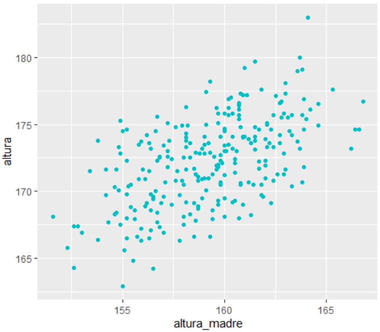
Clasificación

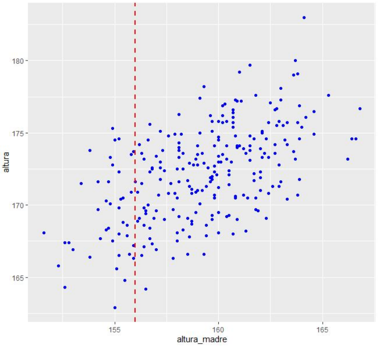
IECD

Problema de Alturas: Problema de Predicción

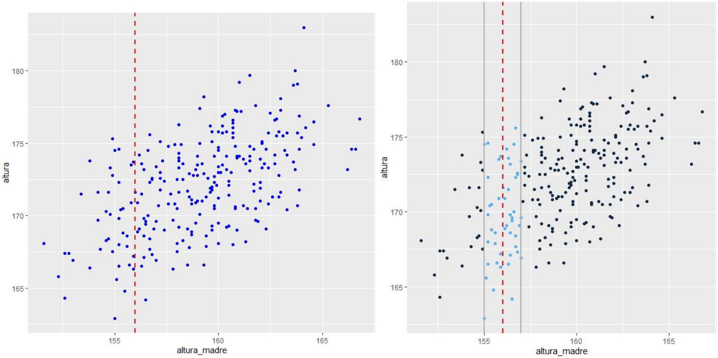


- Marco de Regresión
  1. **Variable de respuesta (Outcome)**
     Y: altura de la hija
  2. **Covariable (independiente, explicativa, regresora, predictora, feature)**:
     X: altura de la madre

Estimación No Paramétrica de la Regresión



Estimación No Paramétrica de la Regresión



Otro escenario: ¿Le damos un crédito?

Se quiere predecir si un cliente de un banco pagará un crédito.

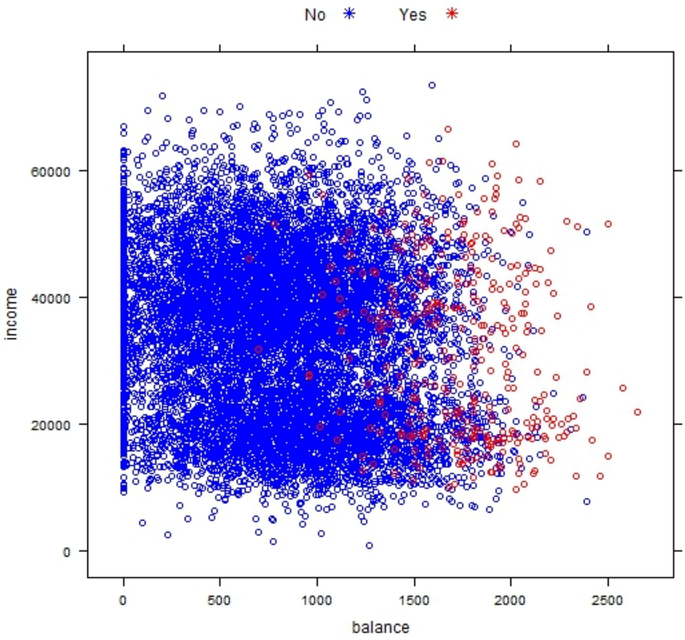
Variables registradas:

- X_1: **balance** (saldo tarjeta)
- X_2: **income** (ingreso anual)
- X_3: **student** (si - no)

- $Y = \begin{cases} 1 & \text{(Yes)} \quad \text{default} \\ 0 & \text{(No)} \quad \text{c.c.} \end{cases}$

