisión curvos por lo que puede aplicarse a situaciones en las que la separación entre grupos no es lineal.

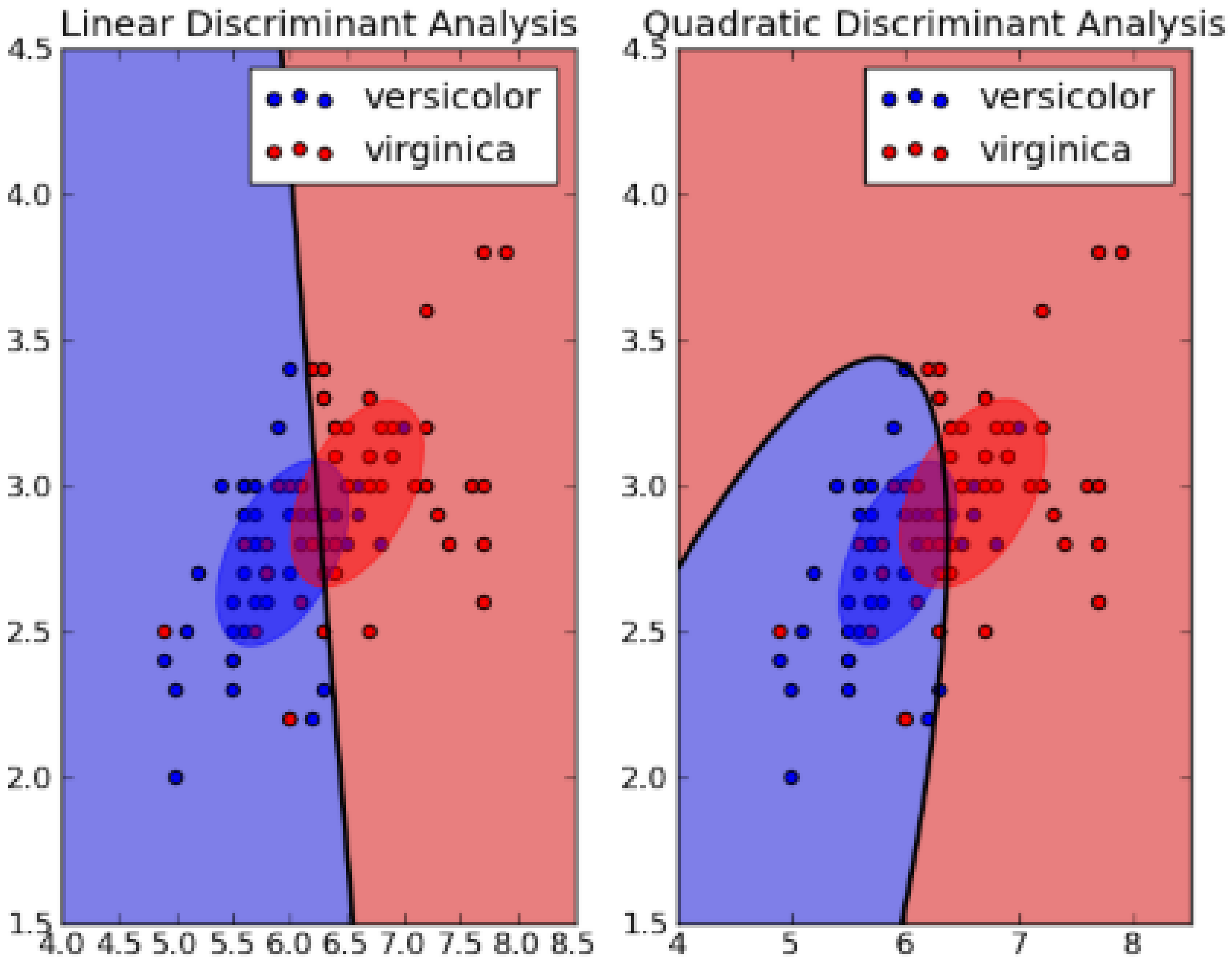
Análisis Discriminante Cuadrático: dos predictores

La gran diferencia es que para el QDA, veremos que puede que la separación idónea no sea lineal, y más bien podría existir alguna curvatura que indique la mejor aproximación de una separación polinomial. Los siguientes dos gráficos dan a entender lo anterior.





Análisis Discriminante Cuadrático



Índice

Comparación entre LDA y QDA

- ¿Qué clasificador es más adecuado? Depende de las implicaciones que tiene, en el balance *bias-varianza*, el asumir que todos los grupos comparten una matriz de covarianza común. *LDA* produce límites de decisión lineales, lo que se traduce en menor flexibilidad y por lo tanto menor problema de varianza. Sin embargo, si la separación de los grupos no es lineal, tendrá un bias grande. El método *QDA* produce límites cuadráticos y por lo tanto curvos, lo que aporta mayor flexibilidad permitiendo ajustarse mejor a los datos, menor bias pero mayor riesgo de varianza.

