
Clase N°8

Regresión logística y
Clasificación

Motivación

Problemas como:

- mora:

Ingresos, antecedentes previos de clientes vs van a caer en mora o no

Motivación

Problemas como:

- mora:

Ingresos, antecedentes previos de clientes vs van a caer en mora o no

- decisiones médicas de urgencia:

llega paciente a una guarda, luego de pocos estudios se lo lleva a intensivos o no

Motivación

Problemas como:

- mora:

Ingresos, antecedentes previos de clientes vs van a caer en mora o no

- decisiones médicas de urgencia:

llega paciente a una guarda, luego de pocos estudios se lo lleva a intensivos o no

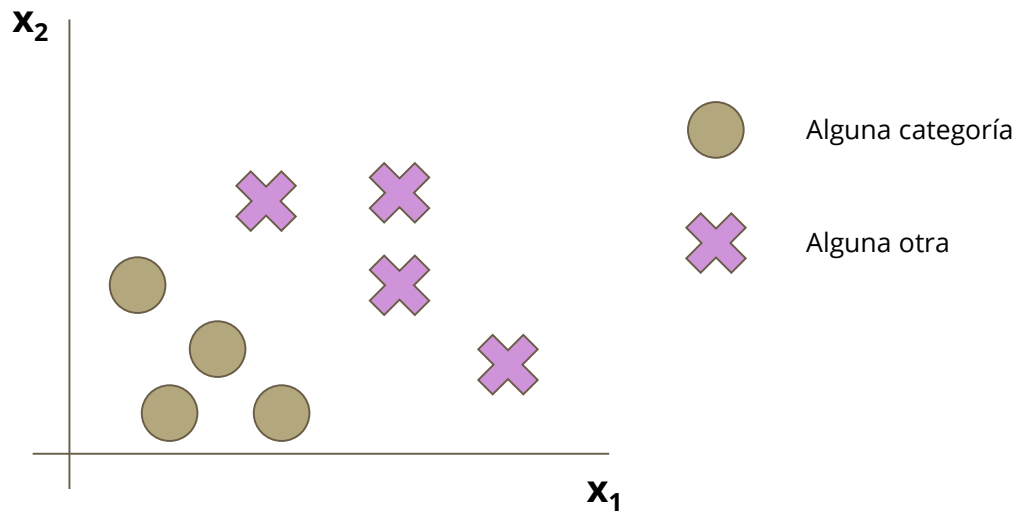
- el clima:

dadas las temperaturas y la humedad, lloverá o no lloverá

Motivación

Tenemos entonces:

- x_1, x_2 características numéricas
- y una etiqueta categórica (binaria en ppio)



Temario

La hipótesis de representación

La frontera de decisión

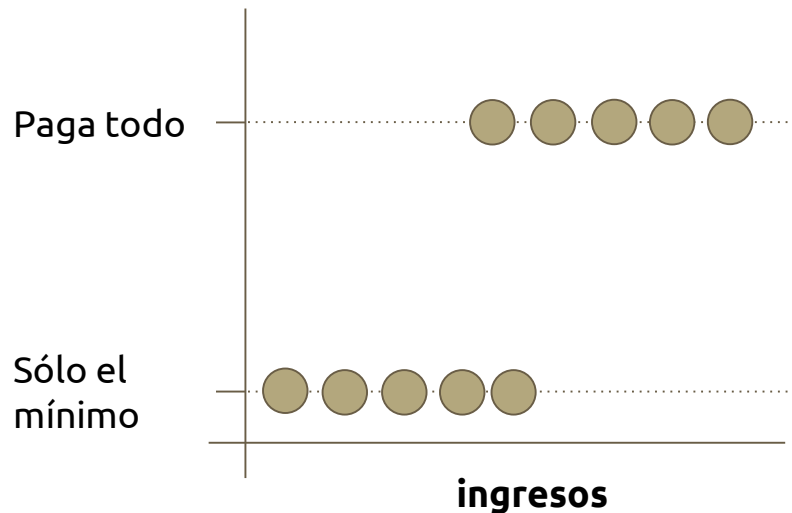
La búsqueda de coeficientes

La regresión múltiple

Un poquito de colab

La hipótesis de representación

Supongamos el problema de clasificación que sigue:

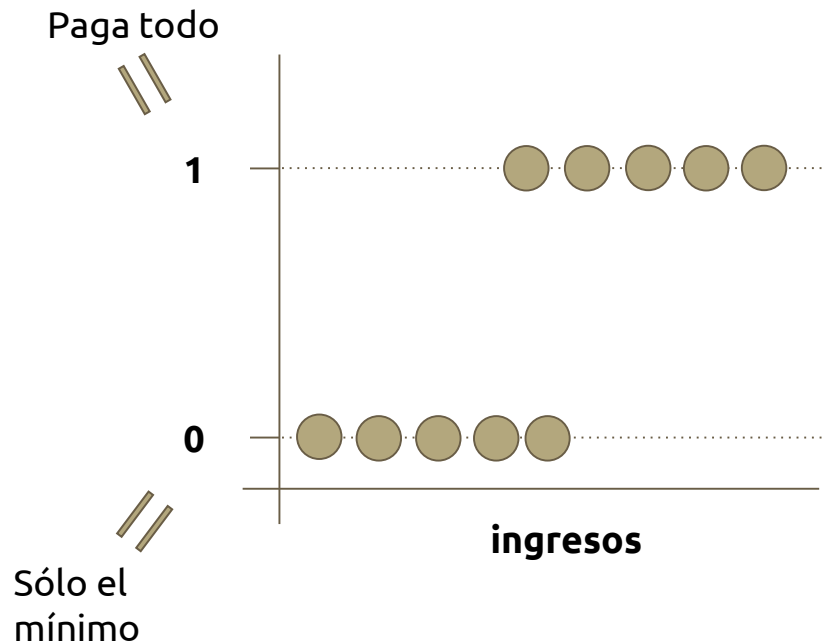


La hipótesis de representación

Supongamos el problema de clasificación que sigue:

- Llevemos las categorías a la representación binaria

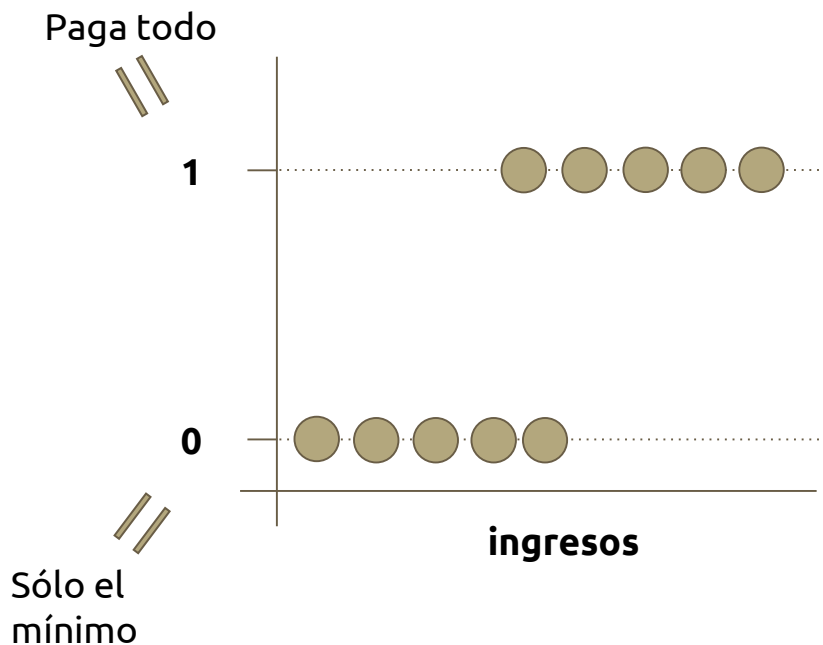
$$y = \begin{cases} 1 & \text{para } y \in \text{Paga todo} \\ 0 & \text{para } y \in \text{sólo el mínimo} \end{cases}$$



La hipótesis de representación

¿Qué pasa si tratamos de ajustar con una lineal?

$$y = f(x, \beta) = \beta_0 + \beta_1 \cdot x$$



La hipótesis de no multicolinealidad

¿Qué pasa si no podemos ajustar con una lineal?

