

Taller Visión Computacional e IA

Sesión 2 – Machine Learning

8° Escuela de Invierno en Ciencia de Datos

ARTIFICIAL INTELLIGENCE IN BIOMEDICINE
GROUP



Machine Learning

Supervised Learning

Unsupervised Learning

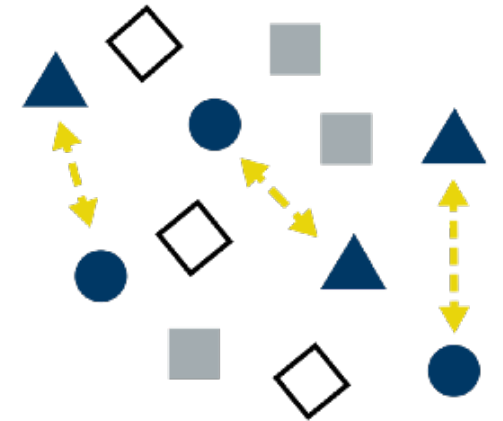
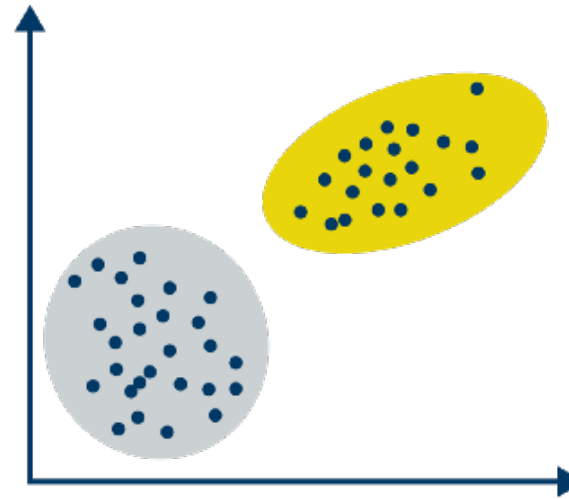
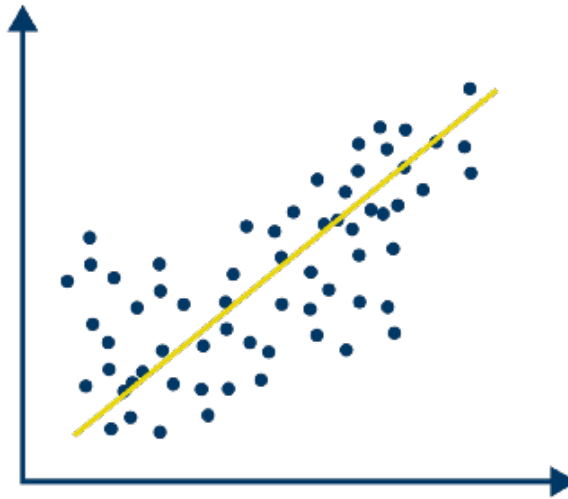
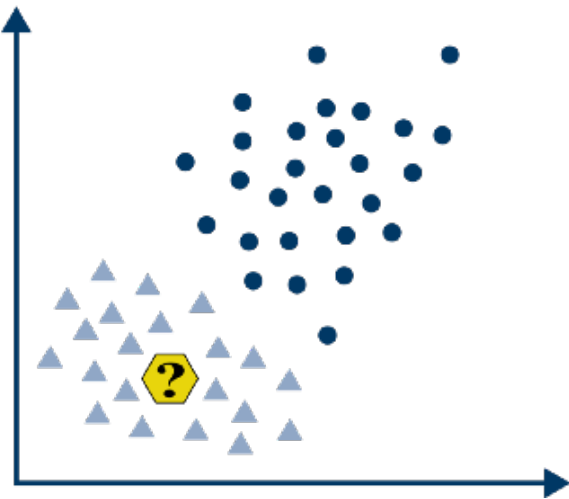
Reinforcement Learning

Classification

Regression

Clustering

Association



Aprendizaje

Extracción de
características
(Experto)