- En términos generales, *LDA* tiende a conseguir mejores clasificaciones que *QDA* cuando hay pocas observaciones con las que entrenar al modelo, escenario en el que evitar la varianza es crucial. Por otra parte, si se dispone de una gran cantidad de observaciones de entrenamiento o si no es asumible que existe una matriz de covarianza común entre clases, *QDA* es más adecuado.

- Si se dispone de p predictores, calcular una matriz de covarianza común requiere estimar $p(p+1)/2$parámetros, mientras que calcular una matriz diferente para cada grupo requiere de $Kp(p+1)/2$. Para valores de p muy altos, la elección del método puede estar limitada por la capacidad computacional.

Índice

Índice

6

Aplicación en R

- Funciones
- Estadísticas descriptivas y supuestos
- Partición de los datos
- Estimación DA: LDA y QDA
- Reglas de discriminación
- Partición y proyección
- Tablas de confusión

Resultados en R: funciones

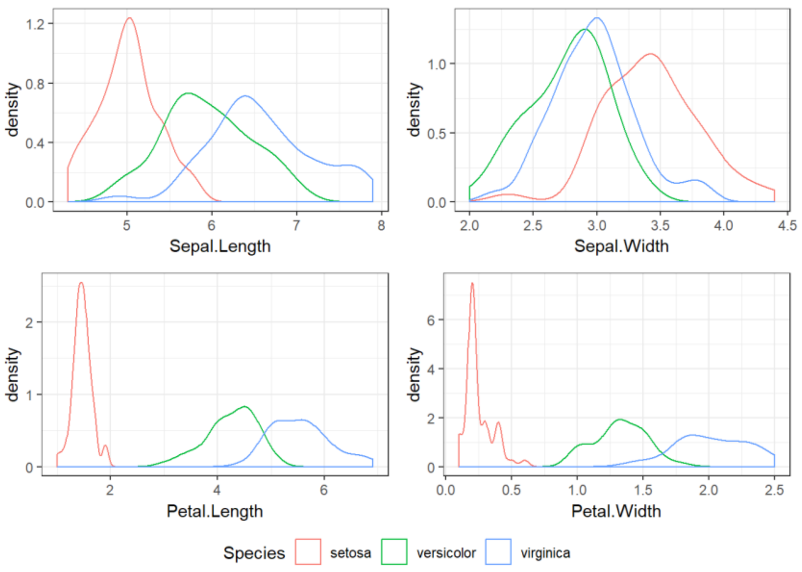
- Se utilizarán las funciones lda() “linear discrimination analyis” y qda() “quadratic discrimination analyis”, ambas de la librería MASS.

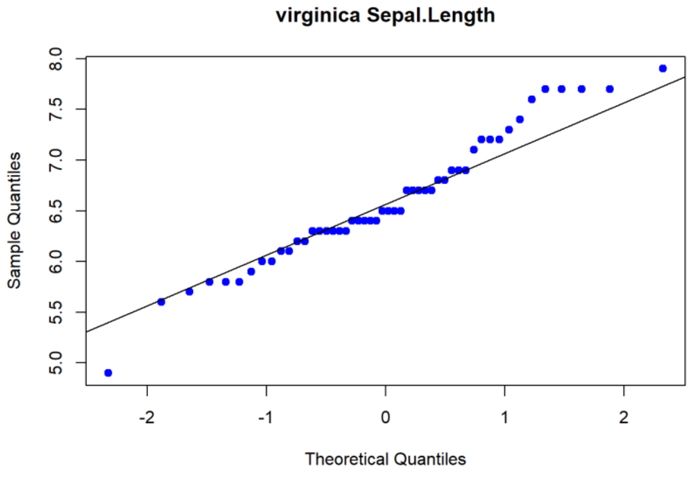
```
# lda(x, …)
#  S3 method for formula
# lda(formula, data, …, subset, na.action)

# S3 method for default
# lda(x, grouping, prior = proportions, tol = 1.0e-4,
#    method, CV = FALSE, nu, …)

# S3 method for data.frame
# lda(x, …)

#  S3 method for matrix
#  lda(x, grouping, …, subset, na.action)
```