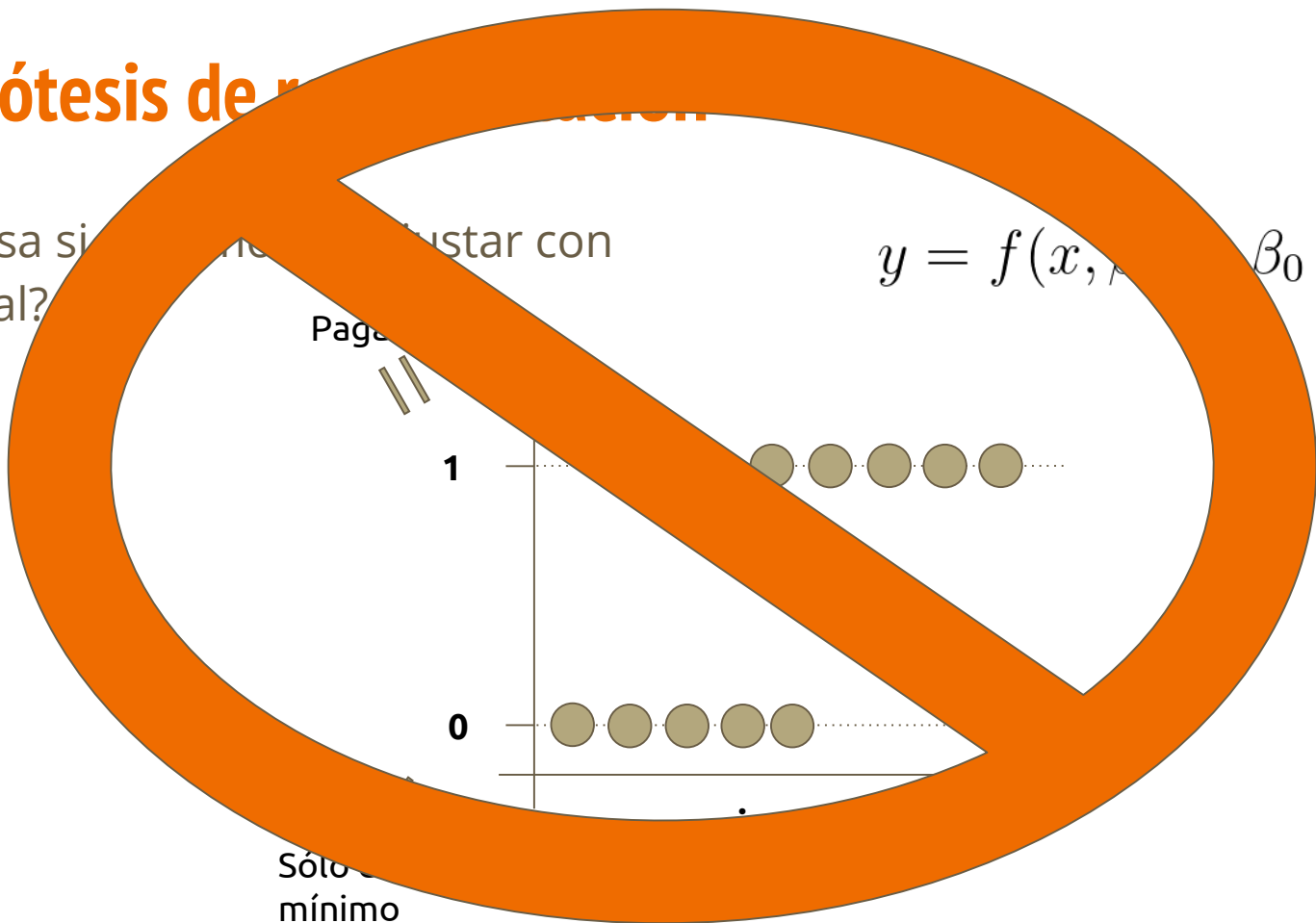
$$y = f(x, \beta_0 + \beta_1 \cdot x)$$

Page

1

0

Sólo el
mínimo



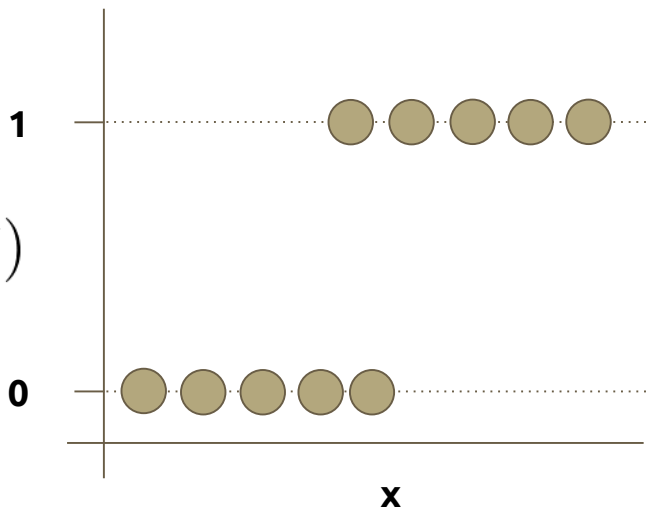
La hipótesis de representación

Pensemos a f como una probabilidad
y usemos una función con imagen
acotada entre cero y uno

función sigmoidea

$$P(y|x, \beta) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot x)}}$$

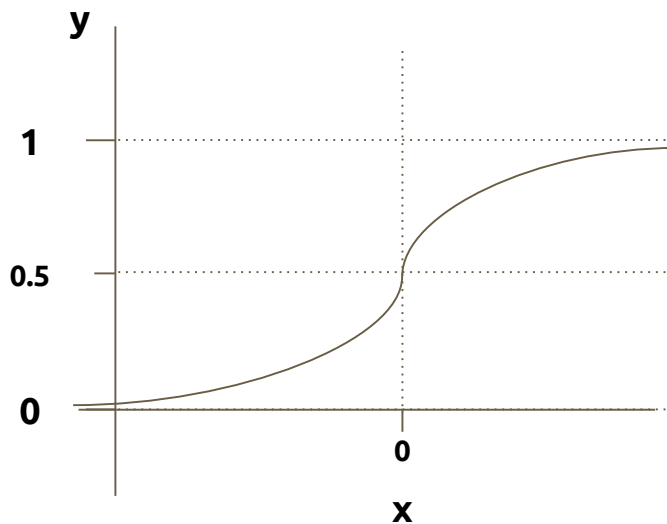
$$y = f(x, \beta) \equiv P(y|x, \beta)$$



La hipótesis de representación

$$P(y|x, \beta) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot x)}}$$

función sigmoidea



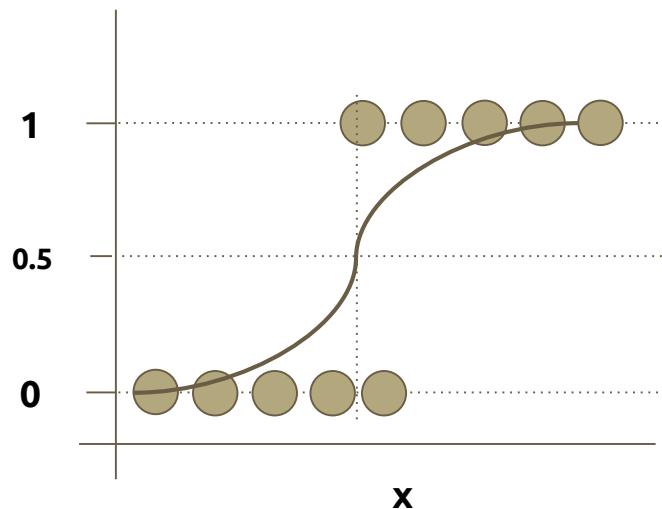
La hipótesis de representación

Entonces, el modelo nos permite encontrar la probabilidad de que, dadas ciertas características y parámetros, el registro corresponda a la categoría **1**

Situamos una probabilidad quiebre en **0.5**

función sigmoidea

$$P(y|x, \beta) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot x)}}$$



La frontera de decisión

$$P(y|x, \beta) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot x)}}$$

Trabajemos entonces con la probabilidad de corte en **0.5**^[1]

$$y = 1, \text{ cuando } P(y|x, \beta) \geq 0.5$$

$$y = 0, \text{ cuando } P(y|x, \beta) < 0.5$$

[1] Si tomáramos otro valor, tendríamos una modificación en β_0 . Al margen, esto es un hiperparámetro del modelo

La frontera de decisión

$$P(y|x, \beta) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot x)}}$$

Trabajemos entonces con la probabilidad de corte en **0.5**^[1]

$$P(y|x, \beta) = 0.5 \qquad \beta_0 + \beta_1 \cdot x_{corte} = 0 \Rightarrow x_{corte} = -\frac{\beta_0}{\beta_1}$$