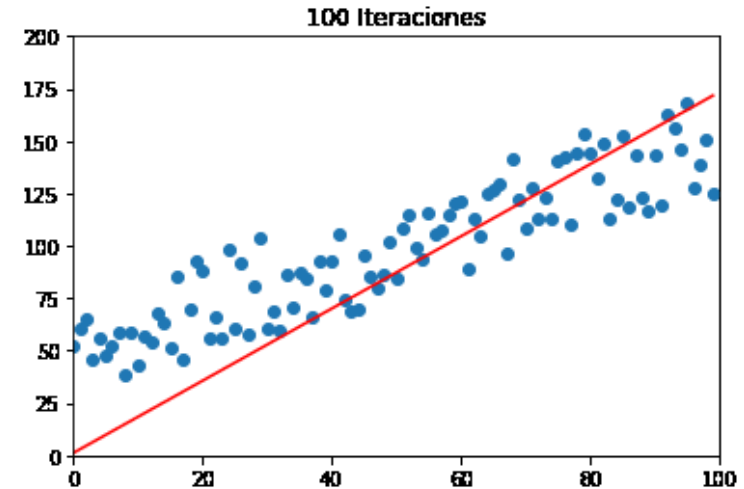
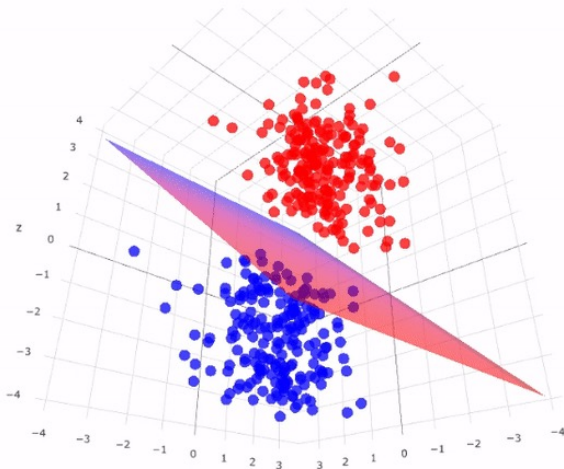
Entrenamiento

Validación

Prueba final
(Uso)