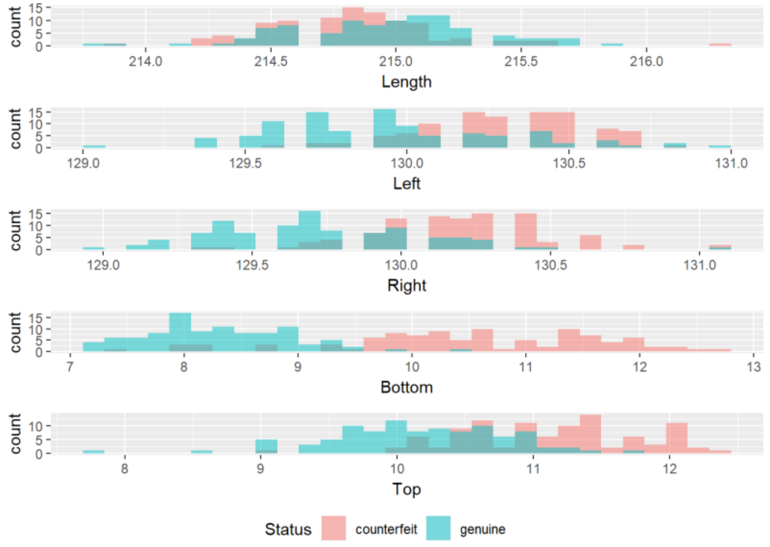
```
# qda(x, …)
#  S3 method for formula
# qda(formula, data, …, subset, na.action)

# S3 method for default
# qda(x, grouping, prior = proportions,
#     method, CV = FALSE, nu, …)

#  S3 method for data.frame
#  qda(x, …)

# S3 method for matrix
#  qda(x, grouping, …, subset, na.action)
```

Resultados en R: Estadísticas descriptivas y supuestos

- Se suelen analizar diversas estadísticas descriptivas (correlaciones, etc.), además del cumplimiento de los supuestos de normalidad y variancia constante (homogeneidad).