[1] Si tomáramos otro valor, tendríamos una modificación en β_0 . Al margen, esto es un hiperparámetro del modelo

La frontera de decisión - Más de una característica

Trabajemos entonces con la probabilidad de corte en **0.5**^[1]

$$P(y|\mathbf{x}, \boldsymbol{\beta}) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot x_1^c + \beta_2 \cdot x_2^c)}} = 0.5$$

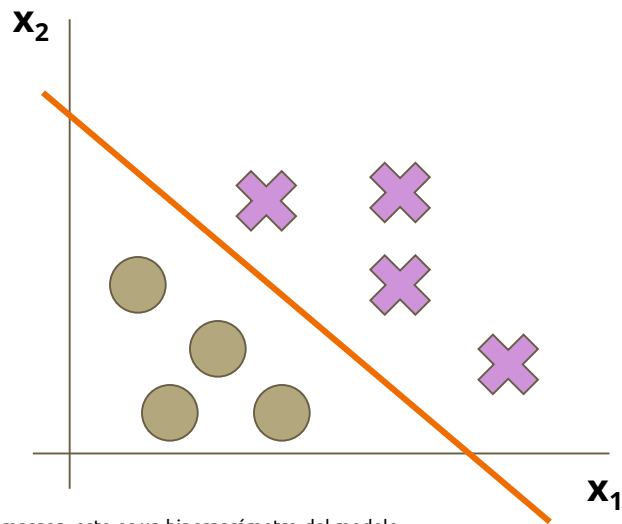
$$\beta_0 + \beta_1 \cdot x_1^c + \beta_2 \cdot x_2^c = 0 \Rightarrow x_2^c = \frac{-(\beta_0 + \beta_1 \cdot x_1^c)}{\beta_2}$$

[1] Si tomáramos otro valor, tendríamos una modificación en β_0 . Al margen, esto es un hiperparámetro del modelo

La frontera de decisión - Más de una característica

Trabajemos entonces con la probabilidad de corte en **0.5**^[1]

$$\beta_0 + \beta_1 \cdot x_1^c + \beta_2 \cdot x_2^c = 0 \Rightarrow x_2^c = -(\beta_0 + \beta_1 \cdot x_1^c)$$



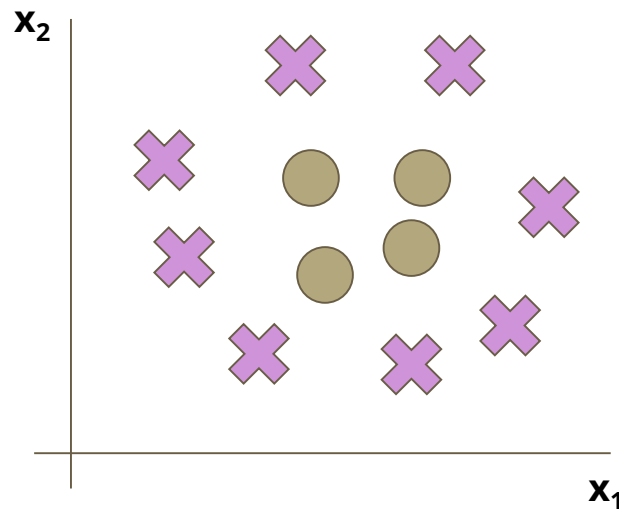
[1] Si tomáramos otro valor, tendríamos una modificación en β_0 . Al margen, esto es un hiperparámetro del modelo

¿Esto termina acá?

No, al igual que con la regresión lineal, podemos proponer polinomios de la pinta:

$$\beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_1 \cdot x_2 + \beta_4 \cdot x_1^2 \dots$$

$$\frac{1}{1 + e^{-(\beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \beta_3 \cdot x_1 \cdot x_2 + \beta_4 \cdot x_1^2 \dots)}}$$



La búsqueda de coeficientes - Máxima verosimilitud

A diferencia de lo que teníamos en el caso lineal, acá no tenemos cuadrados mínimos.

La búsqueda de coeficientes - Máxima verosimilitud

A diferencia de lo que teníamos en el caso lineal, acá no tenemos cuadrados mínimos.

