

# 1. Unsupervised Learning

## 1. Unsupervised Learning

- a) Clustering
  - 1) K-means.
  - 2) Clustering jerárquico.
  - 3) DBSCAN.
  - 4) Mezcla de Gaussianas.
- b) Association
  - 1) Reglas de asociación.
  - 2) algoritmo Apriori
- c) Dimensionality Reduction

### 1.1. Clustering

En este tipo de problemas se trata de agrupar los datos. Agruparlos de tal forma que queden definidos N conjuntos distinguibles, aunque no necesariamente se sepa que signifiquen esos conjuntos. El agrupamiento siempre será por características similares.

Cuantos clusters elegir:

1. Regla del codo (Elbow Method). fer
2. Método de Silhouette. fer
3. Estadística de Hopkins. fer

**Regla del codo (Elbow Method)** En el grafico buscamos un 'codo', el lugar donde baja abruptamente.

- Elegimos un rango, ejemplo 1 a 10, y para cada valor:
  - Para cada centroide calculamos la distancia promedio.

#### 1.1.1. K-means

1. El usuario decide la cantidad de grupos.
2. K-Means elige al azar K centroides.
3. Decide qué grupos están más cerca de cada centroide. Esos puntos forman un grupo.
4. K-Means recalcula los centroides al centro de cada grupo
5. K-Means vuelve a reasignar los puntos usando los nuevos centroides. Calcula nuevos grupos
6. K-means repite punto 4 y 5 hasta que los puntos no cambian de grupo.

Un ejemplo de este algoritmo seria el conjunto de datos Iris, donde se tiene 5 columnas (Largo de sépalo, Ancho de sépalo, Largo de pétalo, Ancho de pétalo, Especies). La columna Especies no la usamos por que estamos en Unsupervised Learning. Con las columnas restantes tenemos que buscar cuantos clusters hay y luego podria compara con la columna Especies.