Paquete ISLR

```
> Default
   default student    balance    income
1       No      No  729.52650 44361.625
2       No     Yes  817.18041 12106.135
3       No      No 1073.54916 31767.139
4       No      No  529.25060 35704.494
5       No      No  785.65588 38463.496
6       No     Yes  919.58853  7491.559
7       No      No  825.51333 24905.227
8       No     Yes  808.66750 17600.451
9       No      No 1161.05785 37468.529
10      No      No    0.00000 29275.268
11      No     Yes    0.00000 21871.073
12      No     Yes 1220.58375 13268.562
13      No      No  237.04511 28251.695
14      No      No  606.74234 44994.556
15      No      No 1112.96840 23810.174
16      No      No  286.23256 45042.413
17      No      No    0.00000 50265.312
18      No     Yes  527.54018 17636.540
19      No      No  485.93686 61566.106
20      No      No 1095.07274 26464.631
21      No      No  228.95255 50500.182
22      No      No  954.26179 32457.509
23      No      No 1055.95660 51317.883
24      No      No  641.98439 30466.103
25      No      No  773.21172 34353.314
```



No
Yes
income
balance
0
20000
40000
60000
0
500
1000
1500
2000
2500

Foco: Predicción

- $\mathbf{X} = (X_1, X_2, \ldots, X_p)$: vector de covariables

- Y: respuesta $\longrightarrow Y = \begin{cases} \text{cuantitativa} & \rightsquigarrow & \text{Regresión} \\ \text{cualitativa} & \rightsquigarrow & \text{Clasificación} \end{cases}$

Clasificador

- Información disponible $x \in \mathcal{X}$.
- Posibles *etiquetas*. Caso binario $\mathcal{Y} = \{0, 1\}$
- Posibles *etiquetas*. Caso general $\mathcal{Y} = \{y_1, \ldots, y_k\}$
- Clasificador: Regla que asigna a $x \in \mathcal{X}$ un posible valor $y \in \mathcal{Y}$.

Clasificación: Marco Teórico

- $X \in \mathcal{X}$, $Y \in \mathcal{Y}$.
- (X, Y) vector aleatorio, con puntual p_{XY}.
- Clasificador: Regla (de clasificación) que asigna a cada $x \in \mathcal{X}$ un elemento $y \in \mathcal{Y}$

$$\text{Clasificador} \quad g : \mathcal{X} \to \mathcal{Y}$$

- Error de Clasificación Medio (verdadero - poblacional) del clasificador g

$$L(g) = \mathbb{P}(g(X) \neq Y)$$

- Objetivo (teórico): Encontrar g que minimice el error de clasificación medio

$$g^{op}$$

En la práctica, ¿cómo hacemos?

(X, Y) vector aleatorio

Observamos los datos $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

En la práctica, ¿cómo hacemos?

depredadosclasificadas: Bloc de notas

Archivo Edición Formato Ver Ayuda

X	Y
2	0
4	1
5	1
1	0
2	0
3	0
3	0
3	0
2	0
4	0
3	0
4	0
1	0
4	0
3	0
0	0
3	0
2	0
1	0
3	0
6	1
3	0
3	0
2	0

- Datos:
$(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

En la práctica, ¿cómo hacemos?

- $X \in \mathcal{X}$, $Y \in \mathcal{Y}$.
- Clasificador: $g : \mathcal{X} \to \mathcal{Y}$

Datos: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

- **Error de Clasificación Empírico** del clasificador g: proporción de pares mal clasificados según g.

en matemática: $$\widehat{L}(g) = \frac{1}{n} \sum_{i=1}^{n} I_{\{g(x_i) \neq y_i\}}$$

En la práctica, ¿cómo hacemos?

- $X \in \mathcal{X}$, $Y \in \mathcal{Y}$.
- Clasificador: $g : \mathcal{X} \to \mathcal{Y}$

<u>Datos</u>: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

- **Error de Clasificación Empírico** del clasificador g: proporción de pares mal clasificados según g.

en matemática: $$\widehat{L}(g) = \frac{1}{n} \sum_{i=1}^{n} I_{\{g(x_i) \neq y_i\}}$$

en R: mean$(g(x)! = y)$

Clasificación - Ejemplo

Y	0	1
p_Y	0.90	0.10

Clasificación - Ejemplo

Y	0	1
p_Y	0.90	0.10

- $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, (X, Y) vector aleatorio, con puntual p_{XY}.

Y	0	1
$p_{Y\|X=2}$	0.80	0.20

Y	0	1
$p_{Y\|X=3}$	0.30	0.70

¿Cómo clasificaríamos con esta información?

Clasificación - Ejemplo

Y	0	1
p_Y	0.90	0.10

- $X \in \mathcal{X}$, $Y \in \mathcal{Y}$, (X, Y) vector aleatorio, con puntual p_{XY}.

