

1 Ejercicio Prácticas No Supervisado

Fecha límite de entrega: 11 de julio de 2018 23:59.

1.1 Ejercicio 1 (PCA + K-medias)

Consideremos el dataset **wine**, con información sobre una muestra de diversos tipos de vino. Proceda a graficar en un mismo plot los *scores* de las dos primeras componentes principales del dataset, discriminando con colores en función de un procedimiento de clustering de k-medias con $k = 3$.

Nota: es recomendable un posible primer paso de estandarización de los datos.

1.2 Ejercicio 2 (DBSCAN)

Considere el siguiente esquema de generación de datos:

1. Primer conjunto (“el planeta”), generar 100 datos en \mathbb{R}^2 con la siguiente receta:
 - (a) Generar $n = 100$ radios R_i con distribución uniforme $U(0, 1)$.
 - (b) Generar $n = 100$ ángulos θ_i con distribución uniforme $U(0, 2\pi)$.
 - (c) En base a eso, definir $X_i \in \mathbb{R}^2$ como $X_i = (R_i \cos(\theta_i), R_i \sin(\theta_i))$.
2. Segundo conjunto (“la órbita”), generar 100 datos en \mathbb{R}^2 con la siguiente receta:
 - (a) Generar $n = 100$ radios R_i con distribución uniforme $U(1.5, 2.5)$.
 - (b) Generar $n = 100$ ángulos θ_i con distribución uniforme $U(0, 2\pi)$.
 - (c) En base a eso, definir $X_i \in \mathbb{R}^2$ como $X_i = (R_i \cos(\theta_i), R_i \sin(\theta_i))$.

Plotear en un mismo gráfico ambos datasets. Implemente un esquema de DBSCAN, con parámetros bien elegidos, para poder “separar” los dos conjuntos (“planeta” y “órbita”) en dos clusters.