HUB IQ

Inteligencia Artificial & Machine Learning

Applicaciones en movilidad

Dr. Iván S. Razo Zapata

EngineeringX
Founded by the Royal Academy of Engineering and Lloyd's Register Foundation

Aprendizaje Supervisado

K-vecinos más cercanos (KNN)

Métodos

- Paramétricos
  - Asume una forma para y
  - Regresión lineal
  - Estimación de parámetros
- No paramétricos
  - No asume una forma para y

$$y = f(x) + \epsilon$$

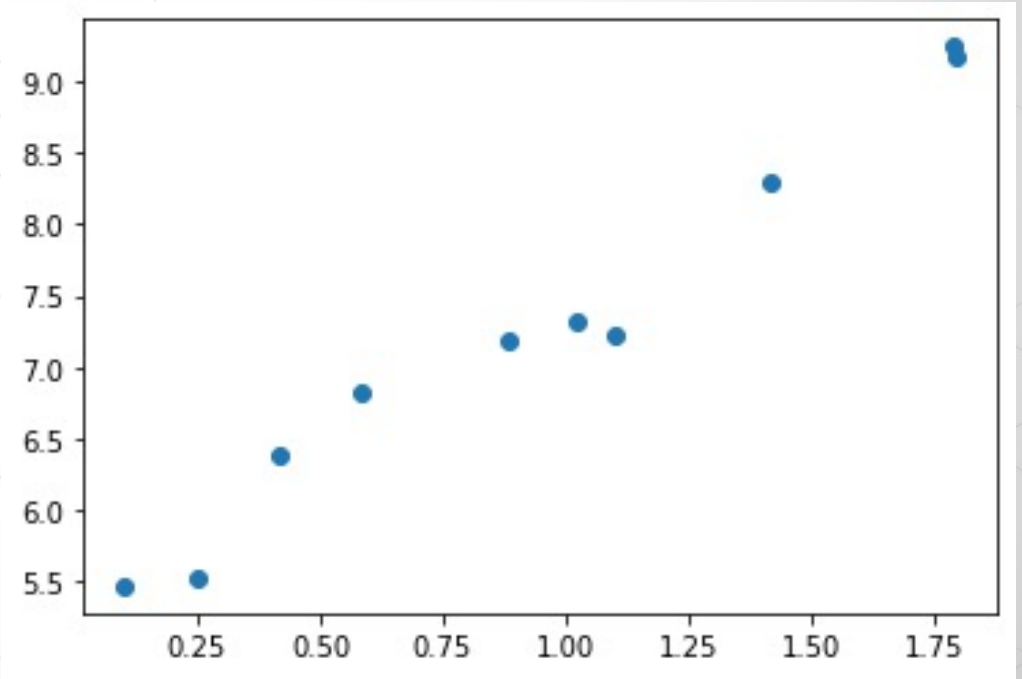
$$y = \theta_0 x_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n + \epsilon$$

$$\theta_0, \theta_1, \theta_2 \ldots \theta_n$$

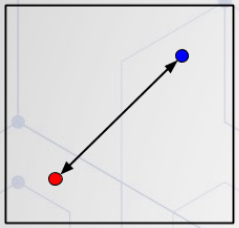
$$y = f(x) + \epsilon$$

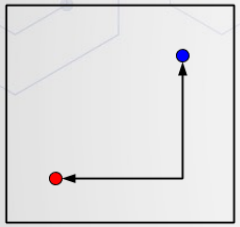
Idea general

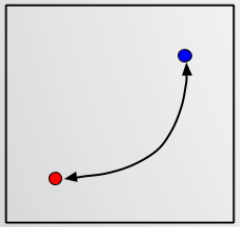
- Los valores de los vecinos más cercanos son una “mejor” aproximación del valor real para una nueva observación
- ¿Vecinos más cercanos?
  - Distancia
- ¿Cuántos vecinos más cercanos?