Y	0	1
$p_{Y\|X=2}$	0.80	0.20

Y	0	1
$p_{Y\|X=3}$	0.30	0.70

¿Cómo clasificaríamos con esta información?

- Clasificador g:

x	.	.	.	2	3	.	.	.	.
clasificador	.	.	.	0	1	.	.	.	.

g^{op} Optimo: Regla de Bayes - Caso binario

$$g^{op}(x) \; = \; \begin{cases} 1 & \text{si} \quad \mathbb{P}(Y=1 \mid X=x) \geq \mathbb{P}(Y=0 \mid X=x) \\ 0 & \text{si} \quad \mathbb{P}(Y=0 \mid X=x) > \mathbb{P}(Y=1 \mid X=x) \end{cases}$$

g^{op} Optimo: Regla de Bayes - Caso binario

$$g^{op}(x) \;=\; \begin{cases} 1 & \text{si} \quad \mathbb{P}(Y=1 \mid X=x) \geq \mathbb{P}(Y=0 \mid X=x) \\ 0 & \text{si} \quad \mathbb{P}(Y=0 \mid X=x) > \mathbb{P}(Y=1 \mid X=x) \end{cases}$$

Teorema: $\mathbb{P}\left(g^{op}(X) \neq Y\right) \leq \mathbb{P}\left(g(X) \neq Y\right)$, para todo g.

g^{op} Optimo: Regla de Bayes - Caso binario

$$
\begin{aligned}
L(g) &= \mathbb{P}\left(g(X) \neq Y\right) = E_{XY}[\mathcal{I}_{(g(X)\neq Y)}] \\
&= E_X E_{Y|X}\left[\mathcal{I}_{(g(X)\neq Y)}\right] \\
&= E_X\left[\mathcal{I}_{(g(X)=0)}\mathbb{P}(Y=1 \mid X) + \mathcal{I}_{(g(X)=1)}\mathbb{P}(Y=0 \mid X)\right].
\end{aligned}
$$

Luego, para minimizar $L(g)$ es suficiente con minimizar para todo x

$$
\mathcal{I}_{(g(x)=0)}\mathbb{P}(Y=1 \mid X=x) + \mathcal{I}_{(g(x)=1)}\mathbb{P}(Y=0 \mid X=x)
$$

Teniendo en cuenta que las dos indicadoras son mutuamente excluyentes, alcanza con tomar

$$
g^{op}(x) = \begin{cases} 1 & \text{si} \quad \mathbb{P}(Y=1 \mid X=x) \geq \mathbb{P}(Y=0 \mid X=x) \\ 0 & \text{si} \quad \text{caso contrario} \end{cases}
$$

g^{op} Optimo: Regla de Bayes - Caso binario

$$g^{op}(x) = \begin{cases} 1 & \text{si} \quad \mathbb{P}(Y = 1 \mid X = x) \geq \mathbb{P}(Y = 0 \mid X = x) \\ 0 & \text{si} \quad \mathbb{P}(Y = 0 \mid X = x) > \mathbb{P}(Y = 1 \mid X = x) \end{cases}$$

g^{op} Optimo: Regla de Bayes - Caso binario

$$g^{op}(x) \;=\; \begin{cases} 1 \quad \text{si} \quad \mathbb{P}(Y=1 \mid X=x) \geq \mathbb{P}(Y=0 \mid X=x) \\ 0 \quad \text{si} \quad \mathbb{P}(Y=0 \mid X=x) > \mathbb{P}(Y=1 \mid X=x) \end{cases}$$

Notemos que

$$P(Y=1 \mid X=x) \geq \mathbb{P}(Y=0 \mid X=x) \Leftrightarrow \mathbb{P}(Y=1 \mid X=x) \geq 1/2$$

$$g^{op}(x) = \begin{cases} 1 \quad \text{si } \mathbb{P}(Y=1 \mid X=x) \geq 1/2 \\ 0 \quad \text{c. c.} \end{cases}$$

g^{op} Optimo: Regla de Bayes - Caso binario

$$g^{op}(x) \; = \; \begin{cases} 1 & \text{si} \quad \mathbb{P}(Y=1 \mid X=x) \geq \mathbb{P}(Y=0 \mid X=x) \\ 0 & \text{si} \quad \text{c. c.} \end{cases}$$

