

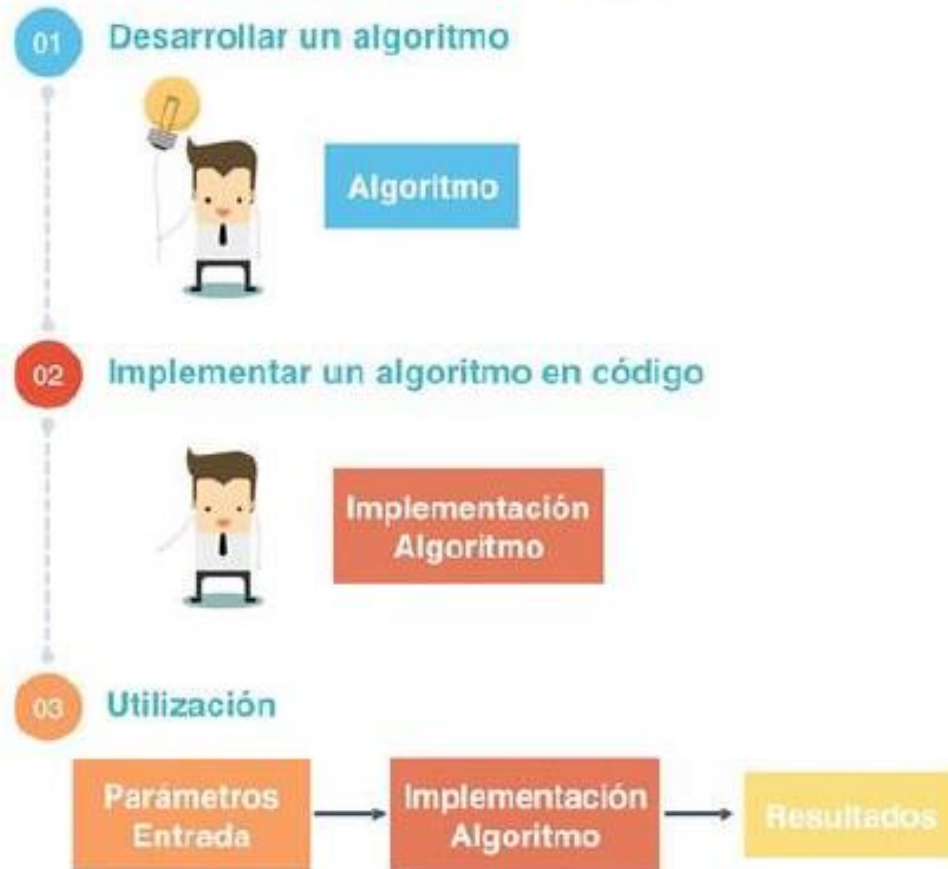
Machine Learning

(Aprendizaje Automático)

