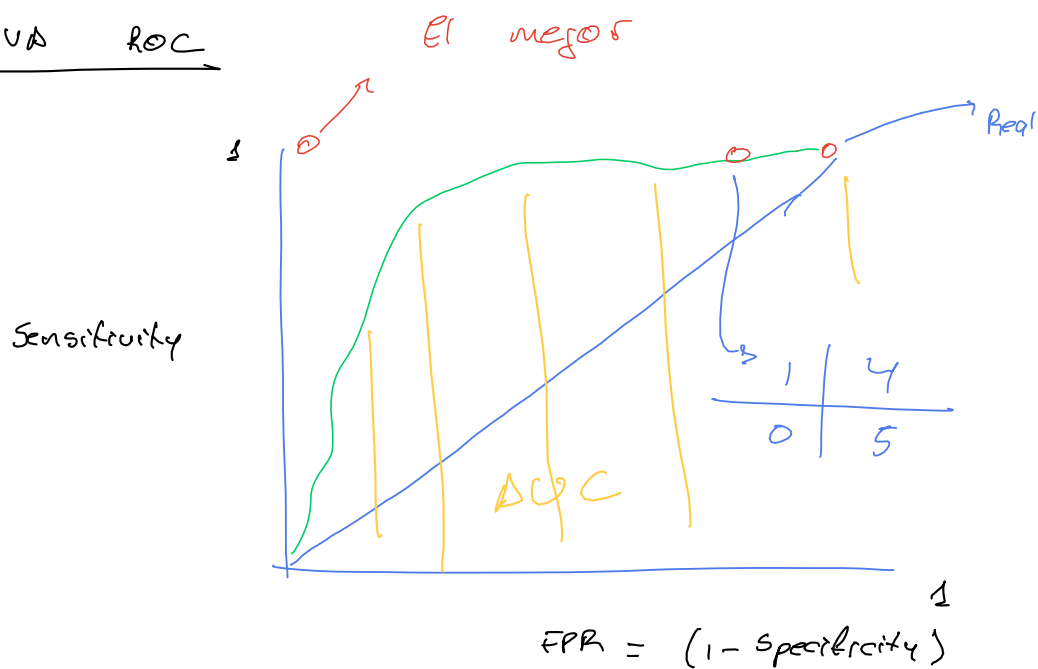


CURVA ROC



Predicted	
0	1
0	5
0	5

$$Sensitivity = \frac{5}{5+0}$$

$$FPR = 1$$

$$Specificity = 0$$

Esta curva define un modelo.

Para un modelo tenemos varias opciones

↓

threshold.