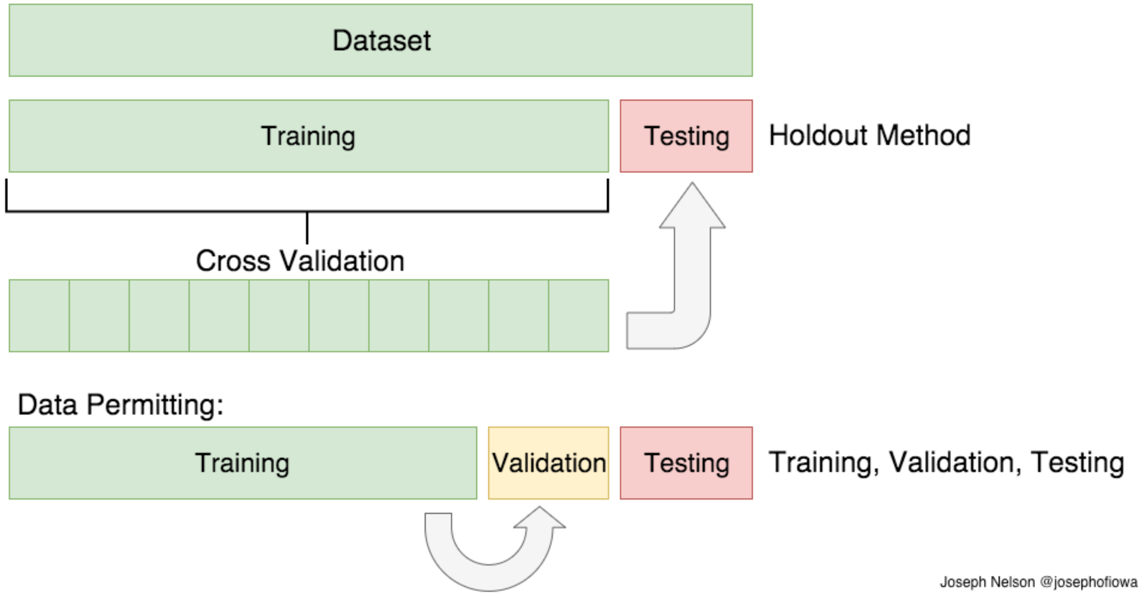






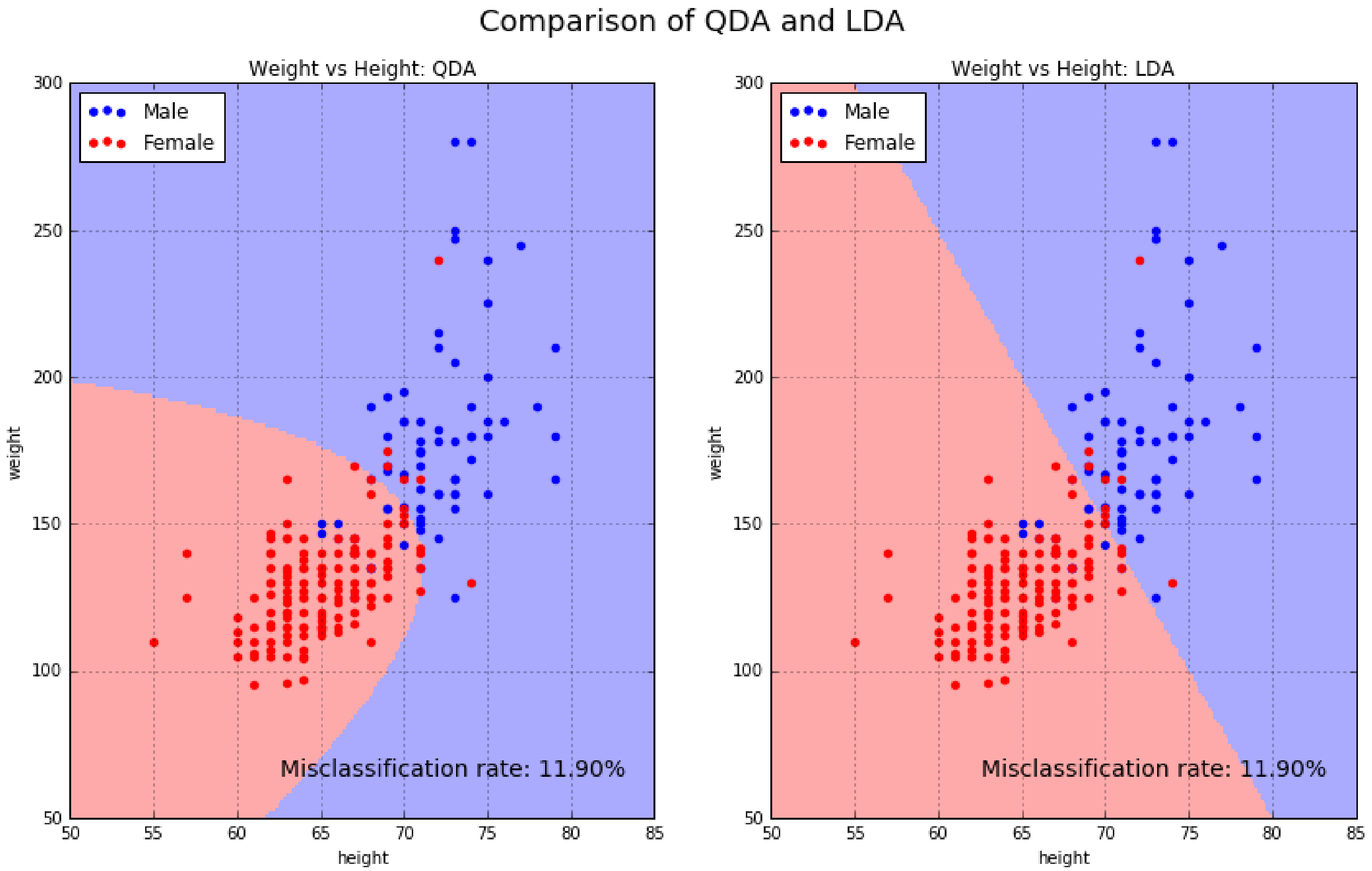
Resultados en R: partición de los datos

- Con tal de validar que el DA es robusto, y por ende separa correctamente, dividimos los datos en dos: entrenamiento y prueba.



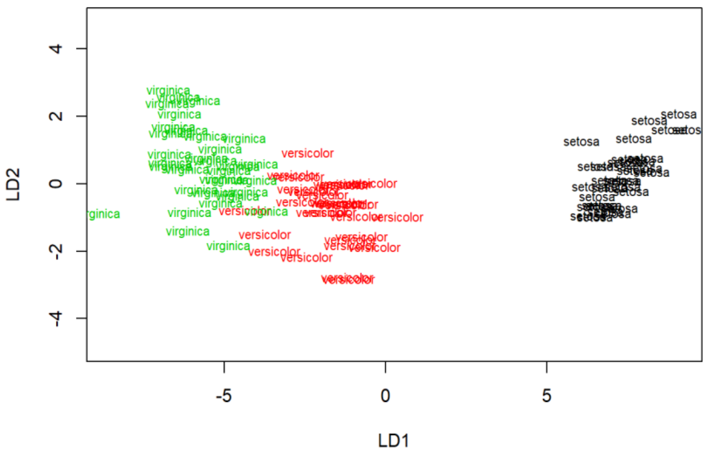
Resultados en R: estimación DA (LDA y QDA)