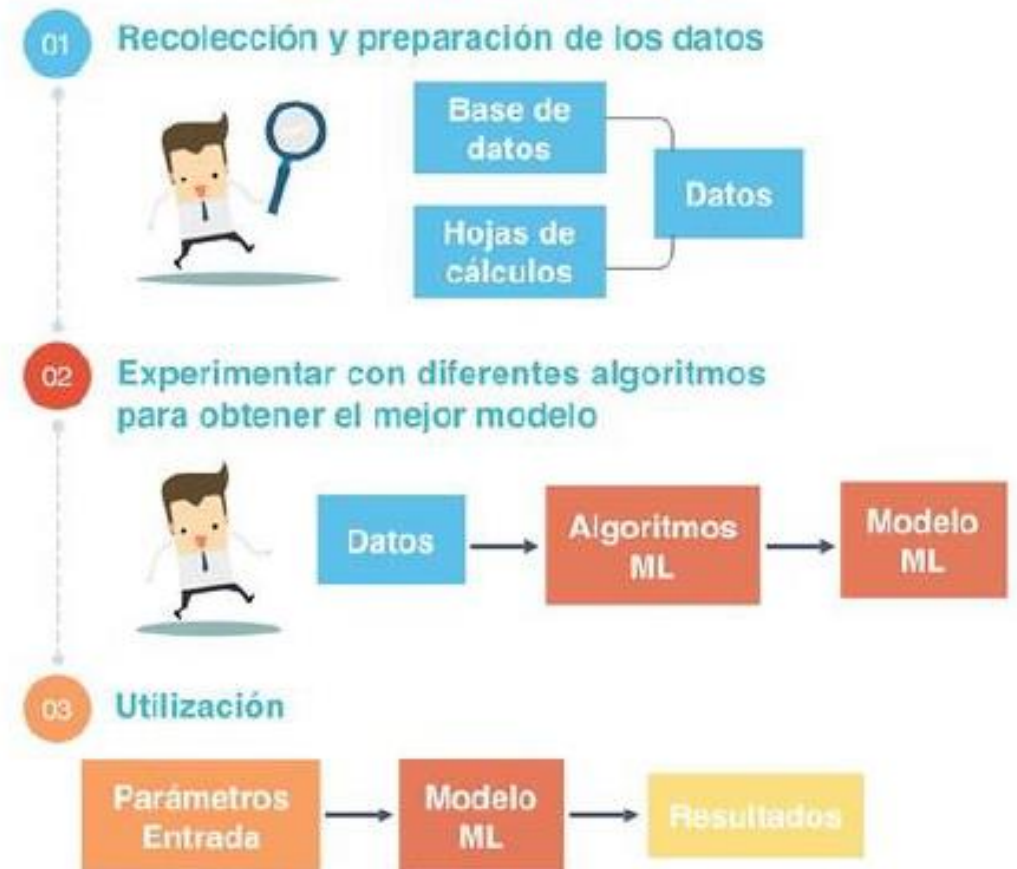
K-Means

Programación tradicional VS Machine Learning

Programación Tradicional



Machine Learning

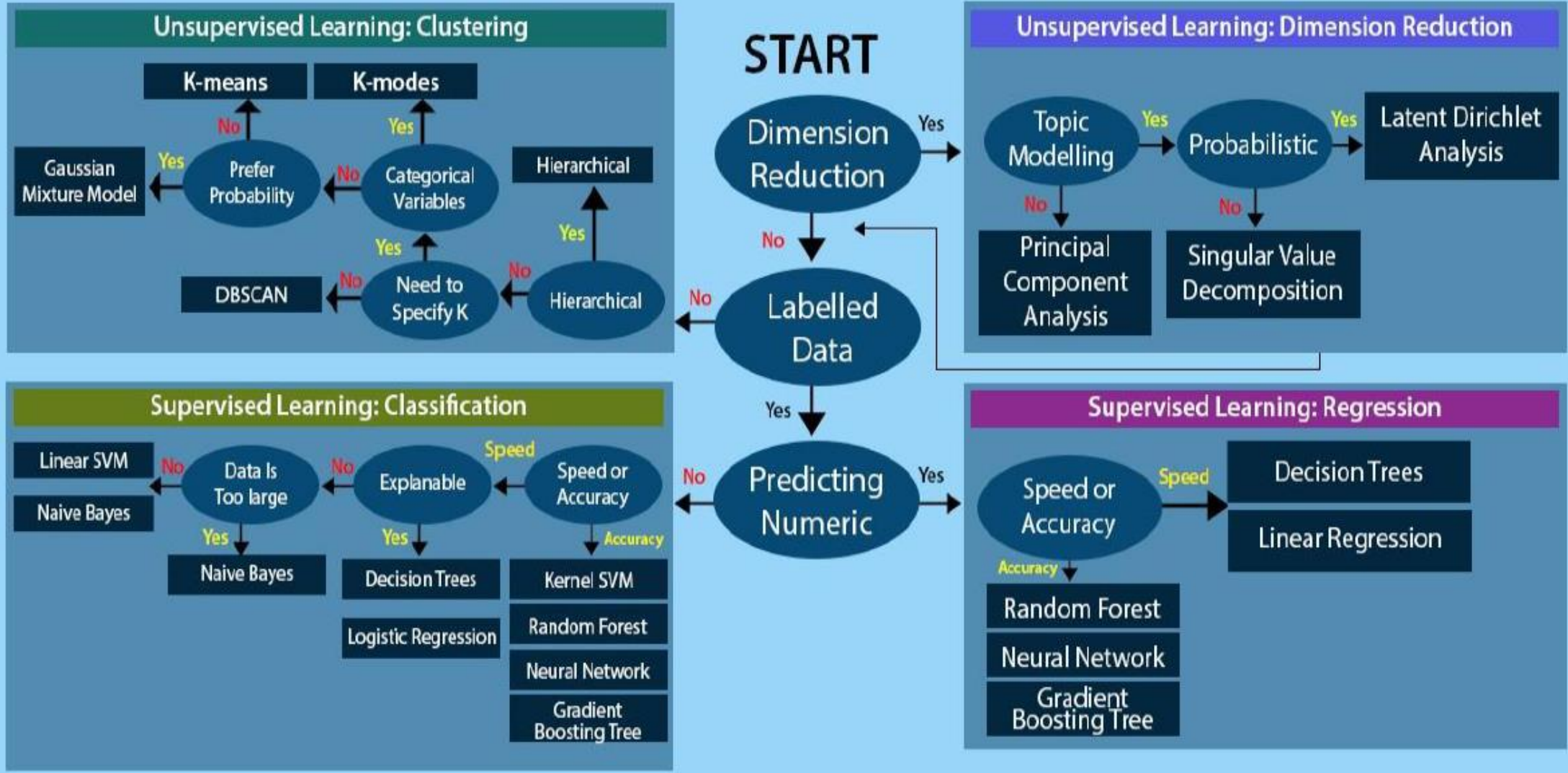


Tipos de Aprendizaje

- **Aprendizaje Supervisado:** Ejemplo: Clasificación
 - **Supervisión:** Los datos (observaciones, medidas, etc.) son etiquetados con clases predefinidas.
- **Aprendizaje No-Supervisado (Clustering/Agrupamiento)**
 - Las etiquetas de clase de los datos son desconocidas.
 - Dado un conjunto de datos, la tarea es establecer la existencia de clases o agrupaciones (clusters) en los datos.