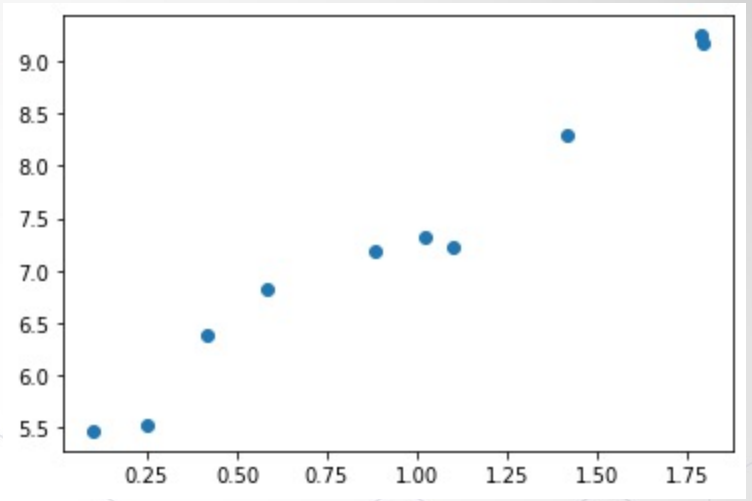


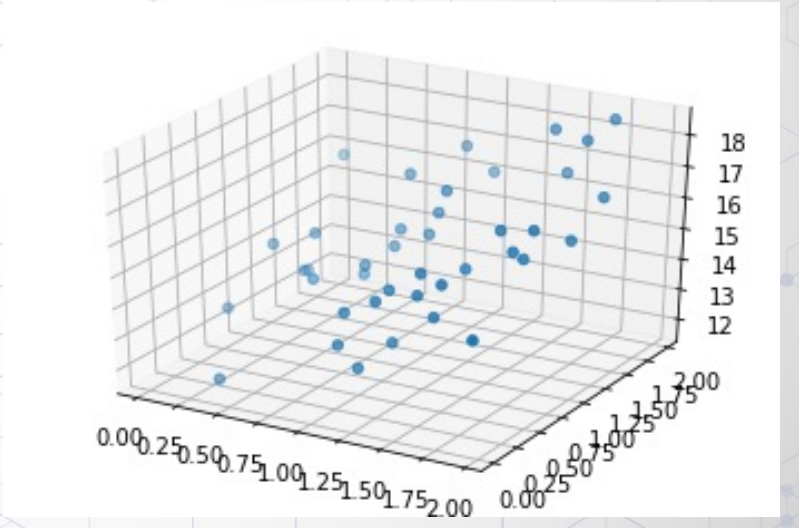
Distancia








9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
0.25
0.50
0.75
1.00
1.25
1.50
1.75


