

Módulo 4: Aprendizaje no Supervisado

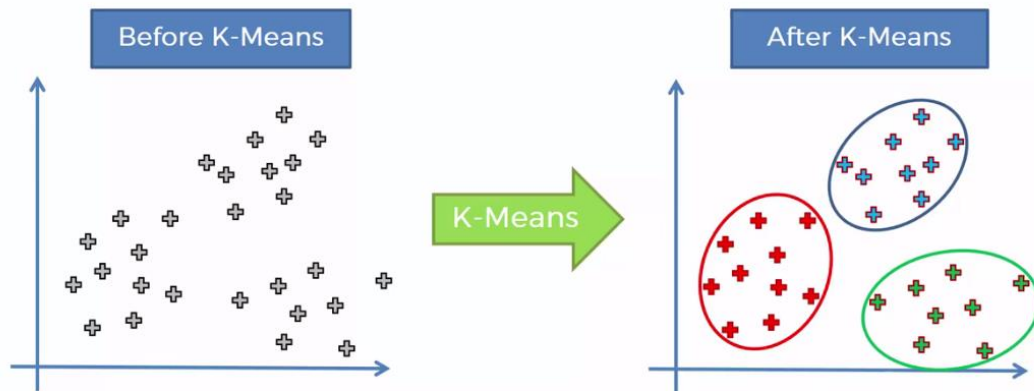
4.1 Algoritmos de Agrupación

Rafael Zambrano

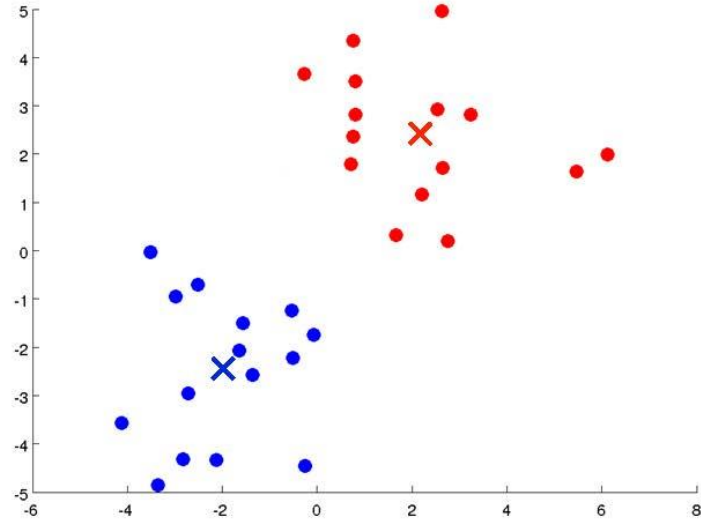
rafazamb@gmail.com

Clustering

- Los algoritmos de clustering forman parte de las técnicas de aprendizaje no supervisado, en las que no hay resultado que predecir y el algoritmo solo trata de encontrar patrones en los datos.
- El algoritmo de agrupación más utilizado es el denominado k-means