Clasificación

Regresión



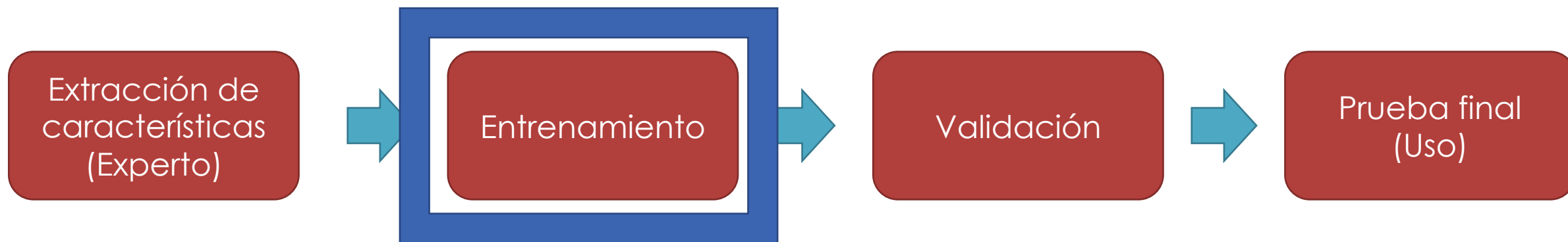
Extracción de Características

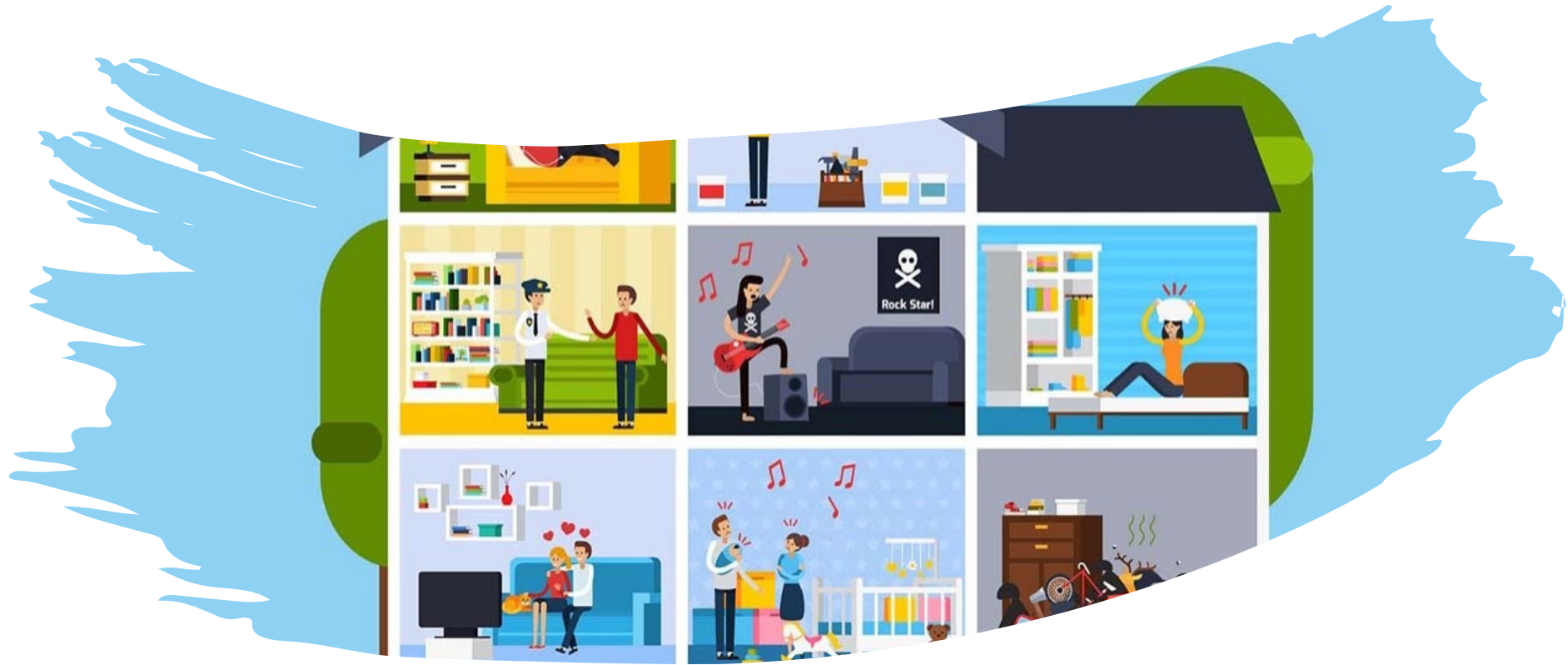




Entrenamiento

Aprendizaje





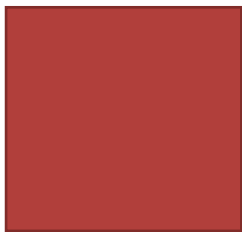
K-NN (K- Nearest Neighbour)

K-NN (K- Nearest Neighbour)

Características:

- Clasificador multiclase (2 clases o más)

Ej. Colores y/o formas

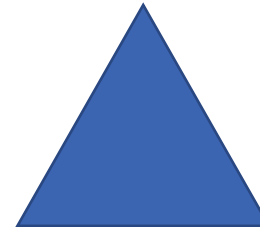


C1



C2

...



CN



K-NN (K- Nearest Neighbour)

Características:

- Clasificador multiclase (2 clases o más)
- Se basa en la cercanía entre vecinos (cercanía entre instancias)

Ej. Asiento de un nuevo aficionado?



K-NN (K- Nearest Neighbour)

Características:

- Clasificador multiclase (2 clases o más)
- Se basa en la cercanía entre vecinos (cercanía entre instancias)

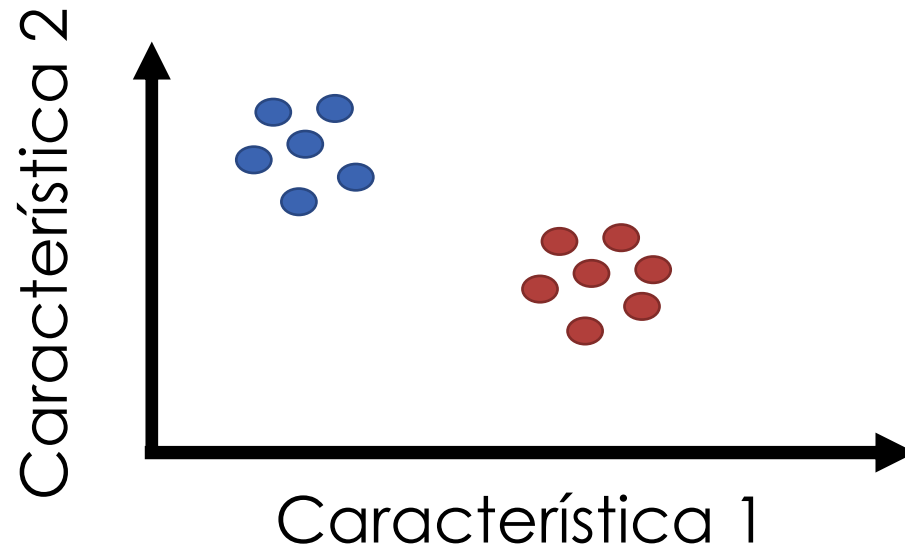
Ej. Asiento de un nuevo aficionado?