De Bayes, tenemos que, si X es discreta, en un x de masa positiva

$$\mathbb{P}\left(Y=y \mid X=x\right) = \frac{\mathbb{P}\left(X=x \mid Y=y\right)\mathbb{P}(Y=y)}{\mathbb{P}(X=x)}$$

$$g^{op}(x) = \begin{cases} 1 & \text{si } \mathbb{P}\left(X=x \mid Y=1\right)\mathbb{P}(Y=1) \geq \mathbb{P}\left(X=x \mid Y=0\right)\mathbb{P}(Y=0) \\ 0 & \text{si } \mathbb{P}\left(X=x \mid Y=0\right)\mathbb{P}(Y=0) > \mathbb{P}\left(X=x \mid Y=1\right)\mathbb{P}(Y=1) \end{cases}$$

g^{op} Optimo: Regla de Bayes - Caso binario

En el caso continuo, si $X|Y = 0 \sim f_0$ y $X|Y = 1 \sim f_1$, resulta

$$g^{op}(x) = \begin{cases} 1 & \text{si } f_1(x)\mathbb{P}(Y = 1) \geq f_0(x)\mathbb{P}(Y = 0) \\ 0 & \text{si c. c.} \end{cases}$$

Estimadores Plug–in

Otra vez en el jardín de los senderos que se bifurcan...

Estimación g^{op} Optimo: Métodos discriminativos

$$g^{op}(x) \;=\; \begin{cases} 1 & \text{si} \quad \mathbb{P}(Y=1 \mid X=x) \geq 1/2 \\ 0 & \text{si} \quad \text{c. c.} \end{cases}$$

Aquí se estima

- $\mathbb{P}(Y=1 \mid X=x)$.

Estimación g^{op} Optimo: Métodos generativos

$$g^{op}(x) = \begin{cases} 1 & \text{si } f_1(x)\mathbb{P}(Y=1) \geq f_0(x)\mathbb{P}(Y=0) \\ 0 & \text{si c. c.} \end{cases}$$

Se estiman

- f_0 y f_1
- $p_1 = \mathbb{P}(Y=1)$ o $p_0 = \mathbb{P}(Y=0)$.

¿Cómo computamos la regla de clasificación?

Estimación de g^{op} en el **caso discriminativo**.

- X v.a. discreta
- Posibles *etiquetas*. Caso binario $\mathcal{Y} = \{0, 1\}$
- Clasificador: Regla que asigna a cada $x \in \mathcal{X}$ un elemento $y \in \mathcal{Y}$
- g^{op} Optimo: Regla de Bayes - Caso binario

$$g^{op}(x) \ = \begin{cases} 1 & \text{si} \quad \mathbb{P}(Y = 1 \mid X = x) \geq 1/2, \\ 0 & \text{si} \quad c.\ c. \end{cases}$$

¿Cómo computamos la regla de clasificación?

Estimación de g^{op} en el **caso discriminativo**.

- X v.a. discreta
- Posibles *etiquetas*. Caso binario $\mathcal{Y} = \{0, 1\}$
- Clasificador: Regla que asigna a cada $x \in \mathcal{X}$ un elemento $y \in \mathcal{Y}$
- g^{op} Optimo: Regla de Bayes - Caso binario

$$g^{op}(x) \; = \begin{cases} 1 & \text{si} \quad \mathbb{P}(Y = 1 \mid X = x) \geq 1/2, \\ 0 & \text{si} \quad c.\ c. \end{cases}$$

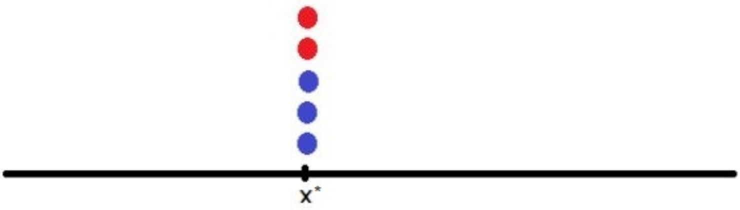
Datos: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

Usaremos el método *Plug–in*, pero....

¿cómo podríamos estimar $\mathbb{P}(Y = 1 \mid X = x)$ o $\mathbb{P}(Y = 0 \mid X = x)$?

