

## Algoritmo K-means

Es un algoritmo de clasificación no supervisada que agrupa objetos en  $k$  grupos basándose en sus características. El agrupamiento se realiza minimizando la suma de distancias entre cada objeto y el centroide de su grupo.

Se puede usar una diversidad de métricas de similitud para el cálculo de la distancia. Sin embargo, la distancia euclidiana es la más común.

Pasos de algoritmo:

1. Se carga la instancia de datos a agrupar
2. Se establece el valor de  $K$
3. Se establecen a los  $K$  centroides dentro del conjunto de registros de la instancia. Generalmente son escogidos aleatoriamente. No se deben repetir los centroides
4. Se asigna a los registros al grupo del centroide con el que tengan una menor distancia (centroide más cercano).
5. Para cada centroide “ $i$ ”:
  - 5.1 Se calcula la media “ $m$ ” para cada característica “ $j$ ” del conjunto de los registros que pertenece al centroide “ $i$ ”
  - 5.2 Se utiliza a “ $m$ ” para actualizar a la característica “ $j$ ” del centroide “ $i$ ”.
6. Se limpian los  $K$  grupos
7. Se vuelve al paso 4 hasta que los centroides no se mueven (NO CAMBIEN), o se mueven por debajo de una distancia umbral en cada repetición. También se puede establecer un número de iteraciones de repetición.

La función del algoritmo K-Means es optimizar (minimizar) la suma de las distancias de cada registro al centroide de su grupo.