TP Clustering

Scikit-learn is a free software machine library containing various clustering algorithms in the module cluster.

1) Use functions provided by sklearn.datasets to create a data set with 4 Gaussians tightly overlapped.

2) Apply k-means and plot the result with the centroids. Display the objective function value.

3) For k-means, the number of classes c should be known a priori. Run several k-means with c varying from 2 to 10 and measure the silhouette index. Plot the result and guess the number of classes to use.

4) Generate data sets with the function make_moons with various types of bandwidth. Apply k-means, hierarchical clustering and dbscan on them. What can be observed ?

5) The basic version of k-means is only working with hyperspherical clusters. However, some variant allows to overcome the issue. Test the spectral clustering algorithm and the meanshift method on the data set.