18
17
16
15
14
13
12
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00

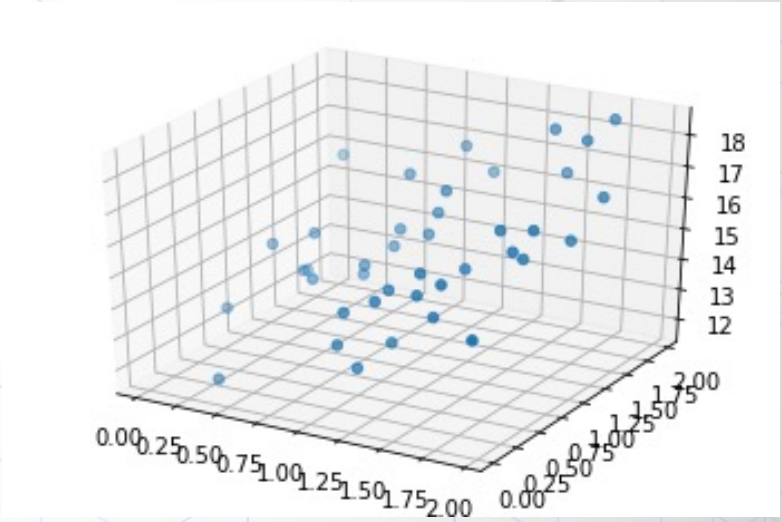
Distancia

$$D(X,Z) = \sqrt{(x_1 - z_1)^2 + (x_2 - z_2)^2}$$

$X = (3,5)$

$Z = (6,9)$

$D(X,Z) = ?$



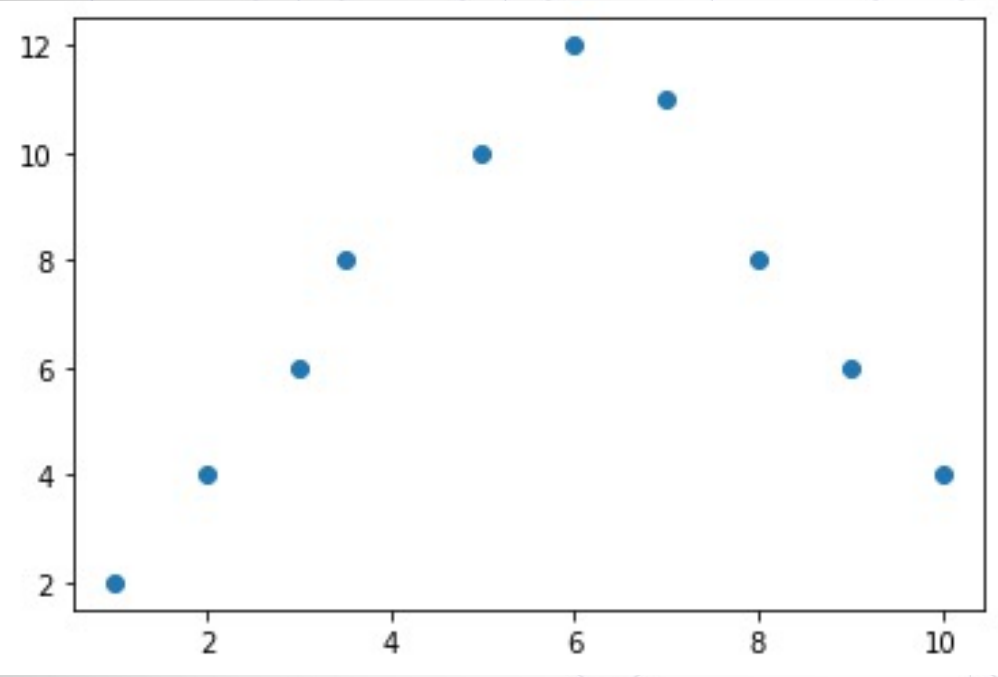
Vecinos

$K = ?$

Regresión

$$\hat{f}(x_0) = \frac{1}{k} \sum_{x^i \epsilon N_k(x_0)} y^i$$

Funcionamiento – Idea general



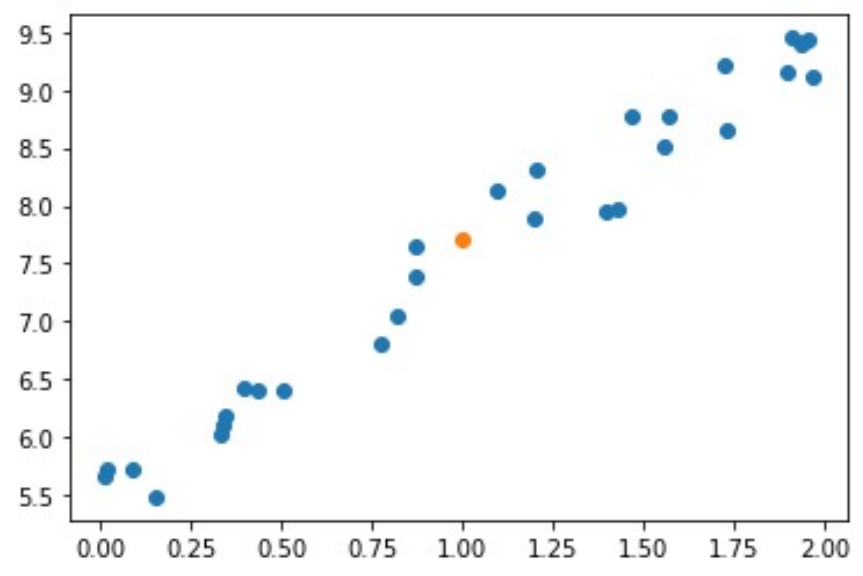
[1, 2]
[2 , 4]
[3 , 6]
[3.4, 8]
[5 , 10]
[6 , 12]
[7 , 11]
[8 , 8]
[9 , 6]
[10 , 4]