 - Aprendizaje por observación

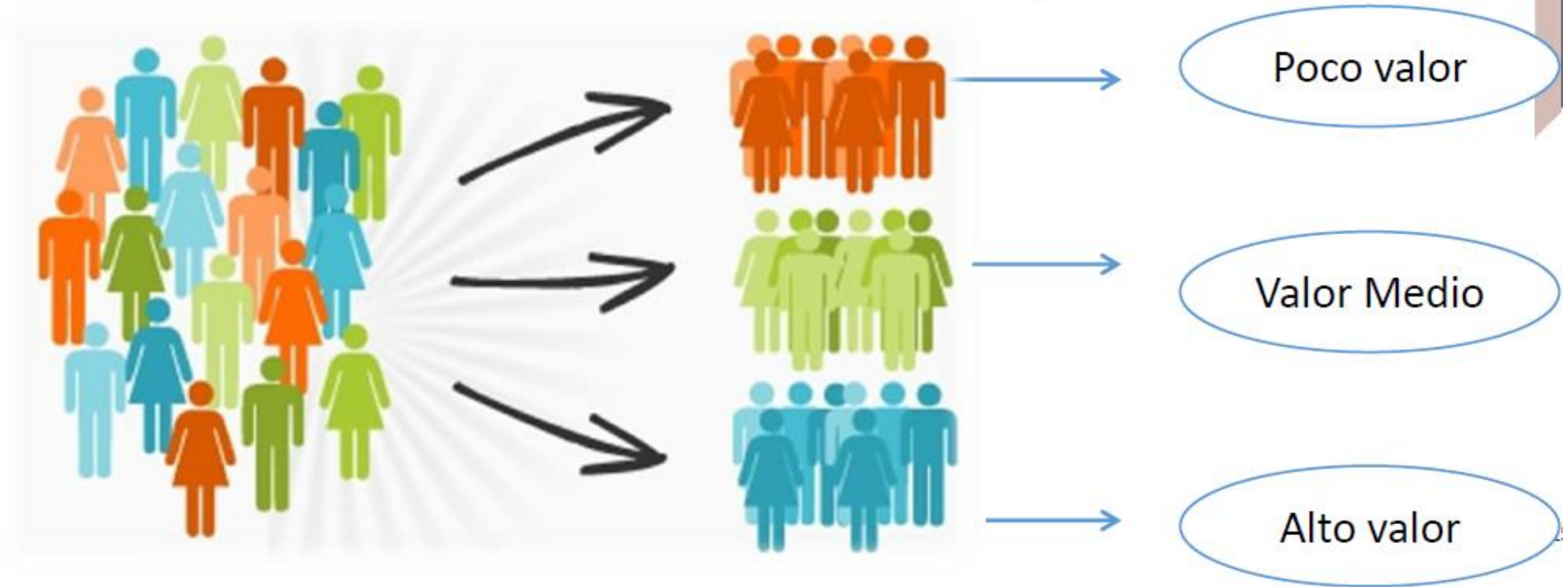
ID	Age	Has_Job	Own_House	Credit_Rating	Class
1	young	false	false	fair	No
2	young	false	false	good	No
3	young	true	false	good	Yes
4	young	true	true	fair	Yes
5	young	false	false	fair	No
6	middle	false	false	fair	No
7	middle	false	false	good	No
8	middle	true	true	good	Yes
9	middle	false	true	excellent	Yes
10	middle	false	true	excellent	Yes
11	old	false	true	excellent	Yes
12	old	false	true	good	Yes
13	old	true	false	good	Yes
14	old	true	false	excellent	Yes
15	old	false	false	fair	No