Δ threshold ⇒ Δ sensitivity

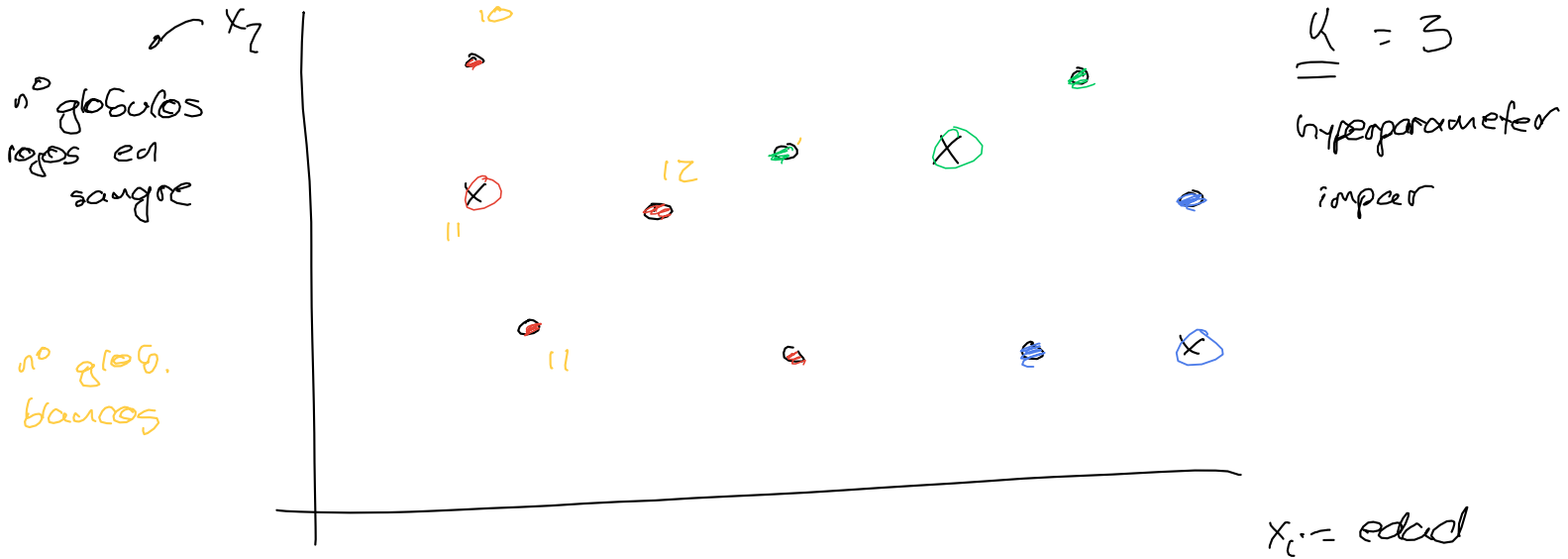
⇒ Δ specificity

AUC → area under curve como max es 1.

Cuanto mayor AUC ⇒ mejor modelo

nearest neighbour.

# KNN - modelo de clasificación (regression)



Cómo medir distancias? - quién es su vecino?

Euclidea:



$$d = x^2 + y^2$$

$$d(A, B) = \sqrt{\sum_i^n (A_i - B_i)^2}$$

esto los puntos en todas las dim.

el de más distancia

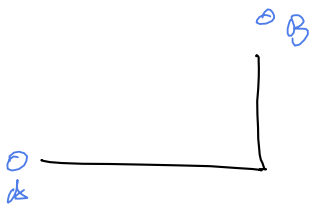
OB 2,3

OA 1,2

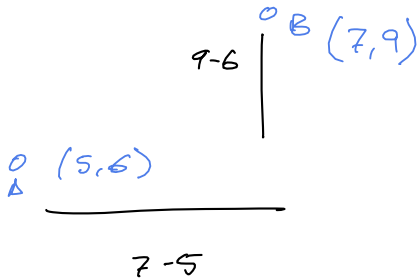
$$\sqrt{(2-1)^2 \oplus (3-2)^2}$$

$\Sigma$  en m dimensiones

## Manhattan



$$d(A, B) = \sqrt{\sum (A_i - B_i)^2}$$



$$d = \sqrt{|9-6| + |7-5|} \quad (B-A)$$

$$d = \sqrt{3 + 2}$$

$$d = \sqrt{|6-9| + |5-7|} \quad (A-B)$$

$$d = \sqrt{3 + 2}$$

Jaccard → se usa para variables no numéricas

	007	Spiderman	Disney	Star Wars	Coco
Berta	1	0	1	1	0
Itaca	0	0	1	1	1
Nico	1	1	0	0	1
Felipe	1	1	1	0	?

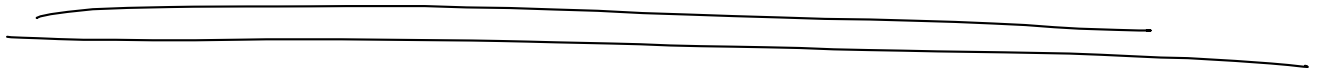
$$d(\text{Nico, Felipe}) = 1 - \frac{\text{cosas comunes}}{\text{cosas de al menos alguno}} = 1 - \left| \frac{X \cap Y}{X \cup Y} \right|$$

$$= 1 - \frac{2}{3} = 0.33 \rightarrow \begin{array}{l} \text{menor} \\ \text{dist} \end{array}$$

