
Algoritmo de K-Means

A large, faint, light blue watermark of the BDA logo is centered in the background of the slide. It consists of a cloud shape with a mountain peak inside, and the text "BDA" and "BIG DATA ACADEMY" overlaid on it.

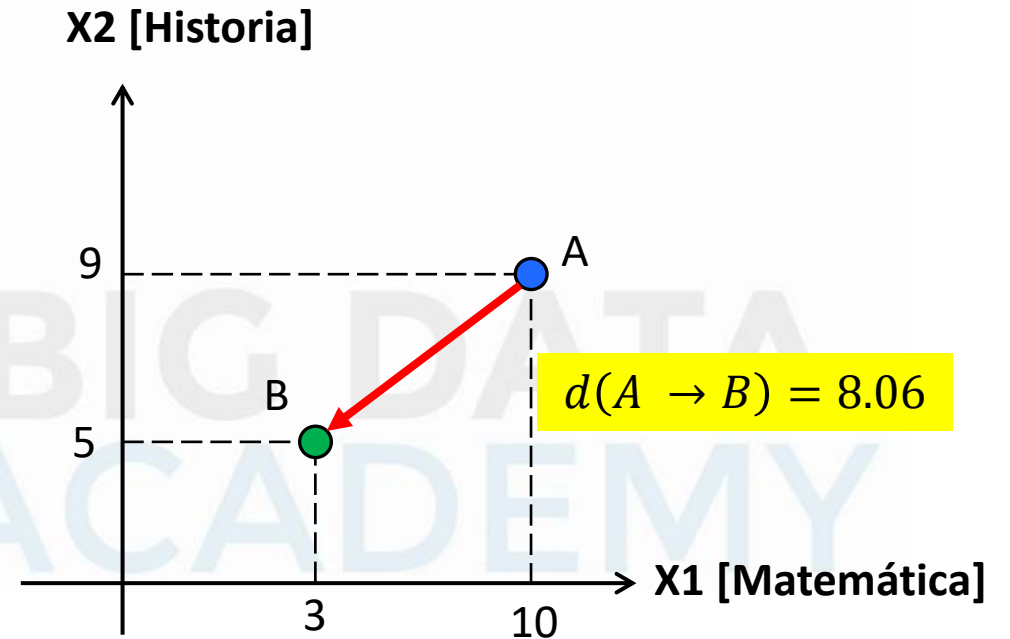
Objetivo

Modelo de Machine Learning
descriptivo que **crea clústers**
(segmentos) de registros que se
parecen entre sí

Concepto de distancia en puntos de dos dimensiones

PERSONA	MATEMÁTICA [X1]	HISTORIA [X2]
A	10	9
B	3	5

Cálculo de la distancia del punto A al punto B



$$d(A \rightarrow B) = \sqrt{(x_1^A - x_1^B)^2 + (x_2^A - x_2^B)^2}$$

Concepto de distancia en puntos de N dimensiones

- Tabla de datos N-Dimensional

PERSONA	MATEMÁTICA [X1]	HISTORIA [X2]	...	LITERATURA [Xn]
A	10	9	...	8
B	3	5	...	1

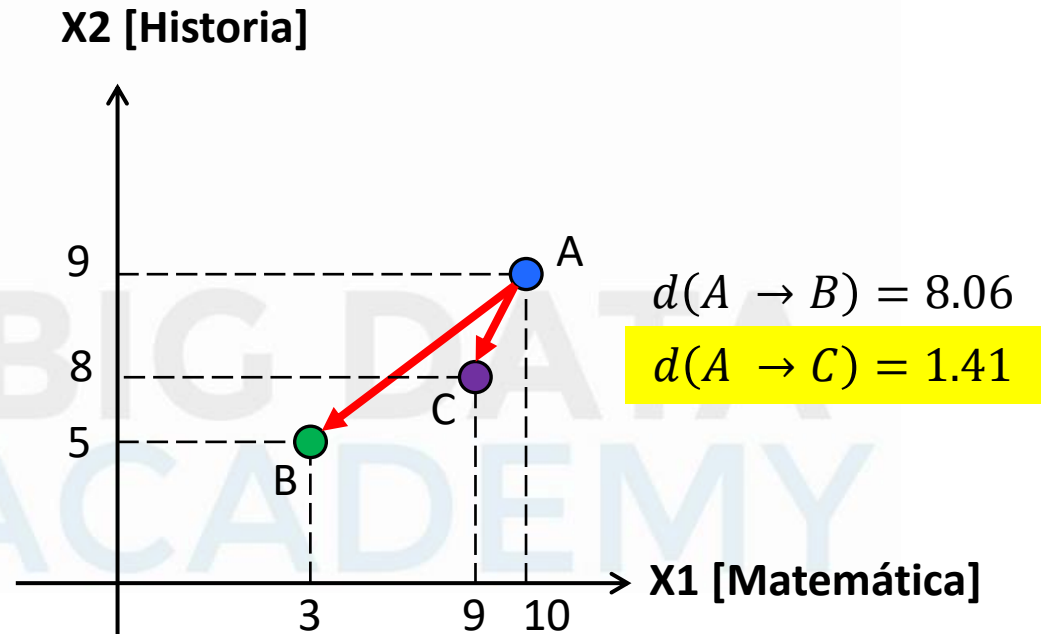
- Cálculo de la distancia del punto A al punto B en N-Dimensiones