Machine Learning Algorithms Cheat Sheet



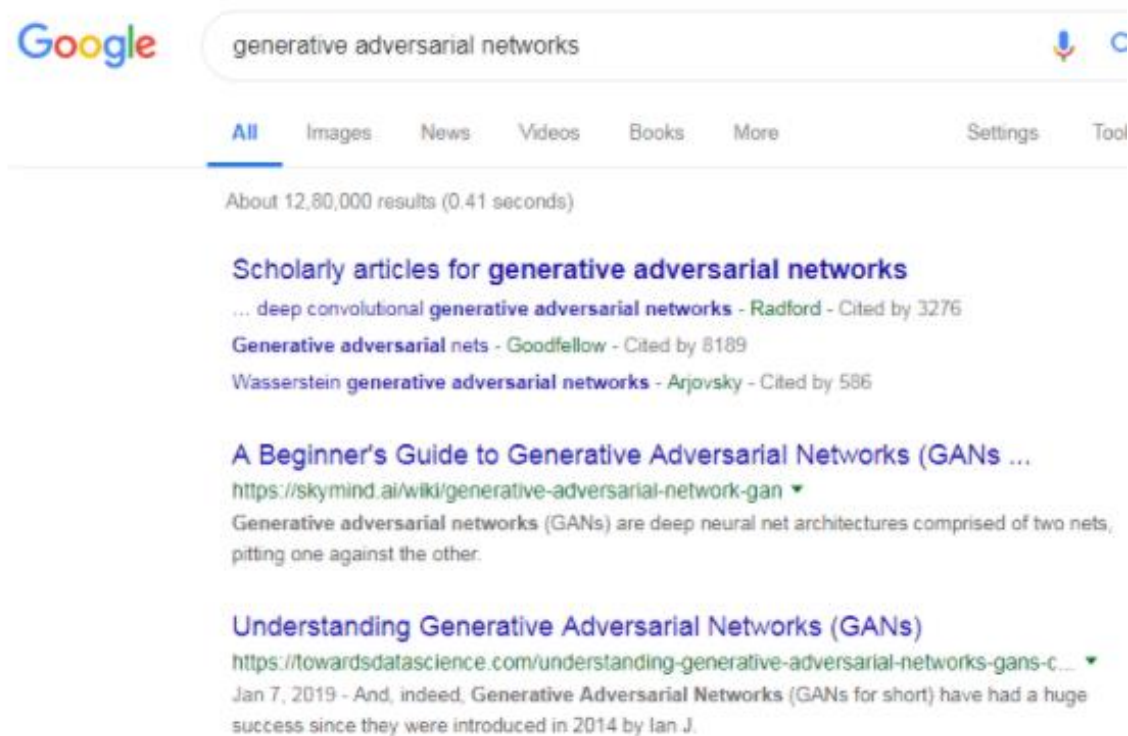
Segmentación (Clustering)