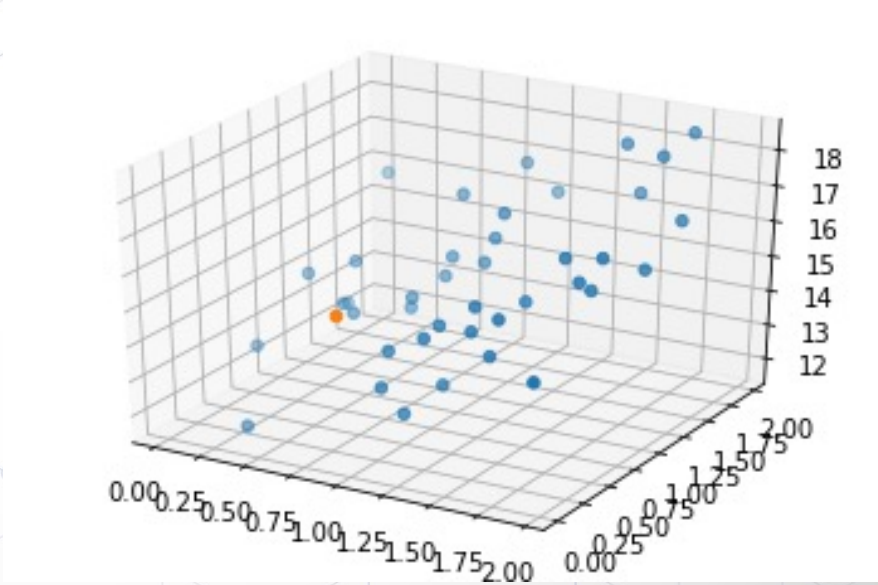
K = 3

x_0 = 5.5

y_0 = ?