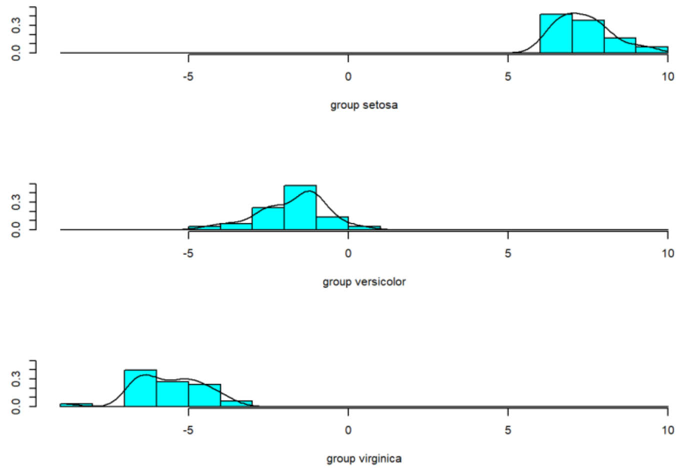
- El análisis por DA se puede llevar a cabo de forma lineal o cuadrática.

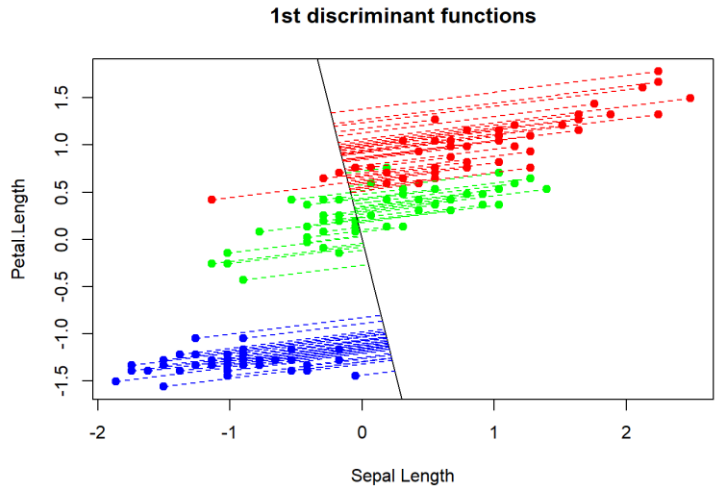


Resultados en R: análisis reglas de discriminación

En el análisis de la regla construida por los LDA y QDA, se puede estudiar las proyección según las reglas en los individuos, la distribución de la separación, o las proyecciones de las reglas.