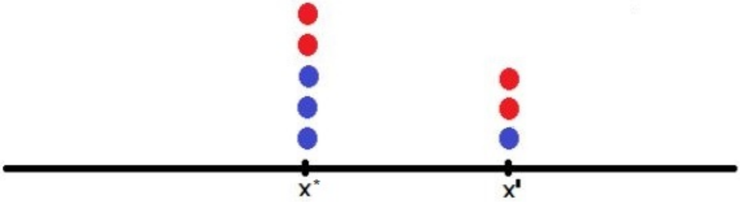
Un método democrático: regla de la mayoría

X discreta



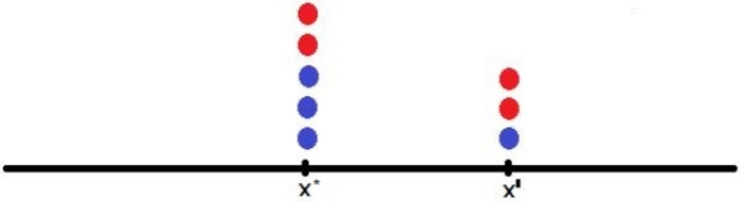
Un método democrático: regla de la mayoría

X discreta



Un método democrático: regla de la mayoría

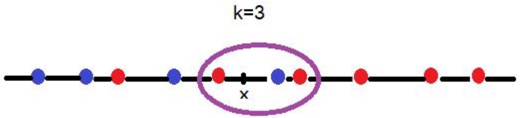
X discreta



$$\frac{\sum_{i=1}^{n} y_i I_{\{x_i=x\}}}{\sum_{i=1}^{n} I_{\{x_i=x\}}}$$

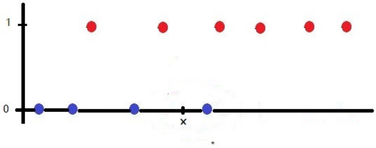
Un método democrático: regla de la mayoría

X continua



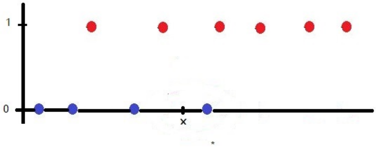
Un método democrático: regla de la mayoría

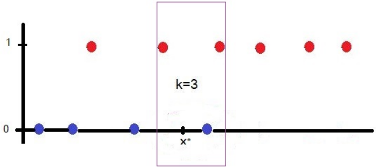
X continua



Un método democrático: regla de la mayoría

X continua





$k-$Vecinos más cercanos (kNN: k-nearest neighbours)

El método de $k-$Vecinos más cercanos es uno de los métodos existentes para estimar la distribución condicional de Y dado X y para después clasificar una observación en la clase con la mayor probabilidad estimada.

- Elegimos k un entero positivo y un punto x para clasificar.
- El clasificador kNN identifica el conjunto de los k puntos más cercanos a x. Sea N_x dicho conjunto.
- Estima a $\mathbb{P}(Y = 1 \mid X = x)$ por la fracción de puntos en N_x cuya etiqueta es igual a 1:

$$\widehat{\mathbb{P}}(Y = 1 \mid X = x) = \frac{1}{k} \sum_{i \in N_x} \mathcal{I}_{\{y_i = 1\}}$$

- Análogamente estimamos $\mathbb{P}(Y = 0 \mid X = x)$

¿Cómo elegimos el parámetro k?

Otra forma: estimación por núcleos

Otra manera de estimar a $\mathbb{P}(Y = 1 \mid X = x)$ podría ser considerar un entorno $(x - h, x + h)$ y repetir el procedimiento anterior.

- Elegimos $h > 0$ y un punto x para clasificar.
- El clasificador identifica en el intervalo $(x - h, x + h)$ los puntos con etiqueta 1 y 0
- Estima a $\mathbb{P}(Y = 1 \mid X = x)$ por la fracción de puntos en $(x - h, x + h)$ cuya etiqueta es igual a 1:

$$\widehat{\mathbb{P}}(Y = 1 \mid X = x) = \frac{\displaystyle\sum_{i=1}^{n} y_i \; I_{(x-h,x+h)}(x_i)}{\displaystyle\sum_{i=1}^{n} I_{(x-h,x+h)}(x_i)}$$

¿Cómo elegimos el parámetro h?

Otra forma: estimación por núcleos

Otra manera de estimar a $P(Y = 1 \mid X = x)$ podría ser considerar un entorno $(x - h, x + h)$ y repetir el procedimiento anterior.