Funcionamiento – Idea general



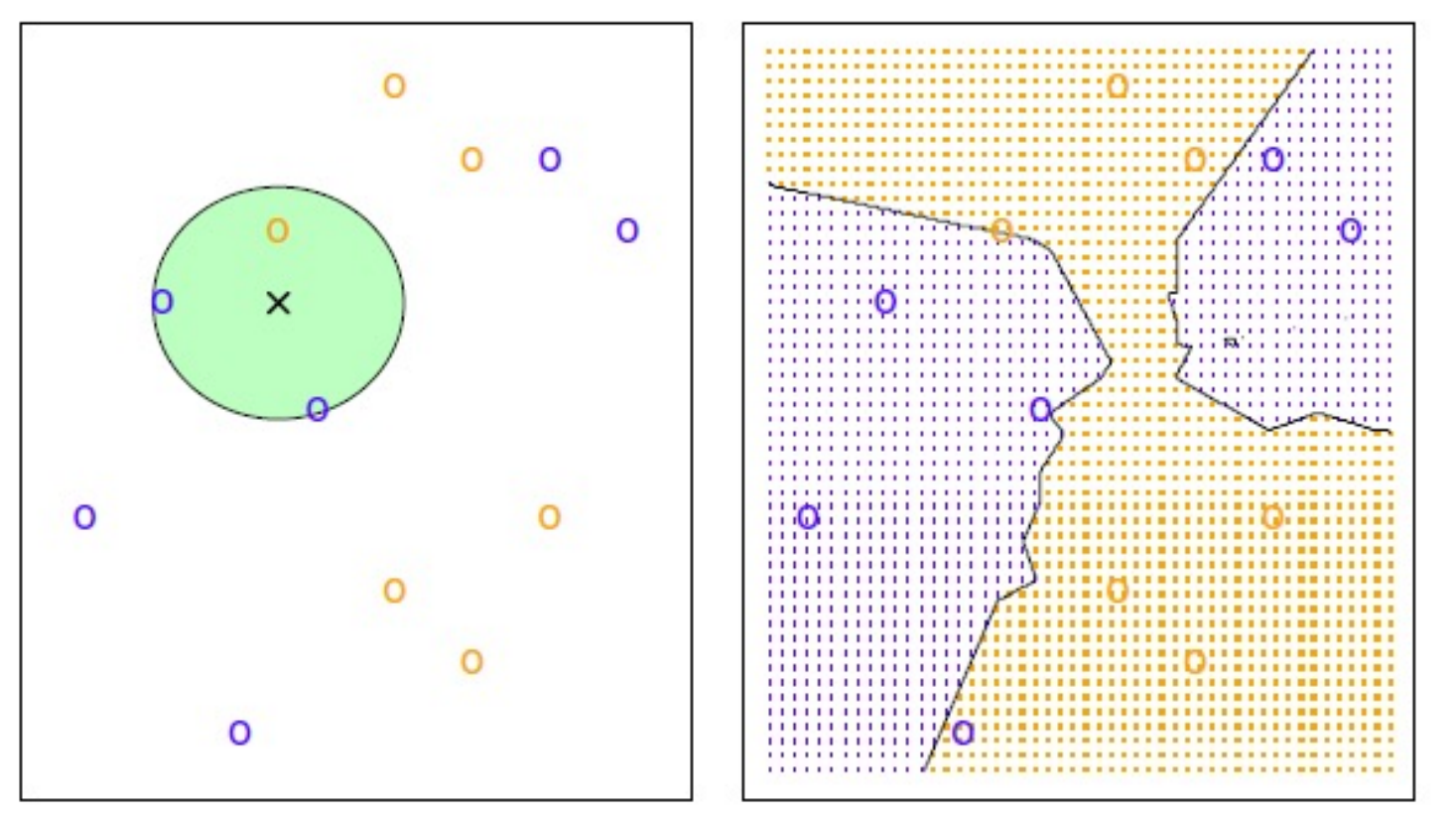


Código

Clasificación

KNN

Idea general



Idea general

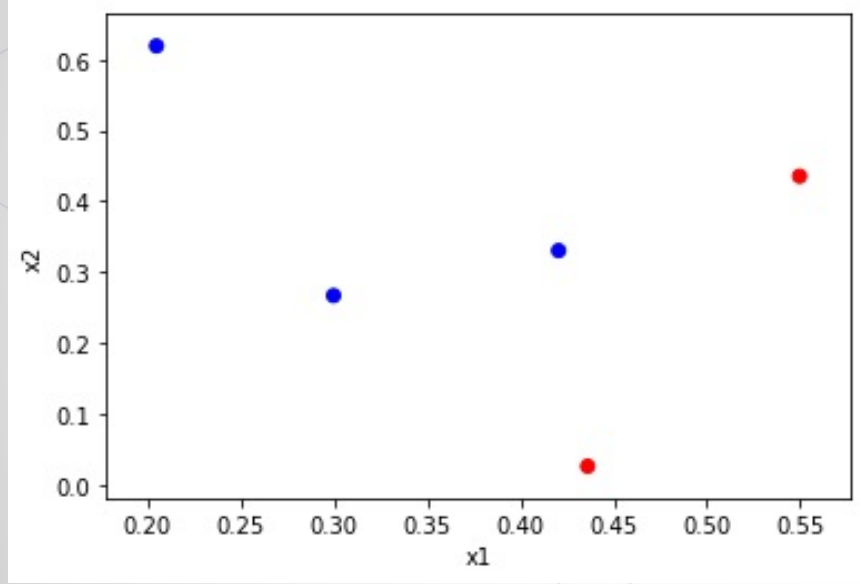
Regresión

$$\hat{f}(x_0) = \frac{1}{k} \sum_{x^i \epsilon N_k(x_0)} y^i$$

Clasificación