K-NN (K- Nearest Neighbour)

Características:

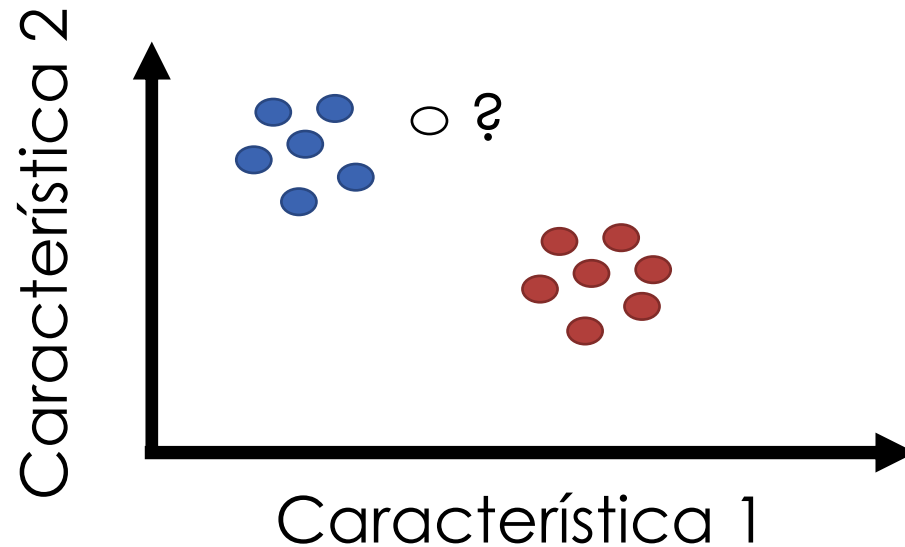
- Se basa en la cercanía entre vecinos (cercanía entre instancias)



K-NN (K- Nearest Neighbour)

Características:

- Se basa en la cercanía entre vecinos (cercanía entre instancias)



K-NN (K- Nearest Neighbour)

Dado un conjunto de datos y sus etiquetas:

		X_1	\dots	X_j	\dots	X_n	C
(\mathbf{x}_1, c_1)	1	x_{11}	\dots	x_{1j}	\dots	x_{1n}	c_1
	\vdots	\vdots		\vdots		\vdots	\vdots
(\mathbf{x}_i, c_i)	i	x_{i1}	\dots	x_{ij}	\dots	x_{in}	c_i
	\vdots	\vdots		\vdots		\vdots	\vdots
(\mathbf{x}_N, c_N)	N	x_{N1}	\dots	x_{Nj}	\dots	x_{Nn}	c_N
\mathbf{x}	$N + 1$	$x_{N+1,1}$	\dots	$x_{N+1,j}$	\dots	$x_{N+1,n}$?

