

Clustering

El plan

1. Lectura de datos
2. Limpieza y preprocesamiento
3. Clustering de los datos
4. Evaluación del clustering resultante.
5. Visualización y análisis de resultados.

Funciones, objetos y documentación útil

- `sklearn.cluster.KMeans`
- `sklearn.cluster.AgglomerativeClustering`
- `sklearn.metrics.silhouette_score`

Dataset de Iris

1. Cargue los datos de Iris usando sklearn.
2. Agrupe los datos utilizando K-Means y Hierarchical Clustering con $k = 2 \dots 10$.
3. Obtenga el Silhouette Index para cada caso del punto anterior.
4. Grafique los clusters resultantes utilizando la clase de cada fila como color en función de la primera y segunda componente y de la segunda y tercera.
5. Responda: ¿Que indica el Silhouette Index en cada caso? ¿Es el resultado esperable, dada la existencia de 3 clases?

Datos de cancelación de clientes de una telefónica

1. Cargue los datos de abandono de clientes usando pandas: elimine las columnas 'State','Area Code','Phone', separe la columna de "Churn?" cómo la clase.
2. Normalice los datos.
3. Agrupe los datos utilizando K-Means y Hierarchical Clustering con $k = 2 \dots 10$.
4. Obtenga el Silhouette Index para cada caso del punto anterior. Identifique el mejor puntaje y guarde las etiquetas correspondientes.
5. Utilizando la información obtenida en los ejercicios de la clase anterior de este dataset, reduzca adecuadamente la dimensionalidad usando PCA e Isomap. Repita el punto anterior y visualice el número de clusters más probable.

Problema abierto: determine características comunes de los clientes agrupados en los clusters que se identifican más claramente.