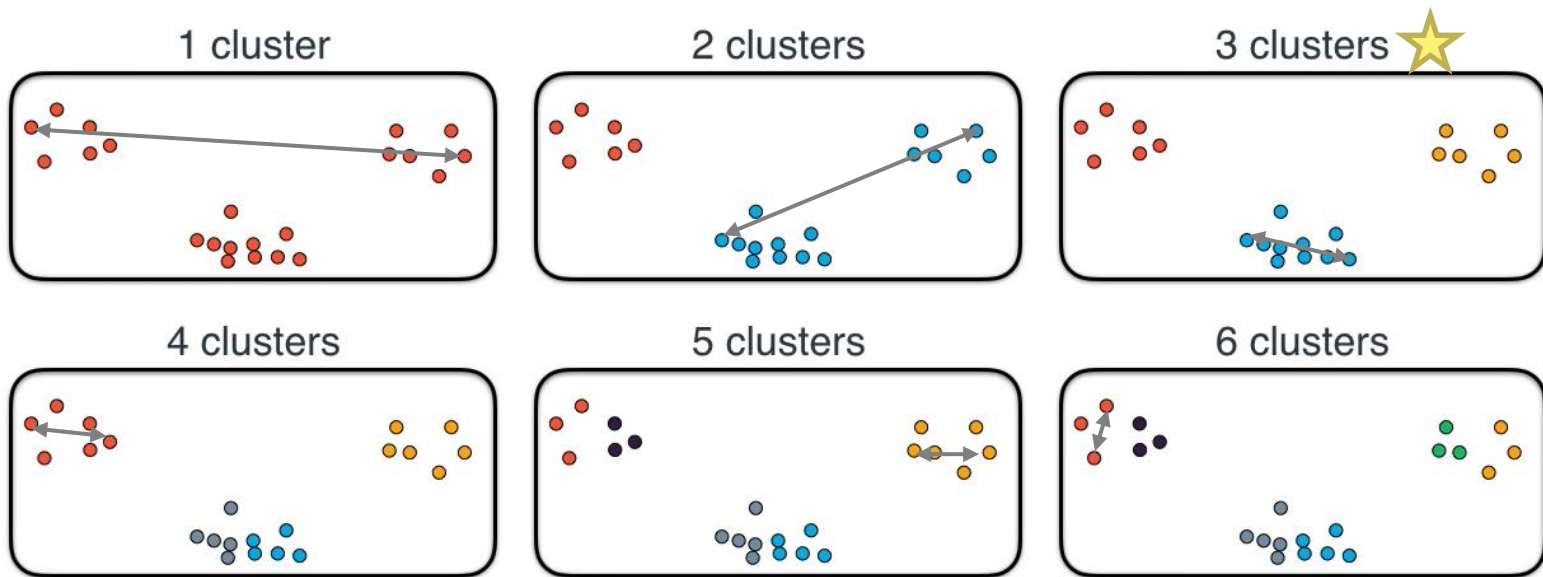


Clustering K-means



Método del codo

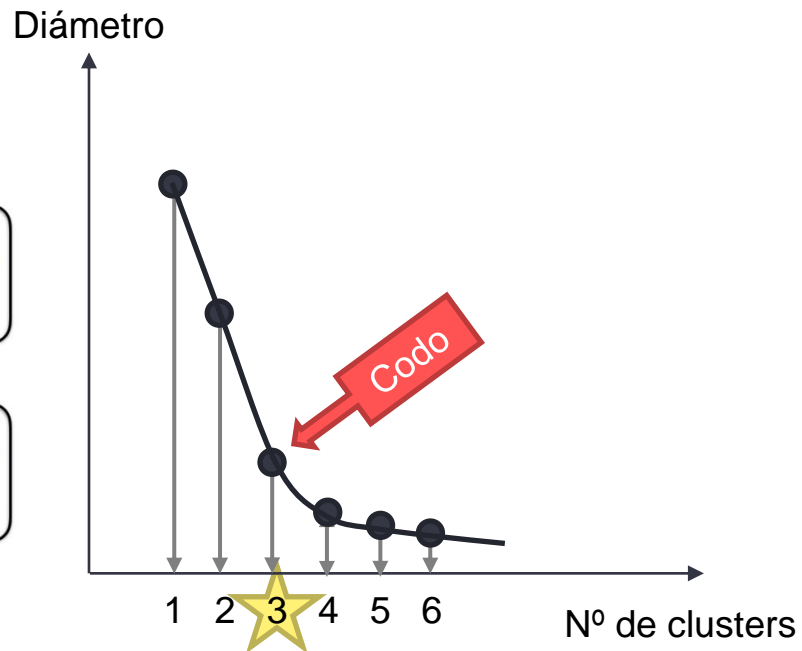
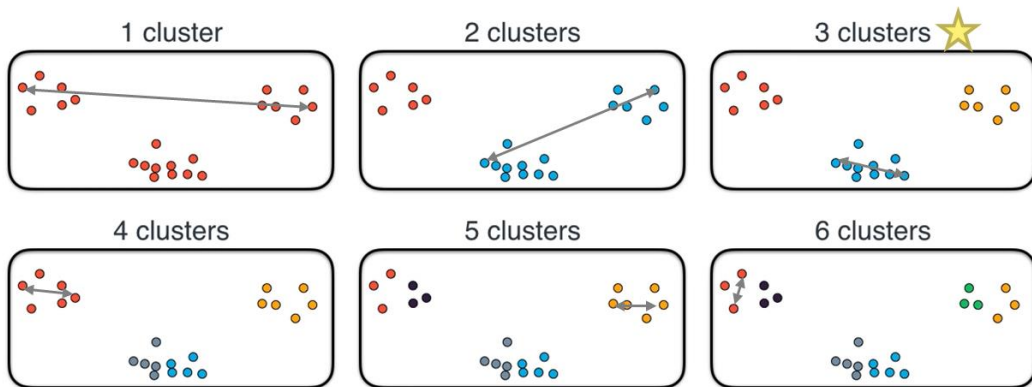
- ¿Cuántos clusters elegimos?