Utilizamos entonces conceptos de **probabilidad**.

La búsqueda de coeficientes - Máxima verosimilitud

A diferencia de lo que teníamos en el caso lineal, acá no tenemos cuadrados mínimos.

Utilizamos entonces conceptos de **probabilidad**.

Buscamos minimizar el producto de probabilidades de que a cada punto se le asigne la categoría correcta

$$l(\boldsymbol{\beta}) = \prod_{i,y=1} P(y|\mathbf{x}, \boldsymbol{\beta}) \prod_{i',y=0} [1 - P(y|\mathbf{x}, \boldsymbol{\beta})]$$

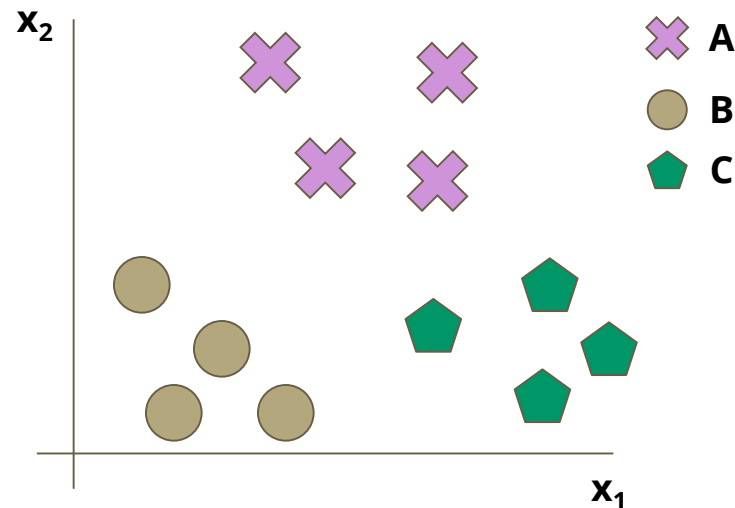
La clasificación múltiple

¿Qué pasa si tenemos más de dos posibles categorías?

La clasificación múltiple

¿Qué pasa si tenemos más de dos posibles categorías?

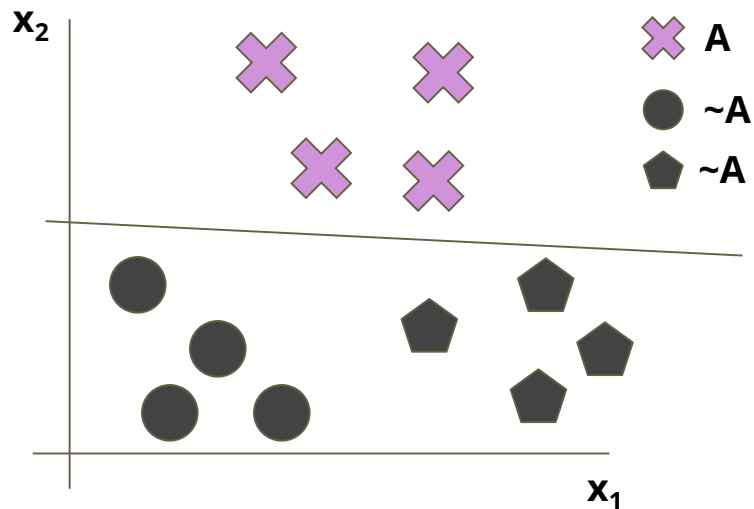
- mora:
paga el total, paga una parte, paga el mínimo
- decisiones médicas de urgencia:
puede esperar, terapia intermedia, terapia intensiva
- el clima:
llueve, nublado, no llueve



La clasificación múltiple - Uno contra todos

Vamos a pensar en una comparación binaria para cada categoría:

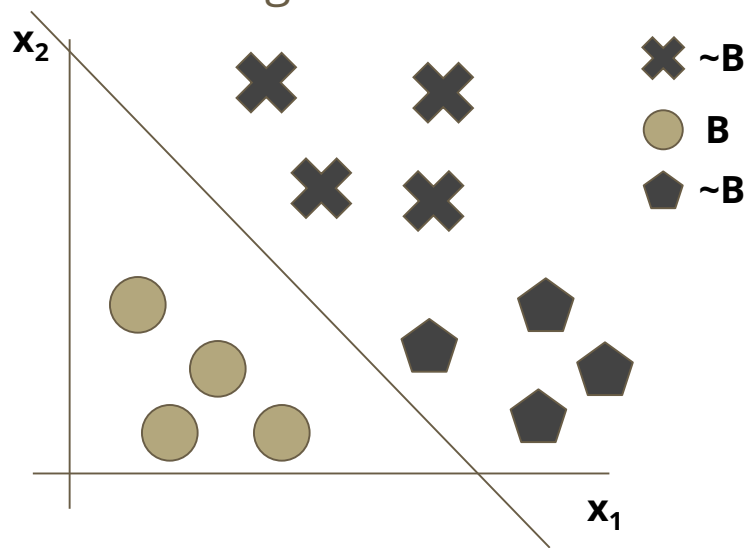
- Que pertenezca a A o no. $P_A(y|\mathbf{x}, \beta_A)$



La clasificación múltiple - Uno contra todos

Vamos a pensar en una comparación binaria para cada categoría:

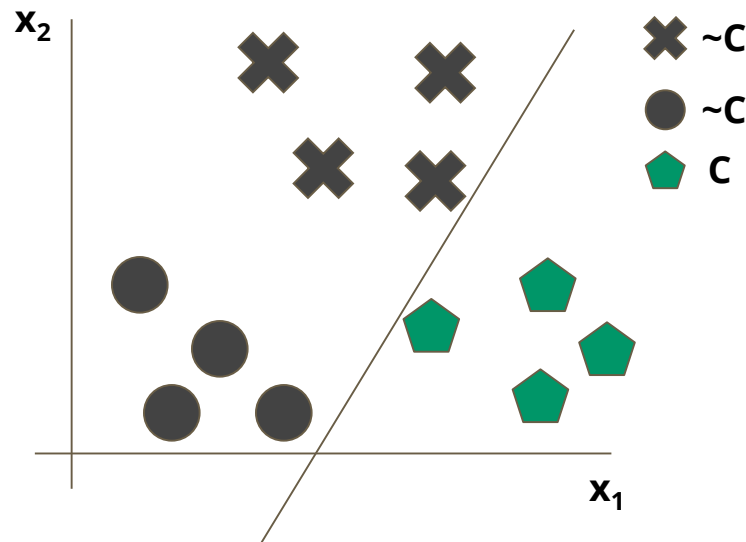
- Que pertenezca a A o no. $P_A(y|\mathbf{x}, \beta_A)$
- Que pertenezca a B o no. $P_B(y|\mathbf{x}, \beta_B)$



La clasificación múltiple - Uno contra todos

Vamos a pensar en una comparación binaria para cada categoría:

- Que pertenezca a A o no. $P_A(y|\mathbf{x}, \beta_A)$
- Que pertenezca a B o no. $P_B(y|\mathbf{x}, \beta_B)$
- Que pertenezca a C o no. $P_C(y|\mathbf{x}, \beta_C)$

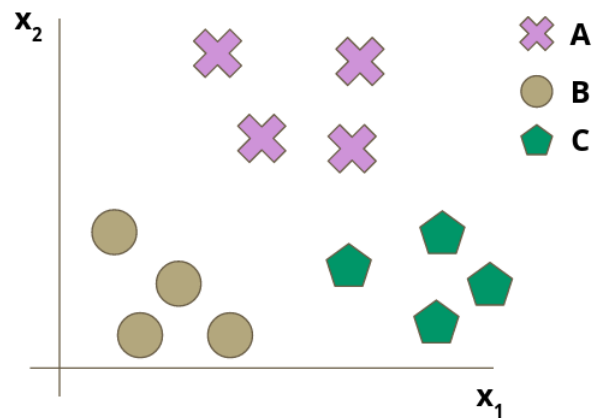


La clasificación múltiple - Uno contra todos

Vamos a pensar en una comparación binaria para cada categoría:

- Que pertenezca a A o no. $P_A(y|\mathbf{x}, \beta_A)$
- Que pertenezca a B o no. $P_B(y|\mathbf{x}, \beta_B)$
- Que pertenezca a C o no. $P_C(y|\mathbf{x}, \beta_C)$

Nos quedamos con la categoría según cuál sea el



$$\max_{j=A,B,C} \{ P_j(y|\mathbf{x}, \beta_j) \}$$

Un poquito de colab