$(\mathbf{x}_1, c_1), \dots, (\mathbf{x}_N, c_N)$

$\mathbf{x}_i = (x_{i,1} \dots x_{i,n})$

$c_i \in \{c^1, \dots, c^m\}$

donde

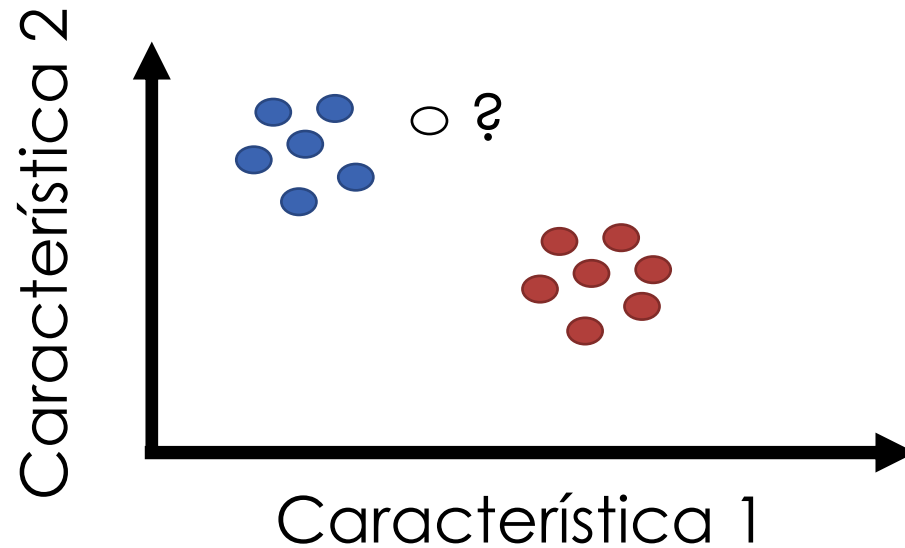
para todo $i = 1, \dots, N$

para todo $i = 1, \dots, N$



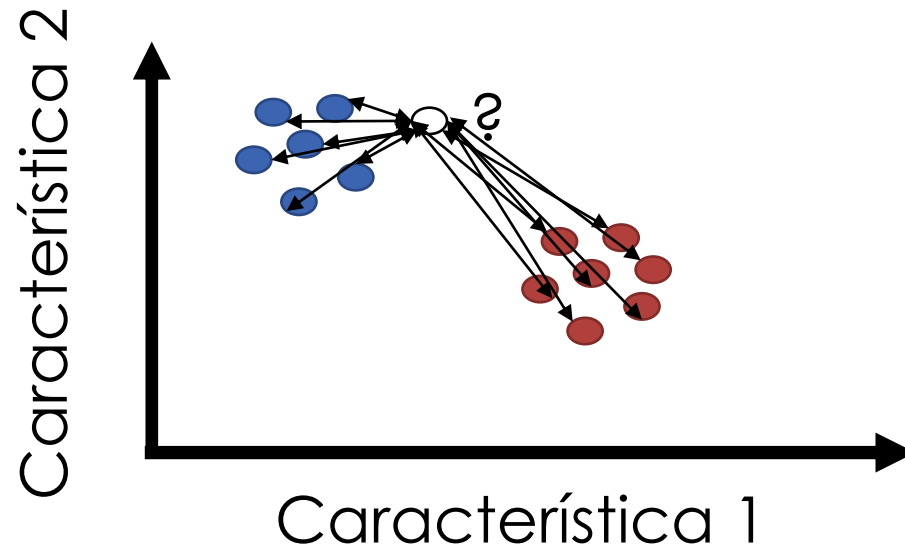
K-NN (K- Nearest Neighbour)

Criterio de clasificación



K-NN (K- Nearest Neighbour)

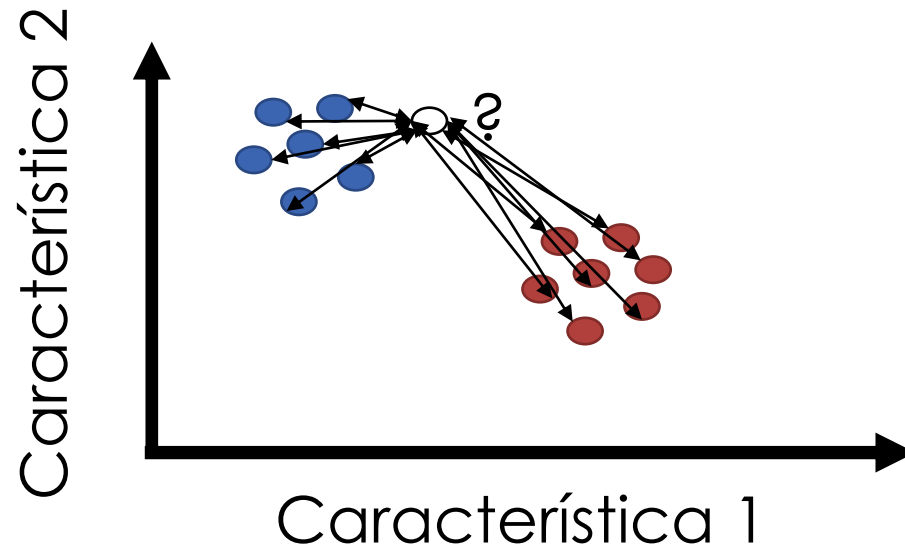
Criterio de clasificación



K-NN (K- Nearest Neighbour)