- Medimos la máxima distancia entre dos puntos del mismo cluster

Método del codo

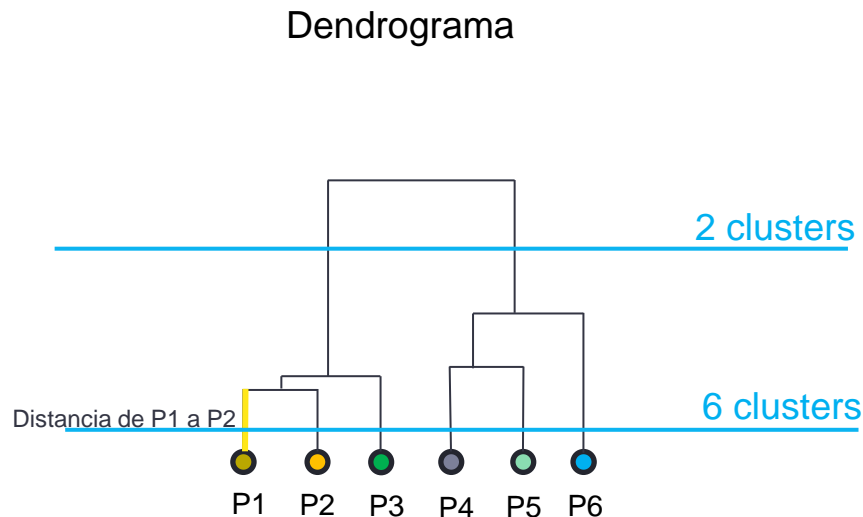
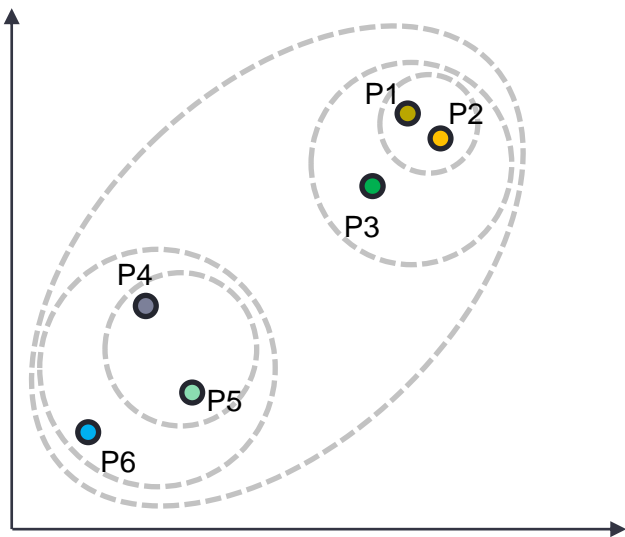
- ¿Cuántos clusters elegimos?



Clustering K-means

- En R, podemos utilizar la función `kmeans`
- Es importante normalizar los datos antes de aplicar el algoritmo