Objetivo: Organizar en grupos homogéneos

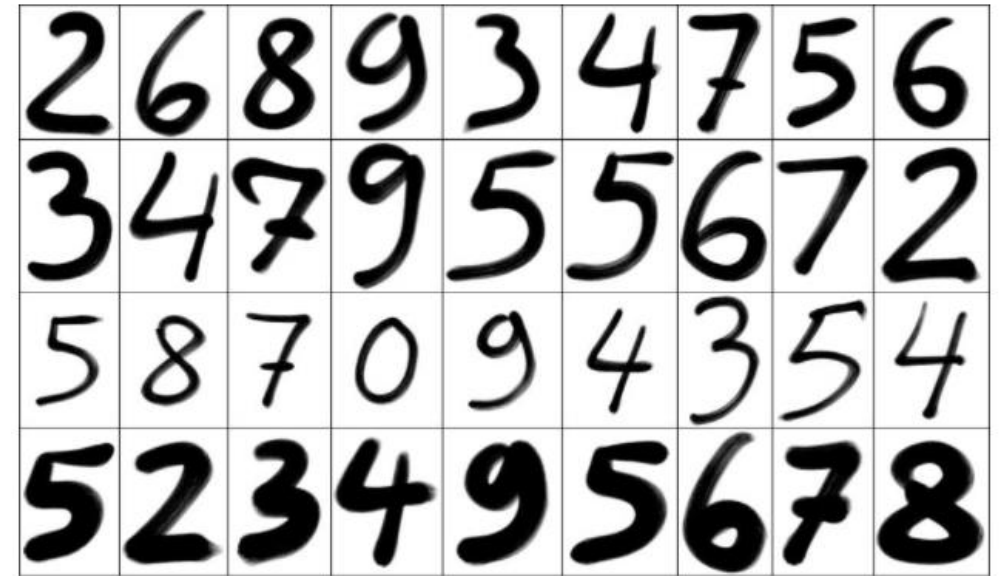


Ejemplos en los que se usa segmentación

Búsqueda en Web



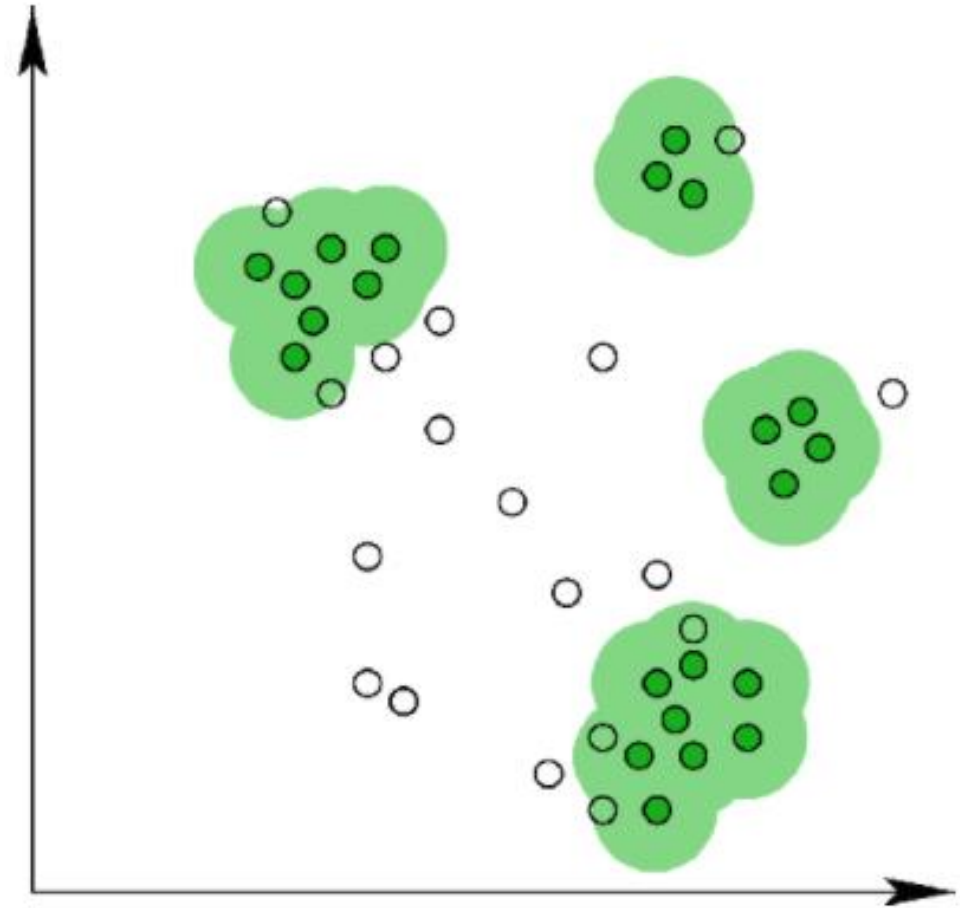
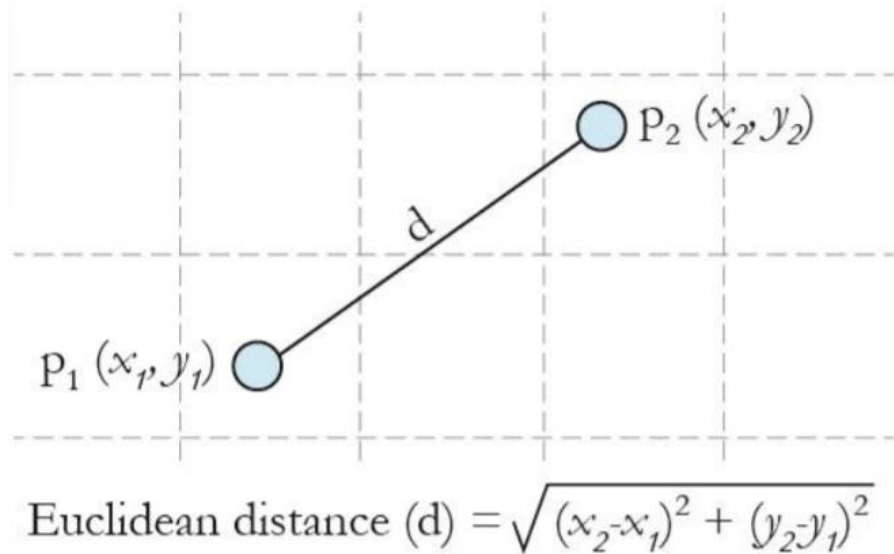
Reconocimiento de imágenes