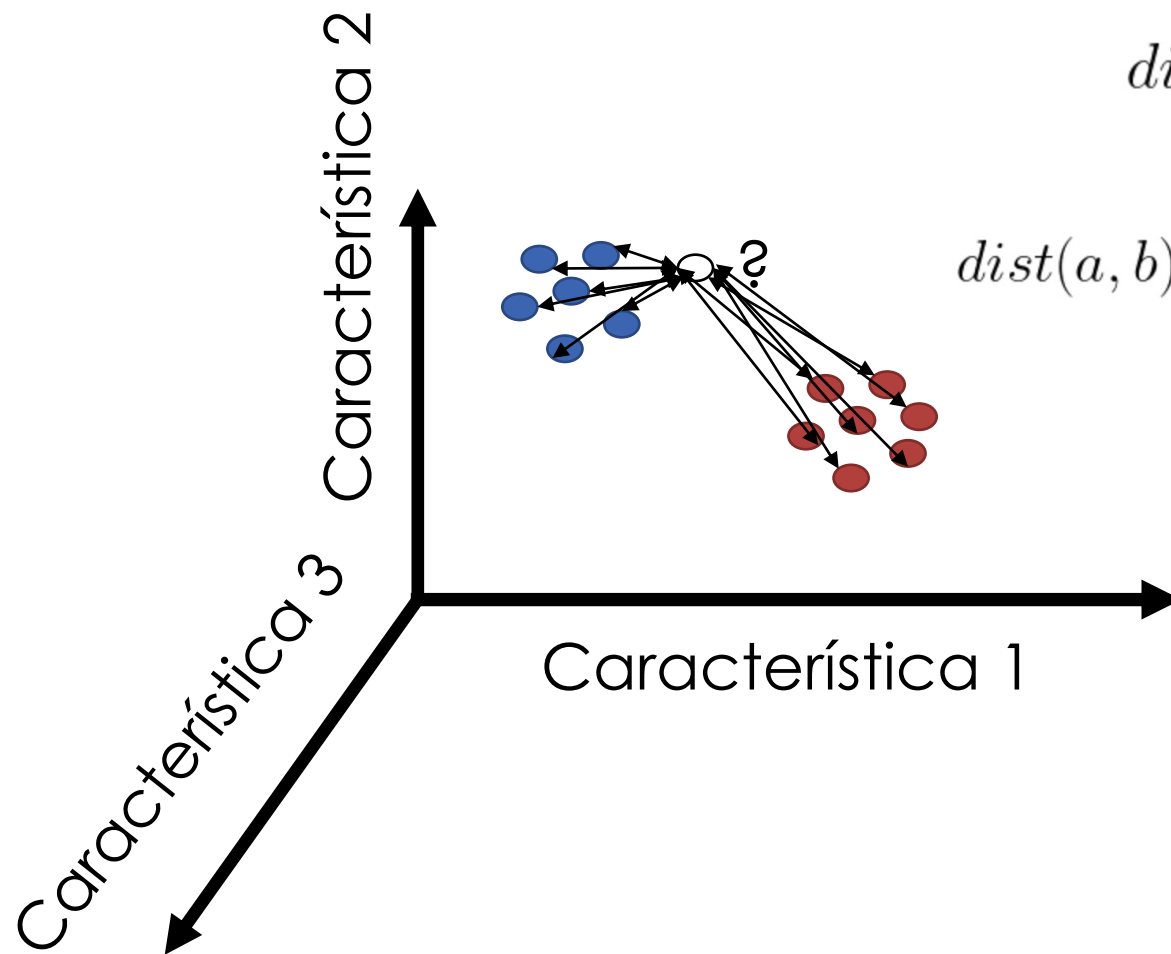
Criterio de clasificación 2D

$$\text{dist}(a, b) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$



K-NN (K- Nearest Neighbour)

Criterio de clasificación 3D



$$dist(a, b) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

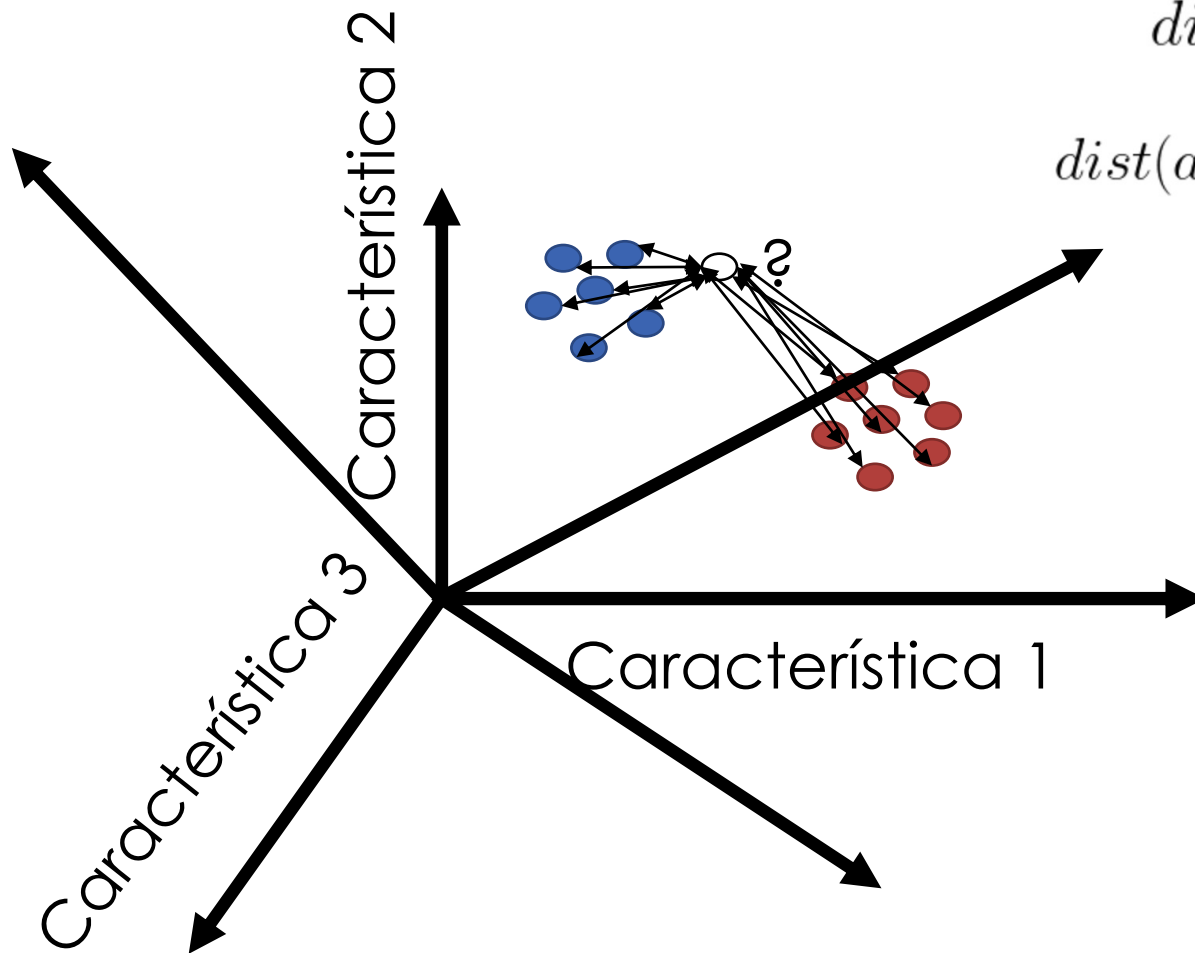
$$dist(a, b) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

