

Methods of machine learning

Exercise sheet IV

May 28th, 2025

In this exercise class we apply different clustering algorithms to the classical *wines dataset*.

1. Task.

- a) Load the data and use `whiten` function from `scipy.cluster.vq` to standardize the data.
- b) Display the dataset with `pairplot` from `seaborn`, showing only the first three features.

2. Task.

- a) Perform hierarchical clustering using the linkage methods *single*, *complete*, *centroid* and *ward*. Then, visualize the resulting clusters using `seaborn`, showing only the first three features.
- b) Display the dendograms for the linkage methods *single*, *complete*, *centroid* and *ward*.
- c) Compare the Calinski-Harabasz index for the linkage methods *single*, *complete*, *centroid* and *ward*.

3. Task.

- a) Perform k-means clustering, visualize the resulting clusters (showing only the first three features) and compare the predicted labels with the true labels.
- b) Use the elbow method to determine the optimal number of clusters for k-means clustering.

4. Task.

- a) Perform spectral clustering, visualize the resulting clusters (showing only the first three features) and compare the predicted labels with the true labels.
- b) Calculate the Calinski-Harabasz scores of spectral clustering with different numbers of clusters, and determine the optimal number of clusters based on these scores.

5. Task.

Compare the Silhouette scores and the Calinski-Harabasz scores for hierarchical clustering with the 'ward' linkage method, k-means clustering and spectral clustering.