Clustering jerárquico

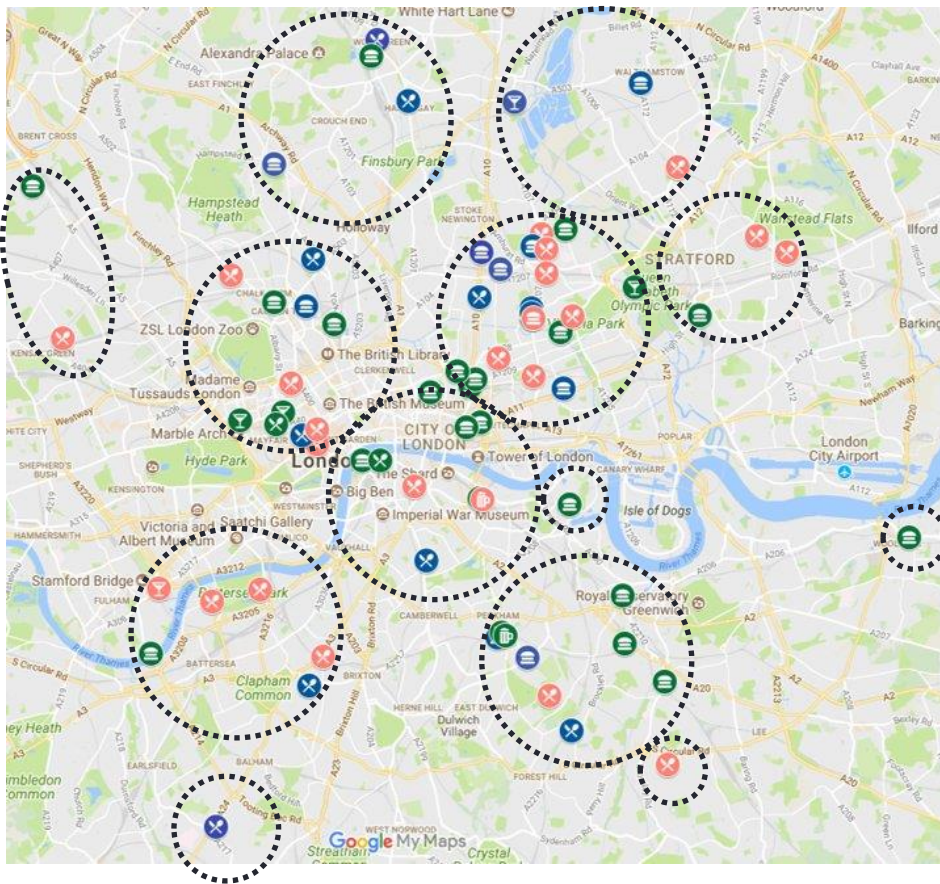


Representa todos los posibles clusters
que puede haber en los datos

Al algoritmo podemos especificarle el número de clusters que queremos o la distancia máxima para agrupar conjuntos

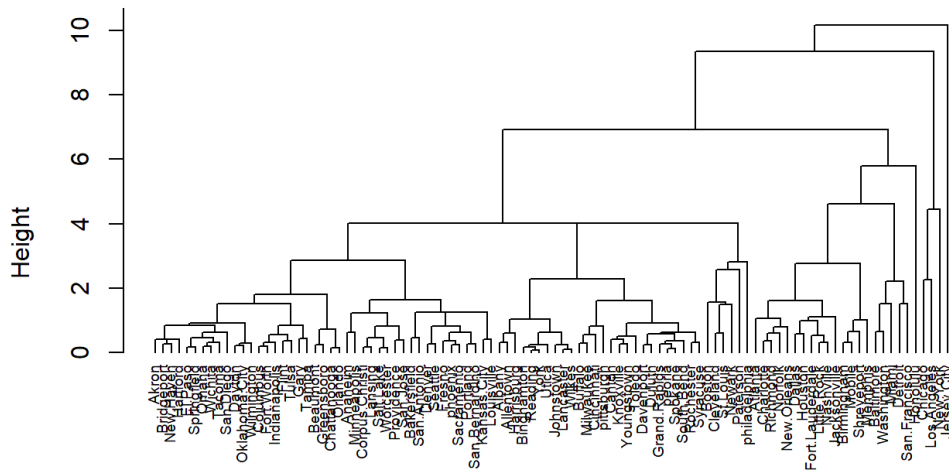
Clustering jerárquico

Ejemplo: Agrupar
restaurantes en clusters con
distancia entre ellos inferior a
1Km



Clustering jerárquico

- En R, existen varias funciones para clustering jerárquico: `hclust`, `agnes`, `diana`...
- Es importante normalizar los datos antes de aplicar el algoritmo



¡Gracias!

Contacto: Rafael Zambrano

rafazamb@gmail.com