$$\Pr(Y = j | X = x_0) = \frac{1}{k} \sum_{x^i \epsilon N_k(x_0)} I(y^i = j)$$

Ejemplo



[0.43, 0.02] = Rojo

[0.54, 0.43] = Rojo

[0.42, 0.33] = Azul

[0.20, 0.61] = Azul

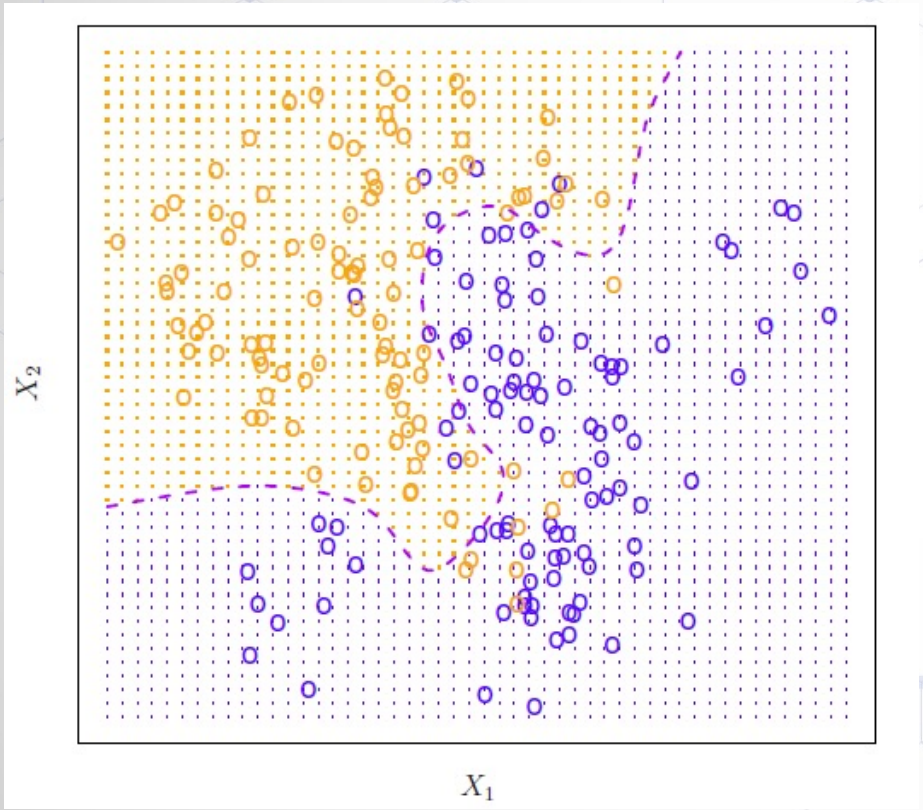
[0.29, 0.26] = Azul

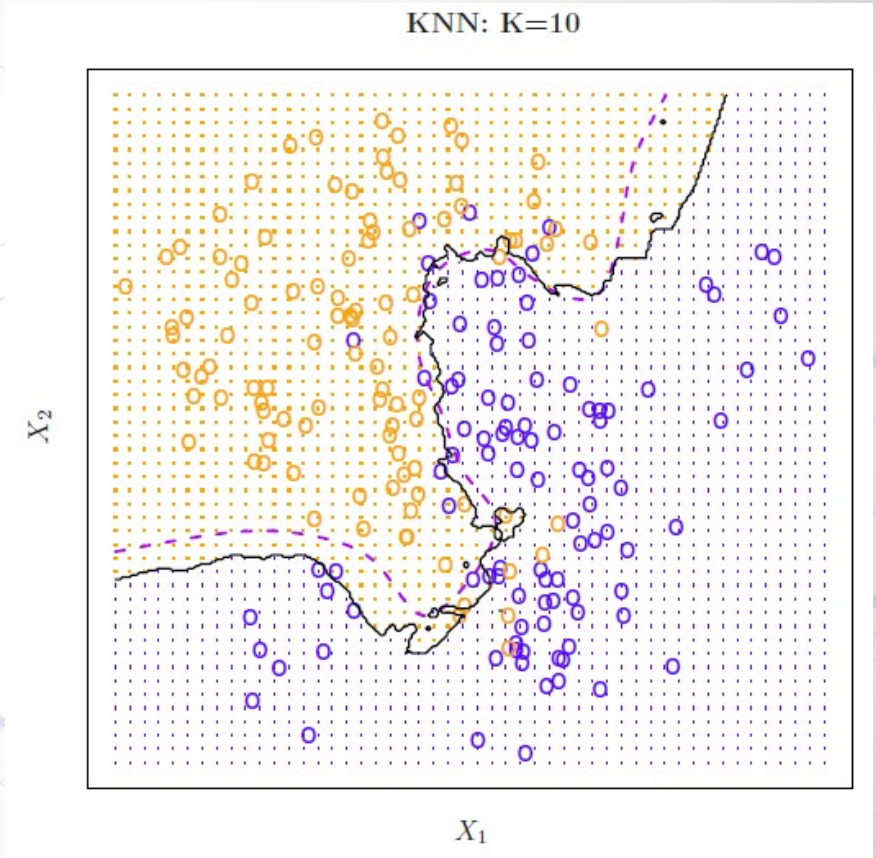
K = 3

x_0 = (0.5,0.5)