$$d(A \rightarrow B) = \sqrt{(x_1^A - x_1^B)^2 + (x_2^A - x_2^B)^2 + \dots + (x_n^A - x_n^B)^2}$$

Concepto de similitud

La similitud determina qué tan parecidos son dos registros

PERSONA	MATEMÁTICA [X1]	HISTORIA [X2]
A	10	9
B	3	5
C	9	8



¿El alumno "A" a quién se parece más, a "B" o "C"?

El alumno "A" se parece al alumno "C"

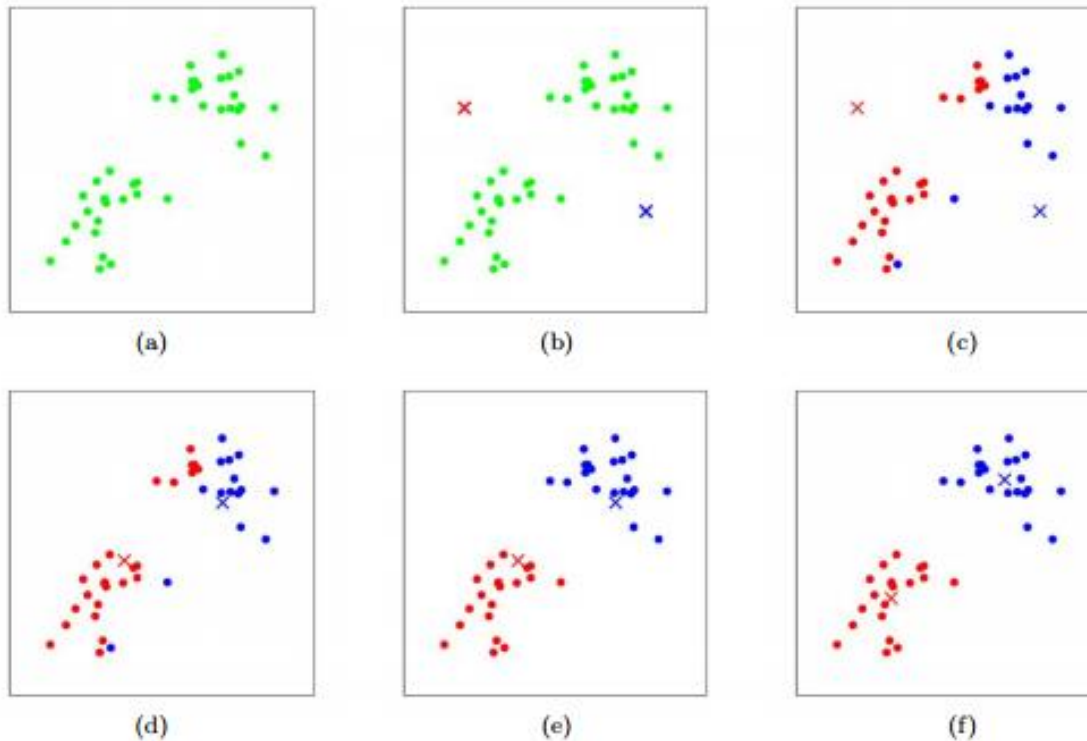
Pasos del algoritmo de K-Means



K-Means

1. Indicar cuántos clústers queremos encontrar
(p.e. $k=2$)
2. El algoritmo aleatoriamente asigna centroides
3. Se empieza a iterar, en cada iteración los registros se re-asignan al nuevo centroide más cercano

Ejemplo



El algoritmo finaliza cuando ya no es necesario re-asignar los elementos a otro clúster