

Práctica 4: Agrupación y evaluación de resultados

Objetivo: El objetivo de esta práctica es introducir los conceptos básicos de agrupación (clustering) de datos y los algoritmos más importantes. Asimismo, se estudiarán los conceptos básicos de evaluación y comparación de resultados. Se tratará, por tanto, de ejecutar diferentes modelos y comparar sus resultados.

Implemente los siguientes ejercicios:

0. Ejecute los programas ejemplo facilitados junto con la práctica para familiarizarse con los conceptos de los algoritmos de clustering de scikit-learn.
1. Seleccione al menos cinco problemas de los disponibles en los repositorios usados en las prácticas anteriores. Use problemas que solo contengan atributos numéricos. No olvide eliminar la información de la clase antes de ejecutar los algoritmos.
2. Seleccione el algoritmo de clustering k -means.
3. Seleccione los algoritmos de clustering jerárquicos *single link*, *complete link* y *average link*.
4. Implemente la medida de evaluación de la calidad de un método de agrupación basada en la correlación entre la matriz de incidencia y la de proximidad. Seleccione una de las medidas de evaluación no supervisada disponibles en scikit-learn.
5. Para cada uno de los problemas seleccionados realice las siguientes tareas:
 1. Ejecute el algoritmo k -means y evalúe su rendimiento para un rango de valores alrededor del número conocido de clases.
 2. Ejecute los algoritmos jerárquicos y evalúe su rendimiento al nivel en el cual tengan el mismo número de grupos que el número de clases del problema.
6. (*) Implemente algoritmo particional *bisecting k*-means y compare sus resultados con el algoritmo k -means.
7. (*) Usando la función para generar datos de forma aleatoria (`make_blobs`) genere diversos grupos de datos y compare de forma gráfica los algoritmos disponibles en scikit de forma similar a los ejemplos proporcionados.

NOTAS:

- Los ejercicios marcados con un * son opcionales.