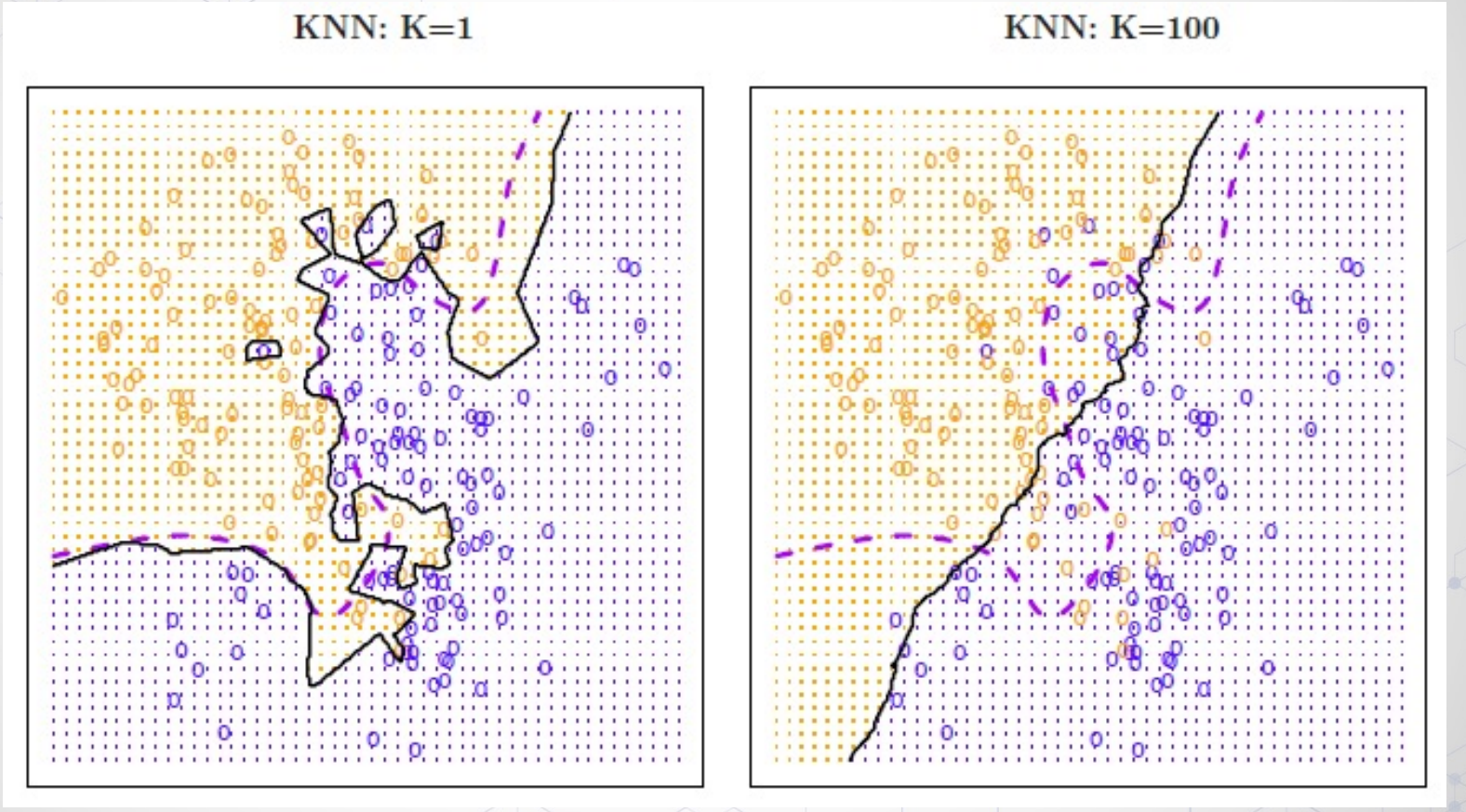
y_0 = ?

