

Tarea 3

Análisis Estadístico Multivariado

Emmanuel Alcalá

Solución de un problema de agrupamiento

Instrucciones

1. Descarga la base de datos [Wholesale customers Data Set](#).
2. Describe brevemente qué información contiene la base de datos.
3. Selecciona solo las variables `Fresh` y `Delicatessen`.
4. Normaliza/escala el nuevo dataset (puedes usar `scale(X)`) 4.1. ¿Qué hace, operativamente, la función `scale()`?
5. Con el dataset normalizado/escalado, obtén el número óptimo de parámetros k^* para K-means: 5.1. Usando el método *elbow* variando de 1 a 12 clusters. Recuerda que para esto debes extraer el valor `tot.withinss` de *cada* valor de k que pruebes. 5.2. Usando el estadístico Gap con la función `clusGap()`. Usa el argumento `B=150`. Una vez que haya terminado el algoritmo, recuerda que puedes obtener el número óptimo k^* con el `print(gap_model, method = XX)`, sustituye en `XX` alguno de los métodos que mencioné en clase.
6. Corre el algoritmo de K-means con el valor que tú decidas del paso 5.
7. Grafica los datos originales usando un color diferente por cluster. Recuerda que la variable `cluster` la puedes obtener de la lista que te arroja `kmeans()` en R.