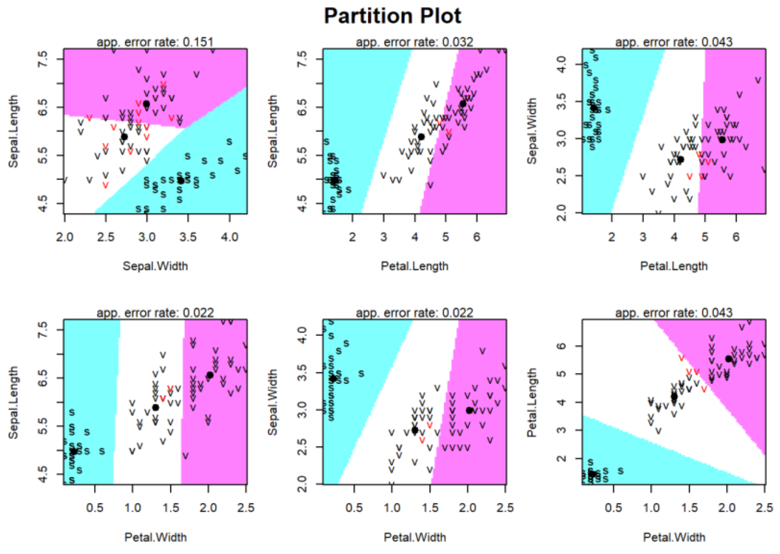


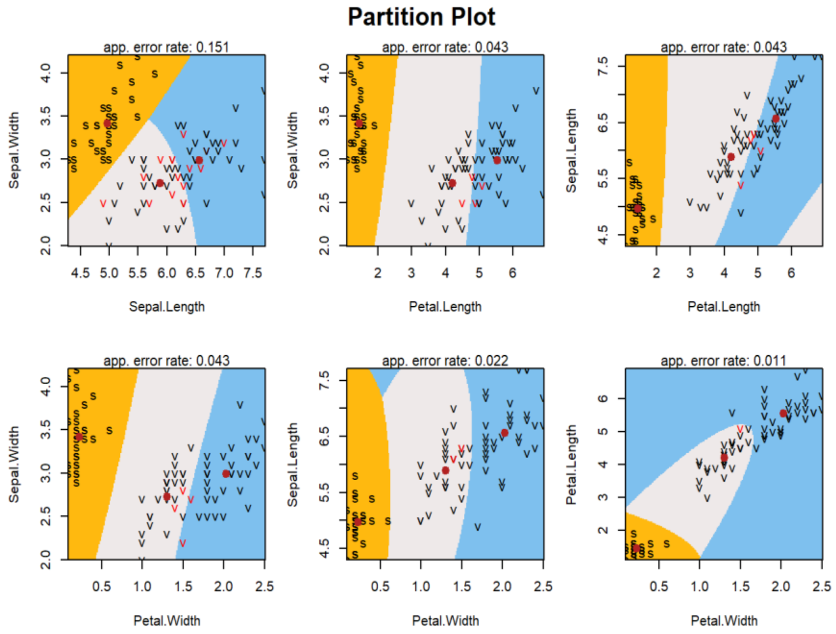


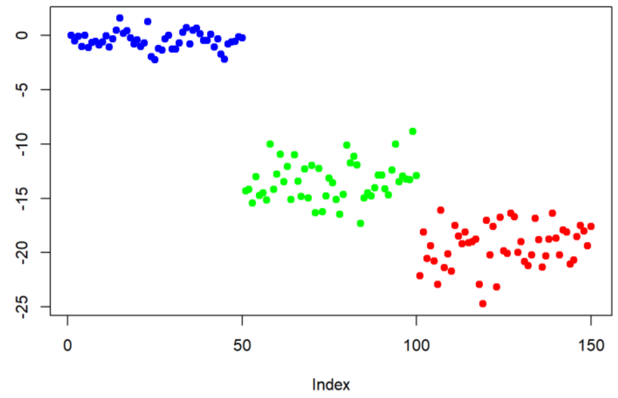


Resultados en R: partición y proyección

El análisis del poder discriminativo de los modelos se puede llevar a cabo mediante gráficos de partición y de proyección







Resultados en R: tablas de confusión

Finalmente, se debe analizar las tablas de confusión tanto del set de entrenamiento como para el de validación.

		Predicted		
		No	Yes	Classification Error
Actual	No	**True Negative:** Predicted No, and target was No.	**False Positive:** Predicted Yes, and target was No.	(TN)/(TN+FP) Model incorrectly predicts 'No' 28% of the time
	Yes	**False Negative**: Predicted No, and target was Yes.	**True Positive:** Predicted Yes, and target was Yes.	(TP)/(FN+TP) Model incorrectly predicts 'Yes' 32% of the time

Índice

Comparación entre LDA y QDA

Leer el artículo "*La importancia de ser humano*" de W.W.Howells, Universidad de Harvard.

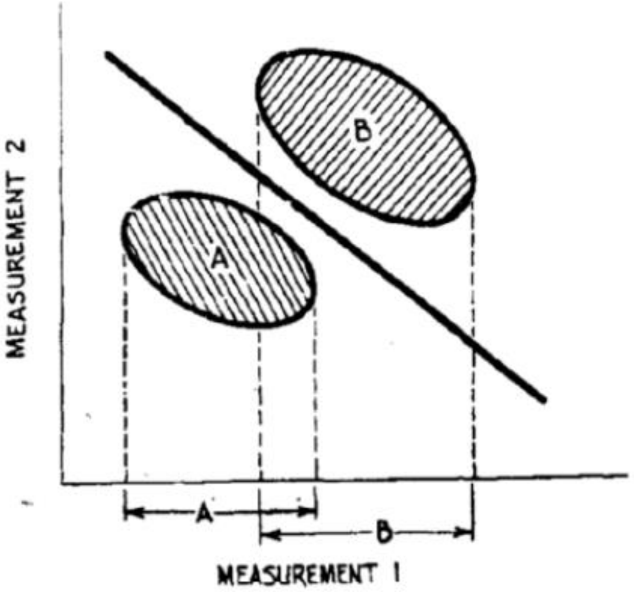


Gráfico 1
Dos mediciones juntas separan los grupos mejor que cualquiera de ellas separadamente