KNN





KNN

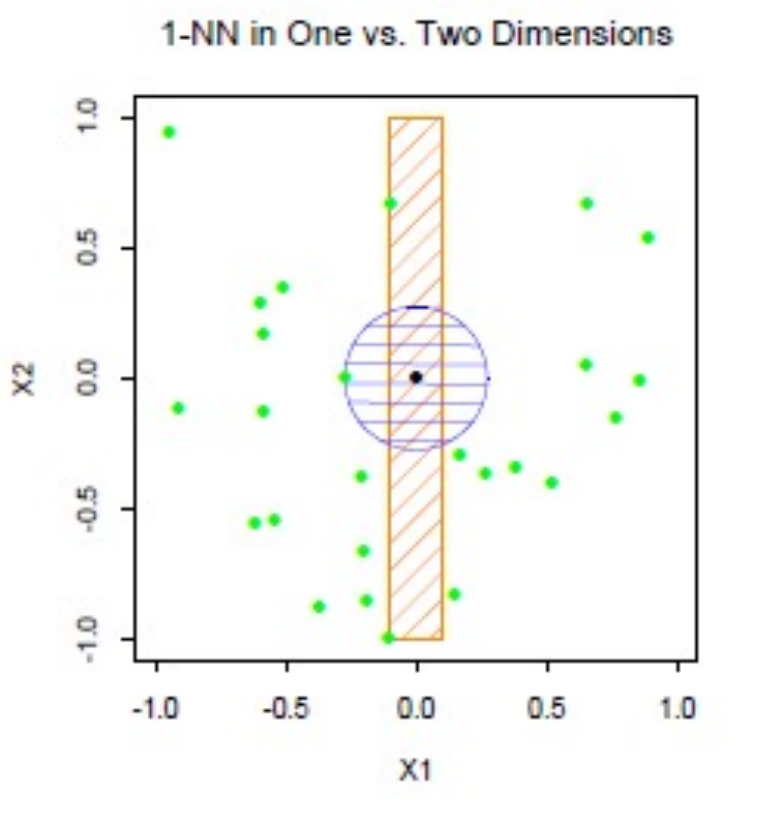


Código

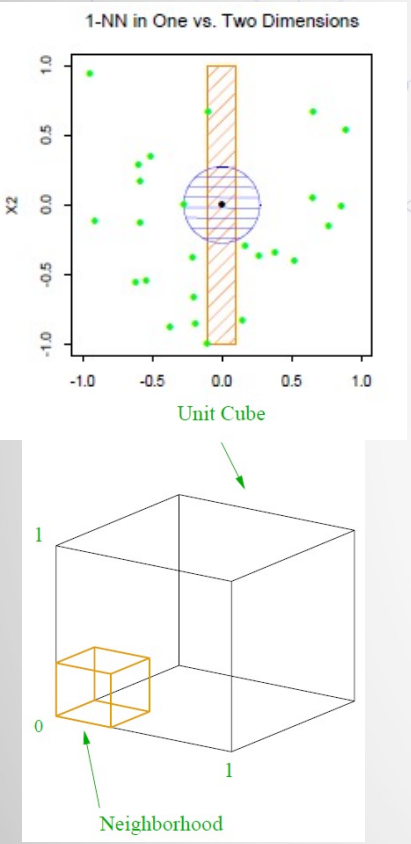
Comentarios finales

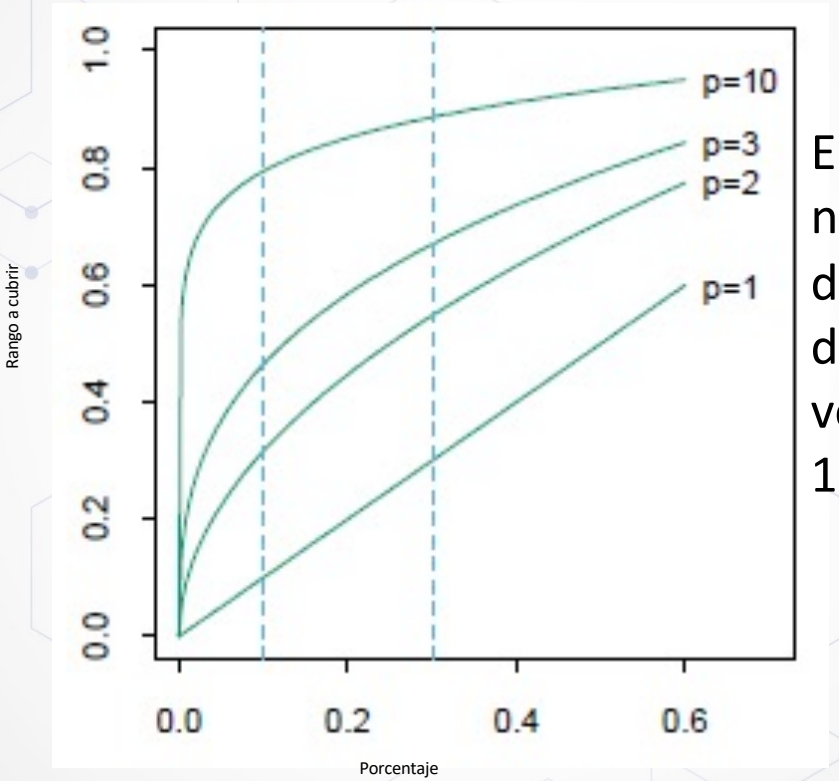
La maldición de la dimensionalidad

1D | 2D



Más dimensiones





En 10 dimensiones necesitamos cubrir 80% del rango de cada dimensión para cubrir una vecindad que contenga el 10% de los datos

EngineeringX

Founded by the Royal Academy of Engineering
and Lloyd's Register Foundation

GRACIAS

https://hubiq.mx/

HUBIQRO HUBIQ HUBIQRO