- Elegimos $h > 0$ y un punto x para clasificar.
- El clasificador identifica en el intervalo $(x - h, x + h)$ los puntos con etiqueta 1 y 0
- Estima a $P(Y = 1 \mid X = x)$ usando un núcleo K

$$\widehat{\mathbb{P}}(Y = 1 \mid X = x) = \frac{\displaystyle\sum_{i=1}^{n} y_i \, K\left(\frac{x - x_i}{h}\right)}{\displaystyle\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right)}$$

¿Cómo elegimos el parámetro h?

Estimación g^{op} Optimo: Métodos generativos

$$g^{op}(x) = \begin{cases} 1 & \text{si } f_1(x)\,\mathbb{P}(Y=1) \geq f_0(x)\,\mathbb{P}(Y=0) \\ 0 & \text{si c. c.} \end{cases}$$

Se estiman

- f_0 y f_1
- $p_1 = \mathbb{P}(Y=1)$ o $p_0 = \mathbb{P}(Y=0)$.

Regla plug–in de Bayes

- (X, Y) vector aleatorio
- Datos: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

- Estimaciones de p_1 y p_0: $\widehat{p}_1 = \frac{n_1}{n}$ y $\widehat{p}_0 = \frac{n_0}{n}$, donde $n_j = \#\{y_i = j\}, j = 0, 1.$

Regla plug–in de Bayes

- (X, Y) vector aleatorio
- Datos: $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$

- Estimaciones de p_1 y p_0: $\widehat{p}_1 = \frac{n_1}{n}$ y $\widehat{p}_0 = \frac{n_0}{n}$, donde $n_j = \#\{y_i = j\}, j = 0, 1$.

- Estimamos $f_1(x)$ mediante $\widehat{f}_1(x)$.

- Estimamos $f_0(x)$ mediante $\widehat{f}_0(x)$.

- Hacemos un plug–in en la regla $g^{op}(x)$:

$$\widehat{g}(x) = \begin{cases} 1 & \text{si } \widehat{f}_1(x)\,\widehat{p}_1 \geq \widehat{f}_0(x)\,\widehat{p}_0, \\ 0 & \text{si } c.c. \end{cases}$$

¿Cómo obtenemos las densidades estimadas $\widehat{f_1}(x)$ y $\widehat{f_0}(x)$?

¿Cómo obtenemos las densidades estimadas $\widehat{f}_1(x)$ y $\widehat{f}_0(x)$?

Opción 1: Enfoque No Paramétrico

- Estimamos $f_1(x)$ no paramétricamente con $\widehat{f}_{1,h_1}(x)$
- Estimamos $f_0(x)$ no paramétricamente con $\widehat{f}_{0,h_0}(x)$
- Hacemos un plug–in en la regla $g^{op}(x)$:

$$\widehat{g}_{h_0,h_1}(x) = \begin{cases} 1 & \text{si } \widehat{f}_{1,h_1}(x)\,\widehat{p}_1 \geq \widehat{f}_{0,h_0}(x)\,\widehat{p}_0, \\ 0 & \text{si } c.c. \end{cases}$$

- $h_0, h_1 =?$

¿Cómo elegimos k, h o (h_0, h_1)?

En cada caso, tenemos una $\widehat{g}_t$, ¿cómo elegimos el parámetro o vector de parámetros t?

¿Cómo elegimos k, h o (h_0, h_1)?

En cada caso, tenemos una $\widehat{g}_t$, ¿cómo elegimos el parámetro o vector de parámetros t?

Dividiendo los datos....

Test Error

Podemos repetir el mismo razonamiento que hicimos para regresión:

- $\widehat{m}_{\mathcal{M}}$: $\widehat{m}$ estimador construido a partir de $\mathcal{M}$

- Test Error

$$\text{Error}_{\mathcal{M}} = \mathbb{E}\left[(Y_o - \widehat{m}_{\mathcal{M}}(X_o))^2 \mid \mathcal{M}\right]$$

Test Error

Podemos repetir el mismo razonamiento que hicimos para regresión:

- $\widehat{g}_{\mathcal{M}}$: $\widehat{g}$ regla construida a partir de $\mathcal{M}$

- Test Error

$$\text{Error}_{\mathcal{M}} = \mathbb{E}\left[(Y_o - \widehat{g}_{\mathcal{M}}(X_o))^2 \mid \mathcal{M}\right]$$

Test Error

Podemos repetir el mismo razonamiento que hicimos para regresión:

- $\widehat{g}_{\mathcal{M}}$: $\widehat{g}$ regla construida a partir de $\mathcal{M}$

- Test Error

$$\text{Error}_{\mathcal{M}} = \mathbb{E}\left[(Y_o - \widehat{g}_{\mathcal{M}}(X_o))^2 \mid \mathcal{M}\right]$$

- equivalente a

$$\mathbb{P}\left(\widehat{g}_{\mathcal{M}}(X_o) \neq Y_o \mid \mathcal{M}\right)$$

Muchos Datos – Menos Datos

Consideramos las mismas estrategias que ya mencionamos para estimación de la densidad y regresión.