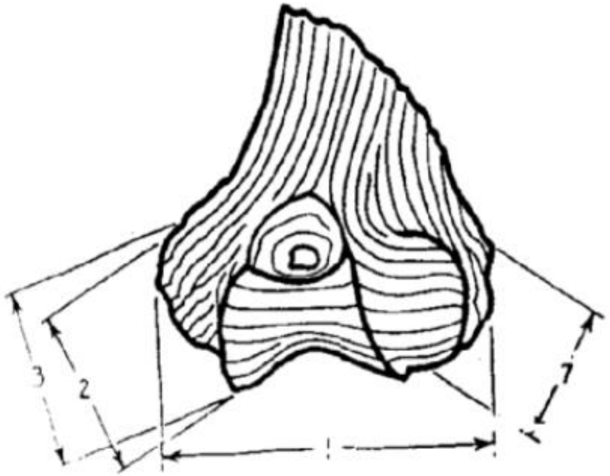


Gráfico 2
Fragmento humeral de Kanapoi y mediciones tomadas

Índice

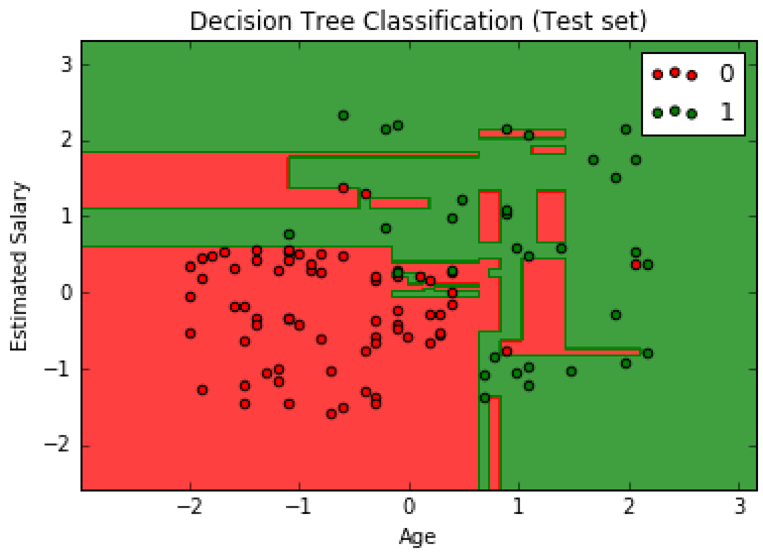
Conclusión: procedimiento del DA

En la realización del DA, se suelen llevar a cabo los siguiente pasos:

1. Estadísticas descriptivas de las variables.
2. Análisis de los supuestos.
3. Partición de los datos.
4. Estimación de los modelos lda y qda.
5. Análisis reglas de discriminación.
6. Gráficos de partición y proyección.
7. Tablas de confusión.
8. Nuevos casos sin categoría de pertenencia.

Conclusión: ¿qué son los nuevos casos?

- Estos representan observaciones que no son clasificadas en ninguna categoría, pero que poseen las observaciones o variables para poder ser asignados a alguna categoría de la variable de clasificación. Al trabajar con este tipo de análisis, es posible que una observación nueva pudiera poseer los valores cuantitativos, y no así un grupo de procedencia. La regla de decisión construida permite determinar el grupo de pertenencia de un nuevo grupo. Por ejemplo, en un nuevo crédito bancario, las características de las personas permiten clasificar si esté podrá ser acreedor de un préstamo.