<https://mathematicaforprediction.wordpress.com/2013/08/26/classification-of-handwritten-digits/>

K-Means: Método basado en particiones

- Agrupan los objetos al medir su distancia de otros objetos en un plano de n dimensiones.
- Se les conoce como métodos basados en distancia. Generalmente se utiliza la distancia Euclidiana

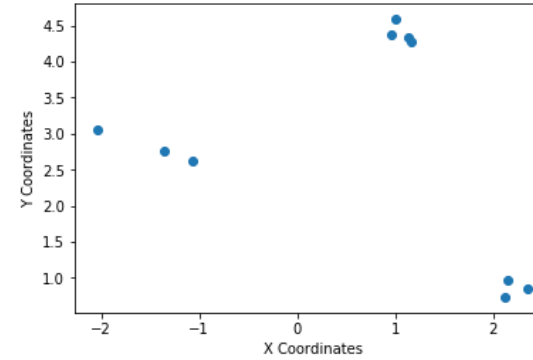
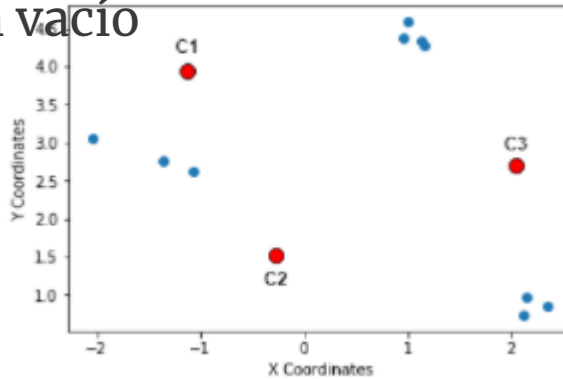


Algoritmo K-means

Datos de entrada

- k : número de grupos (*clusters*) en los que se va a segmentar la población.
- Población de puntos

- 1) Coloca k puntos aleatorios (centroides) e inicializa los clusters en vacío



- 2) Calcula las distancias de cada punto con el centroide. Se puede usar distancia Euclidiana

$$\sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

- 3) Compara las distancias y asigna cada punto al cluster que corresponda con el centroide más cercano.

- 4) Calcula nuevos centroides, en base a la media en x y en y de los n puntos que pertenecen al cluster

$$\frac{\sum_{i=1}^n x_i}{n}, \frac{\sum_{i=1}^n y_i}{n}$$