Validación Cruzada: Leave one out - Fórmulas

t: parámetro de tuneo

$$\mathsf{CV}(t) = \widehat{L(t)} = \frac{1}{n}\sum_{i=1}^{n} L_i(t)$$

- Regresión:

$$L_i(t) = \{Y_i - \widehat{r}_t^{(-i)}(X_i)\}^2$$

$$t_{\mathsf{opt}} = \arg\min_t \widehat{L(t)}$$

- Clasificación:

$$L_i(t) = (Y_i - \widehat{g}_t^{(-i)}(X_i))^2 = \mathcal{I}_{\{\widehat{g}_t^{(-i)}(X_i) \neq Y_i\}}$$

Validación Cruzada: Leave one out - Fórmulas

t: parámetro de tuneo

$$\mathsf{CV}(t) = \widehat{L(t)} = \frac{1}{n}\sum_{i=1}^{n} L_i(t)$$

- Clasificación:

$$L_i(t) = \mathcal{I}_{\{\widehat{g}_t^{(-i)}(X_i) \neq Y_i\}}$$

Validación Cruzada: K folders - Fórmulas

t: parámetro de tuneo

$$\widehat{L(t)} = \frac{1}{K}\sum_{k=1}^{K} L_k(t)$$

- Regresión: para el $k-$ésimo fold

$$L_k(t) = \frac{1}{|\mathcal{T}_k^c|}\sum_{j\in\mathcal{T}_k^c}(Y_j - \widehat{r}_{t,\mathcal{T}_k}(X_i))^2$$

$$t_{\text{opt}} = \arg\min \widehat{L(t)}$$

- Clasificación:

$$L_k(t) = \frac{1}{|\mathcal{T}_k^c|}\sum_{j\in\mathcal{T}_k^c}\mathcal{I}_{\{\widehat{g}_{t,\mathcal{T}_k}(X_j)\neq Y_j\}}$$

¿Cómo obtenemos las densidades estimadas $\widehat{f_1}(x)$ y $\widehat{f_0}(x)$?

Opción 2: Enfoque Paramétrico

Por ejemplo, si fuera razonable asumir que la distribución de la altura X en **cada género** es normal, podríamos estimar los parámetros de cada normal con los datos de altura dentro de cada géneroy hacer un plug–in en la regla de Bayes:

¿Cómo obtenemos las densidades estimadas $\widehat{f}_1(x)$ y $\widehat{f}_0(x)$?

Opción 2: Enfoque Paramétrico
Por ejemplo, si fuera razonable asumir que la distribución de la altura X en **cada género** es normal, podríamos estimar los parámetros de cada normal con los datos de altura dentro de cada géneroy hacer un plug–in en la regla de Bayes:

- Obtenemos $\widehat{\mu}_1$ y $\widehat{\sigma}_1^2$ a partir del género 1: $\Rightarrow \widehat{f}_1(x) = f(x, \widehat{\mu}_1, \widehat{\sigma}_1^2)$.

- Obtenemos $\widehat{\mu}_0$ y $\widehat{\sigma}_0^2$ a partir del género 0: $\Rightarrow \widehat{f}_0(x) = f(x, \widehat{\mu}_0, \widehat{\sigma}_0^2)$.

- Hacemos un plug–in en la regla $g^{op}(x)$:

$$\widehat{g}(x) = \begin{cases} 1 & \text{si } f(x, \widehat{\mu}_1, \widehat{\sigma}_1^2)\, \widehat{p}_1 \geq f(x, \widehat{\mu}_0, \widehat{\sigma}_0^2)\, \widehat{p}_0, \\ 0 & \text{si } c.c. \end{cases}$$

Algunas otras cuestiones en Regresión y Clasificación

- ¿Qué pasa en mayor dimensión? Maldición de la Dimensión
- Clasificación: alternativa inocente Bayes Naive
- Métodos Paramétricos

X de mayor dimensión: ingredientes

Supongamos que disponemos de un conjunto de elementos que pueden venir de dos poblaciones distintas.

