

Clustering

Trabajo de evaluación

José Tomás Palma Méndez

Dept. de Ingeniería de la Información y las Comunicaciones. Universidad de Murcia
email de contacto: jtpalma@um.es

Responde a las siguientes cuestiones utilizando el fichero `echo.kNNImpute.csv`. Puede ser necesario realizar diferentes transformaciones a los datos para alcanzar resultados concluyentes. Prueba con varias transformaciones y comenta los resultados.

Ejercicio 1.

- 1.a) Aplica las técnicas de clustering aglomerativo y divisivo. En este último caso, prueba con las diferentes medidas de distancia entre clusters vistas en clase.
- 1.b) Calcula los índices siluetas para cada uno de las técnicas aplicadas en el apartado anterior, teniendo en cuenta un número amplio de posibles particiones.
- 1.c) Genera un array para cada técnica en el que se almacene el índice silueta para cada posible partición. A partir de dichos arrays genera una gráfica en la que se represente la evolución, de forma conjunta, de los índices silueta medios de todas las técnicas.
- 1.d) Elige la combinación de partición y técnica que mejores resultados proporcione. Selecciona algunos elementos de cada cluster e intenta interpretar la partición generada.

Ejercicio 2.

- 2.a) Utilizando el índice **gap**, calcula el número óptimo de clusters para las técnicas **kmedias**, **kmedoides** y **clara**.
- 2.b) Los representantes obtenidos por las técnicas **kmedoides** y **clara** ¿Permiten inferir algo sobre la interpretación de cada cluster?
- 2.c) Para cada técnica, genera un objeto que contenga el resultado de aplicar dicha técnica para el número óptimo de clusters.
- 2.d) Realiza la misma operación, pero ahora teniendo en cuenta el índice silueta. Muestra gráficamente la evolución de este índice para cada técnica en un mismo gráfico.
- 2.e) Elige la combinación de partición y técnica que mejores resultados proporcione. Selecciona algunos elementos de cada cluster e intenta interpretar la partición generada.

Ejercicio 3.

- 3.a) Crea un objeto como resultado de aplicar la técnica `DBScan()`, eligiendo antes el valor óptimo para la ϵ -distancia. Justifica la elección de dicho valor.
- 3.b) Calcula el índice silueta para la partición generada.
- 3.c) Enumera algunas instancias de cada partición e interpreta los resultados.

Ejercicio 4.

- 4.a) Genera una tabla cuyas filas representen todas las particiones generadas anteriormente (tienes que incluir todas las generadas) y en las columnas el número óptimo de clusters y algunos de los índices calculados (como mínimo el silueta, el gap y algunos de los vistos en clase.)
- 4.b) ¿Cuál crees que es la mejor partición? Justifica tu respuesta.