

# Práctica Clustering

## Introducción al análisis de datos con R A2B2C

### Clustering

Esta práctica tiene como objetivo trabajar distintos métodos de clustering en distintos tipos de datos.

### Iris

El dataset Iris viene precargado en R en un objeto llamado iris (simplemente con poner iris R lo reconoce como cualquier otro dataframe).

(Recordar escalar y centrar los datos)

- a) Observar el dataset iris y notar qué variables posee, cuántas observaciones, contar cantidad de observaciones de cada especie, etc.
- b) Realizar un PCA únicamente con las variables numéricas, sin incluir la especie. ¿Se observan grupos a simple vista? ¿Cuántos?  
Pintar los puntos de acuerdo a la variable 'species'. ¿Coinciden las especies con los grupos?
- c) Usando kmeans, clusterizar los datos para distintos k. Usar tanto el criterio del codo como el criterio de silhouette para elegir un k adecuado.  
¿Los clusters encontrados son compactos?  
¿Todos los puntos de un cluster pertenecen a una misma especie siempre? ¿Qué característica tienen los puntos que quedaron con otras especies?
- d) Usando clustering jerárquico, agrupar nuevamente los datos de iris. Decidir una medida de distancia, una de linkage y una altura para cortar. Graficar el dendrograma.  
¿Cómo son los clusters que quedan respecto a las especies?  
Probar con distintos métodos de linkage y de altura y compararlos entre si.
- e) ¿Qué método de cluster funcionó mejor para estos datos?

### Mamíferos

El dataset mamiferos.csv contiene la información de nutricional de la leche de distintos mamíferos.  
(Recordar escalar y centrar los datos)

- a) Cargar el dataset.
- b) Observar el dataset y notar qué variables posee, cuántas observaciones, etc.

- c) Realizar un PCA únicamente con las variables numéricas, sin incluir el nombre. ¿Se observan grupos a simple vista?  
Graficar los nombres de los animales junto con los puntos. ¿Se encuentra ahora algún patrón?
- d) Usando kmeans, clusterizar los datos para distintos k. Usar tanto el criterio del codo como el criterio de silhouette para elegir un k adecuado.  
¿Los clusters encontrados son compactos?  
¿Los animales que aparecen en un mismo cluster poseen características similares? Interpretar los clusters obtenidos.
- e) Usando clustering jerárquico, agrupar nuevamente los datos. Decidir una medida de distancia, una de linkage y una altura para cortar. Graficar el dendrograma.  
¿Los animales que aparecen en un mismo cluster poseen características similares? Interpretar los clusters obtenidos. Probar con distintos métodos de linkage y de altura y compararlos entre si.
- f) ¿Qué método de cluster funcionó mejor para estos datos?