En cada elemento se ha observado una X variable aleatoria de dimensión p. (por ejemplo registramos altura, ancho, peso, etc.

Se desea clasificar un nuevo elemento, con valores de las variables conocidas.

X de mayor dimensión

Supongamos que para predecir el género del individuo tenemos en cuenta $p = 4$ medidas: altura, peso, perímetro de caderas y de cintura.
$(Y, \mathbf{X})$: donde $\mathbf{X} = (X_1, X_2, \ldots, X_p)^t$, es decir, se registran p covariables.

Igual que antes el clasificador óptimo

$$g^{op}(\mathbf{x}) = \begin{cases} 1 & \text{si } f_1(\mathbf{x})p_1 \geq f_0(\mathbf{x})p_0, \\ 0 & \text{si } c.c. \end{cases}$$

donde $f_1(\mathbf{x})$ y $f_0(\mathbf{x})$ son densidades en dimensión p.

X de mayor dimensión

Opción 1: Enfoque No Paramétrico

Si no sabemos nada sobre las dos densidades, una alternativa es usar el estimador no parámetrico de la densidad al igual que antes Ahora tenemos una dendidad sobre $\mathbb{R}^p$: $f(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^p$.
Dada la

$$\widehat{f}(\mathbf{x}) = \frac{1}{nh^p} \sum_{i=1}^{n} K\left(\frac{\mathbf{x} - \mathbf{X}_i}{h}\right)$$

ahora $\mathbf{x} - \mathbf{X}_i$ es de dimensión p.

Bayes Naive

Una forma de eludir el problema de la maldición de la dimensión es construir un estimador bajo la hipótesis de que las p covariables son independientes.

Bayes Naive

Una forma de eludir el problema de la maldición de la dimensión es construir un estimador bajo la hipótesis de que las p covariables son independientes. Luego, pensamos que

$$d(\mathbf{x}) = d_1(x_1)\, d_2(x_2) \ldots d_p(x_p)$$

y estimamos:

$$\widehat{d}(\mathbf{x}) = \widehat{d}_1(x_1)\, \widehat{d}_2(x_2) \ldots \widehat{d}_p(x_p)$$

Bayes Naive

Una forma de eludir el problema de la maldición de la dimensión es construir un estimador bajo la hipótesis de que las p covariables son independientes. Luego, pensamos que

$$d(\mathbf{x}) = d_1(x_1)\, d_2(x_2) \ldots d_p(x_p)$$

y estimamos:

$$\widehat{d}(\mathbf{x}) = \widehat{d}_1(x_1)\, \widehat{d}_2(x_2) \ldots \widehat{d}_p(x_p)$$

Es decir, en lugar de estimar la densidad multivariada de $\mathbf{X} =$(altura, peso, perímetro de caderas, perímetro de cintura). Estimamos de forma separada la densidad de $X_1 =$ altura, $X_2 =$ peso, $X_3 =$ perímetro de caderas y $X_4 =$ perímetro de cintura.

Bayes Naive

Una forma de eludir el problema de la maldición de la dimensión es construir un estimador bajo la hipótesis de que las p covariables son independientes. Luego, pensamos que

$$d(\mathbf{x}) = d_1(x_1)\, d_2(x_2) \dots d_p(x_p)$$

y estimamos:

$$\widehat{d}(\mathbf{x}) = \widehat{d}_1(x_1)\, \widehat{d}_2(x_2) \dots \widehat{d}_p(x_p)$$

Es decir, en lugar de estimar la densidad multivariada de $\mathbf{X} =$(altura, peso, perímetro de caderas, perímetro de cintura). Estimamos de forma separada la densidad de $X_1 =$ altura, $X_2 =$ peso, $X_3 =$ perímetro de caderas y $X_4 =$ perímetro de cintura.

Bayes Naive

Así, obtenemos $\widehat{f_1}(\mathbf{x})$ y $\widehat{f_0}(\mathbf{x})$ y con esto armamos el clasificador.

$$\widehat{g}(\mathbf{x}) = \begin{cases} 1 & \text{si } \widehat{f_1}(\mathbf{x})\widehat{p_1} \geq \widehat{f_0}(\mathbf{x})\widehat{p_0}, \\ 0 & \text{si } c.c. \end{cases}$$