Clustering en R – parte II

Ejercicio 1: estudio sobre PCA (*Principal Component Analysis*)

Los datos de iris están cargados por defecto en R, basta poner el nombre en R y podrá utilizarlos. Explore brevemente este conjunto de datos y aplíquele un PCA. La función para hacerlo en R es la función: **prcomp(datos)**

- ¿Con cuántos ejes coordenados se obtiene el 90% de la variabilidad de los datos? Función *summary*
- Represente gráficamente el PCA obtenido. (Funciones plot y biplot)
- ¿Cuál es la posición respecto de los nuevos ejes coordenados de los antiguos ejes o atributos? (Vector de rotación)
- Realice un estudio de cómo afecta el escalado, o normalización de los datos, a través de un estudio de cómo afecta éste al PCA.

Ejercicio 2: caso de estudio con datos de cáncer de mama.

Realice un estudio no supervisado, de *clustering*, de unos datos de cáncer de mama suministrados mediante el siguiente enlace:
https://www.kaggle.com/uciml/breast-cancer-wisconsin-data

- Cargue los datos
- Estudie su estructura y construya la matriz numérica para su estudio. La variable nominal, que indica el diagnóstico, no se usa en el estudio de tipo no supervisado sin embargo guarde esos valores en un vector auxiliar.
- Explore los datos convenientemente.
- Aplique un clustering jerárquico.
- Aplique un clustering de tipo kMeans.
- Compare los resultados.
- Repita el proceso realizando previamente un proceso de PCA.

Ejercicio 3: caso de estudio de reglas de asociación.

Realice un estudio de reglas de asociación del fichero que se le proporciona de nombre Titanic.data, que contiene los datos de los pasajeros del famoso transatlántico Titanic que se hundió en 1914.

- Aplique el algoritmo A priori visto en clase. (Función **apriori** del paquete **arules**)
- Siga como modelo el script que se suministró en teoría con un pequeño ejemplo de juguete de la bolsa de la compra.