K-NN (K- Nearest Neighbour)

Criterio de clasificación n-dimensional

$$dist(a, b) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

$$dist(a, b) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$



K-NN (K- Nearest Neighbour)

Criterio de clasificación n-dimensional

$$\mathbf{A} = \{a_1, a_2, \dots, a_n\} \quad \mathbf{B} = \{b_1, b_2, \dots, b_n\}$$

$$dist(\mathbf{A}, \mathbf{B}) = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2}$$

$$dist(\mathbf{A}, \mathbf{B}) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$



K-NN (K- Nearest Neighbour)

COMIENZO

Entrada: $D = \{(\mathbf{x}_1, c_1), \dots, (\mathbf{x}_N, c_N)\}$

$\mathbf{x} = (x_1, \dots, x_n)$ nuevo caso a clasificar

PARA todo objeto ya clasificado (x_i, c_i)

calcular $d_i = d(\mathbf{x}_i, \mathbf{x})$

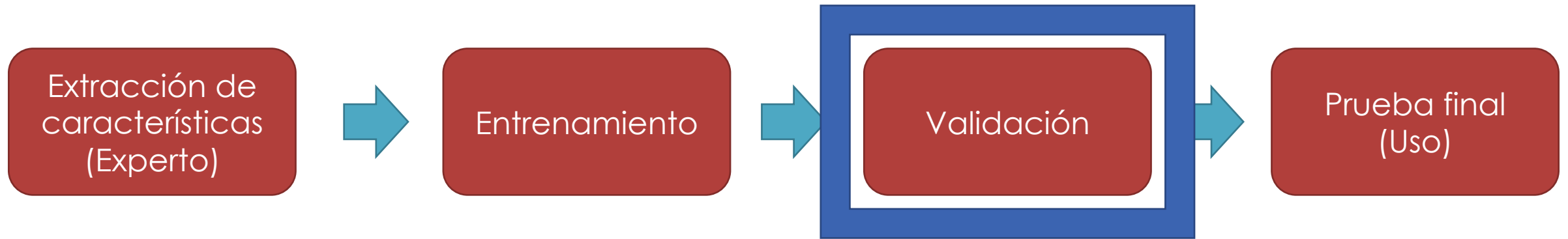
Ordenar $d_i (i = 1, \dots, N)$ en orden ascendente

Quedarnos con los K casos $D_{\mathbf{x}}^K$ ya clasificados más cercanos a \mathbf{x}