$$d(\text{Felipe, Pedro}) = 1 - \frac{2}{4} = 0.5$$

$$d(\text{Felipe, Marco}) = 1 - \frac{1}{4} = 0.75$$

$$\text{Una de } d = 1 \Rightarrow (\text{Felipe, Coco}) = 1$$



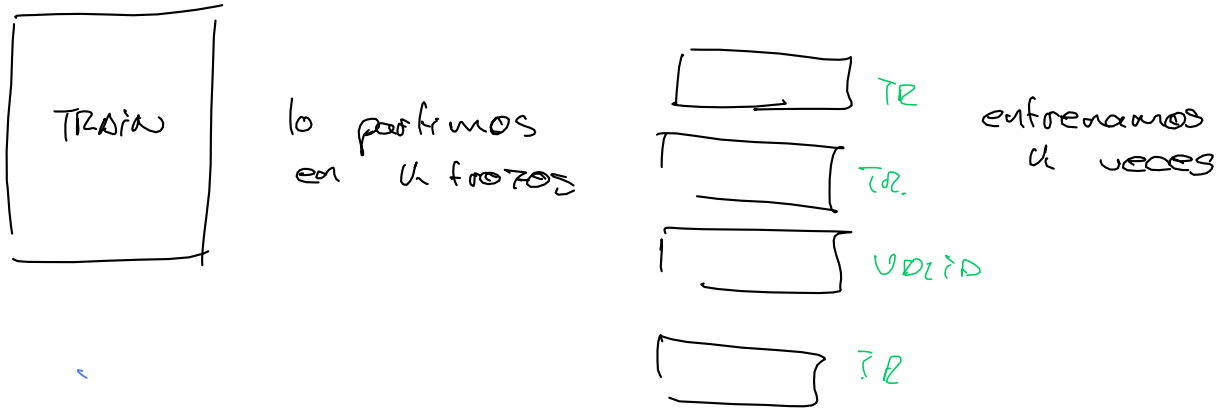
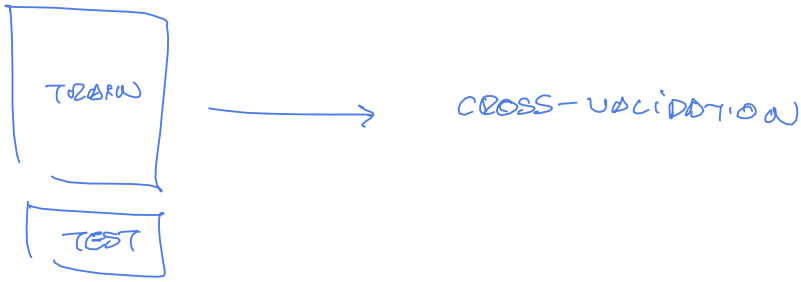
# ¿CÓMO ESCOGER HIPERPARÁMETROS?

threshold

$k$  ( $k, \infty$ )

"nº variables req  $\ln/\log$ "

## 1. CROSS-VALIDATION

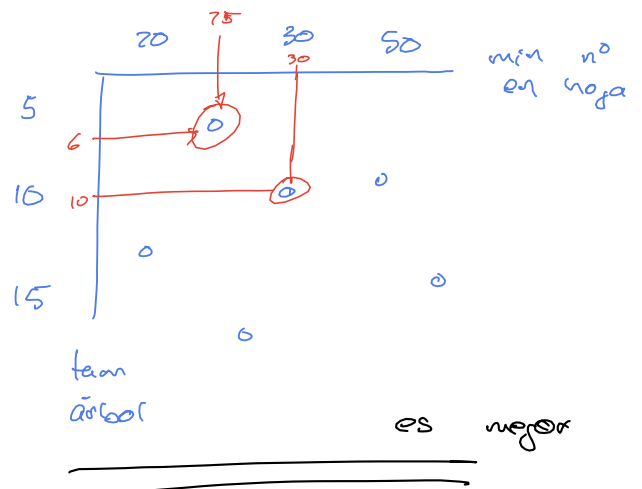


## 2. RANDOM SEARCH & GRID SEARCH

### GRID SEARCH

	$k=3$	$k=5$	$k=7$
Dist Manh.	o	o	o
Dist Euc	o	o	o

### RANDOM SEARCH



o x 6 en cada uno

hemos probado + variables