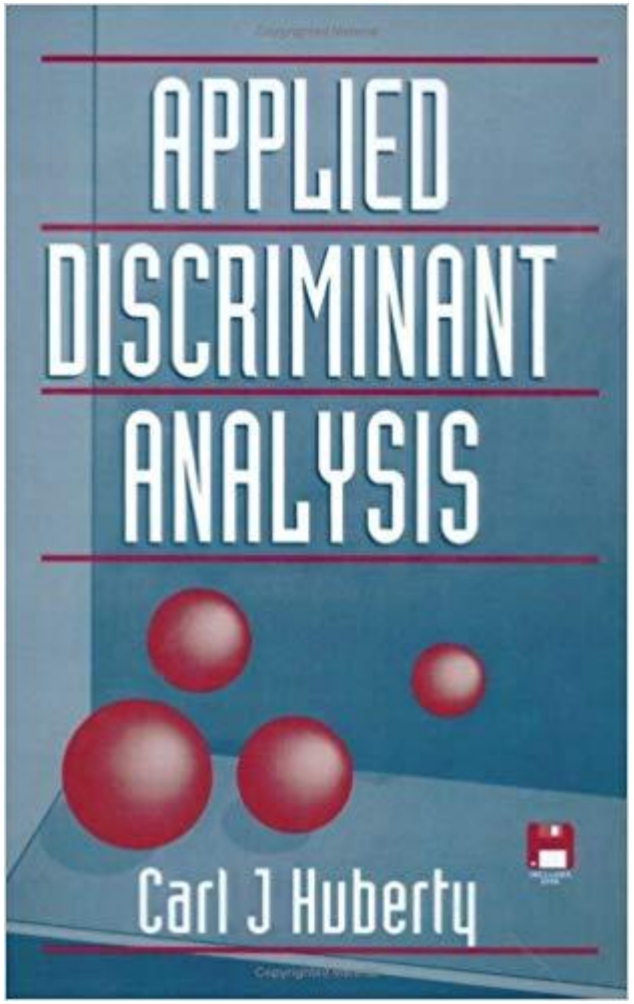
Conclusión: material de consulta

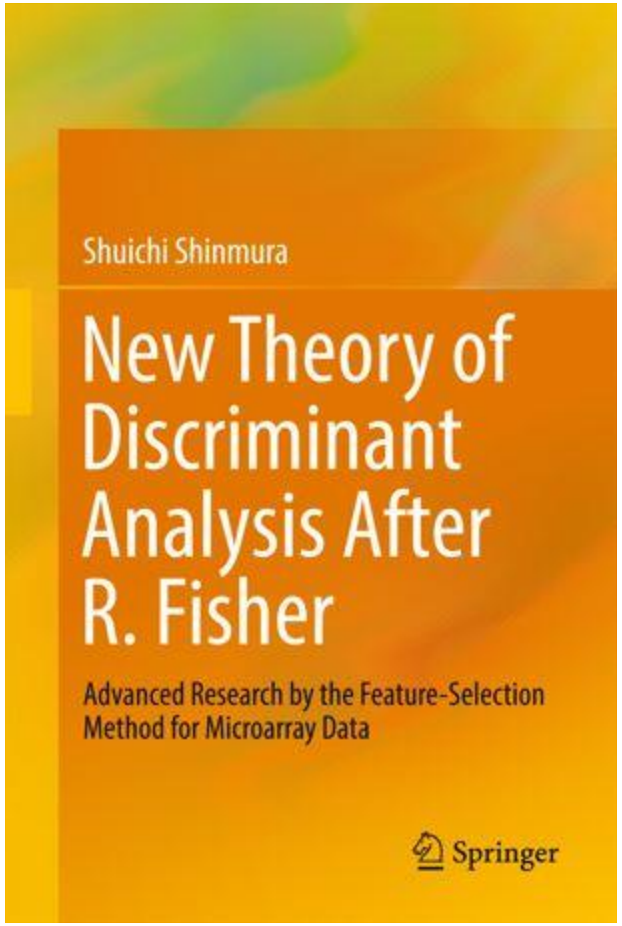
- Se pueden consultar los siguientes enlaces:

https://rpubs.com/Joaquin_AR/233932
https://rpubs.com/shpotes/277311
https://www.ecured.cu/Analisis_discriminante_en_R
http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/DM/tema1dm.pdf
https://www.statmethods.net/advstats/discriminant.html
https://rpubs.com/Nolan/298913
https://rpubs.com/ifn1411/LDA
https://rstudio-pubsstatic.s3.amazonaws.com/35817_2552e05f1d4e4db8ba87b334101a43da.html
https://github.com/avrilcoghlan/LittleBookofRMultivariateAnalysis/blob/master/src/multivariateanalysis.rst

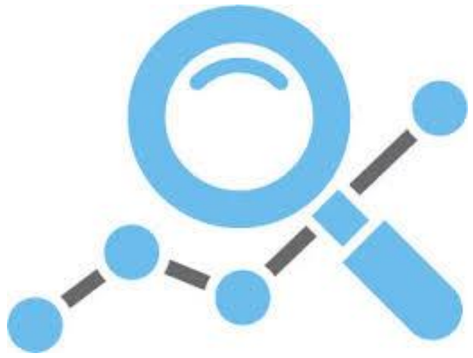
Conclusión: material de consulta

Si desean abordar aún más el DA, se recomiendan los siguientes libros:





Conclusión



- Se presentó el análisis de discriminación como una primera alternativa a la familia de técnicas de separación.
- El interés del DA se centra más en encontrar o separar a las observaciones pertenecientes a cierta categoría con tal de medir dicha faceta de discriminación y así poder generalizar el contexto de estudio.
- Visto como una técnica de análisis supervisado, se recomienda es estudio de otros métodos de clasificación:

- https://medium.com/@Mandysidana/machine-learning-types-of-classification-9497bd4f2e14

The End