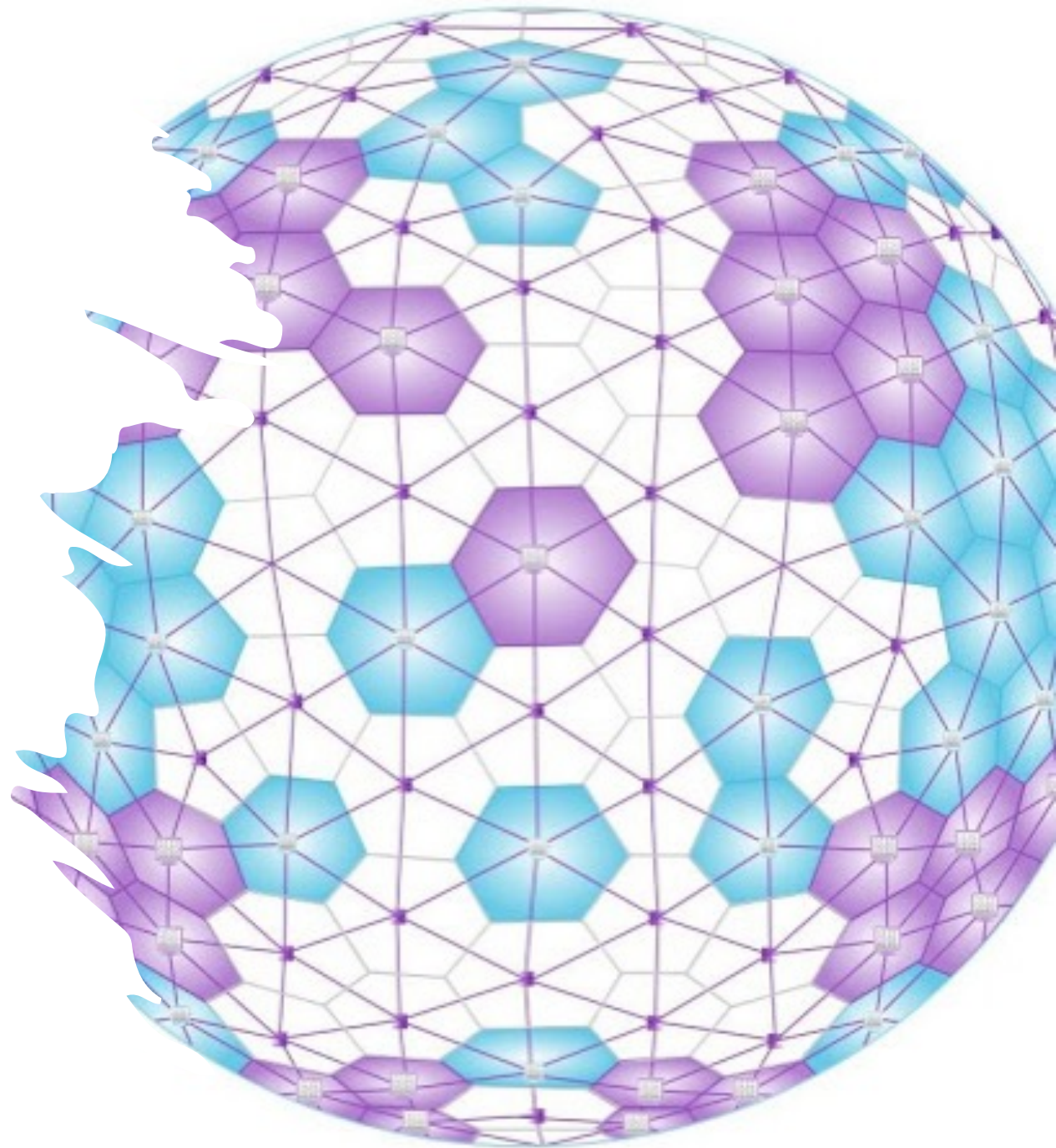
Asignar a \mathbf{x} la clase más frecuente en $D_{\mathbf{x}}^K$

FIN

Aprendizaje



Métricas de Validación



Matriz de Confusión

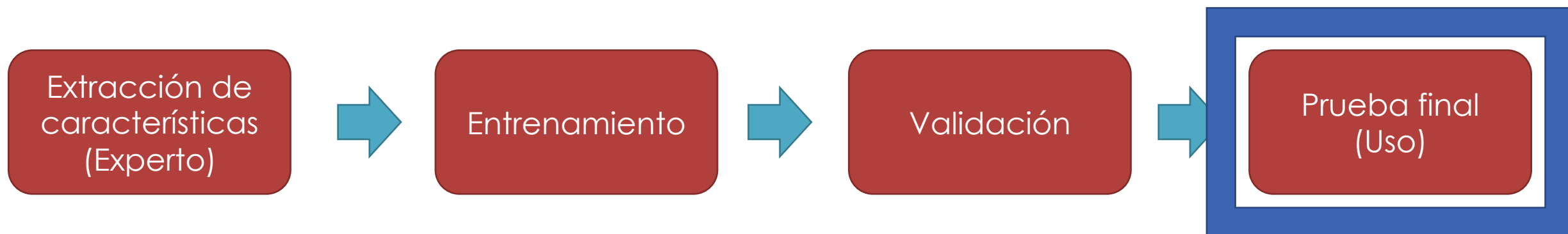
		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

$$\text{Sensibilidad} = \frac{VP}{VP + FN}$$

$$\text{Especificidad} = \frac{VN}{VN + FP}$$

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

Aprendizaje





¿Qué más hay?

- Random forest
- Máquinas de Soporte Vectorial
- Redes Neuronales
- Aprendizaje no Supervisado

Dr. Jorge Luis Perez Gonzalez
jorge.perez@iimas.unam.mx