Assignment on Clustering

In this assignment, you will implement DBSCAN and *k*-means clustering algorithms.

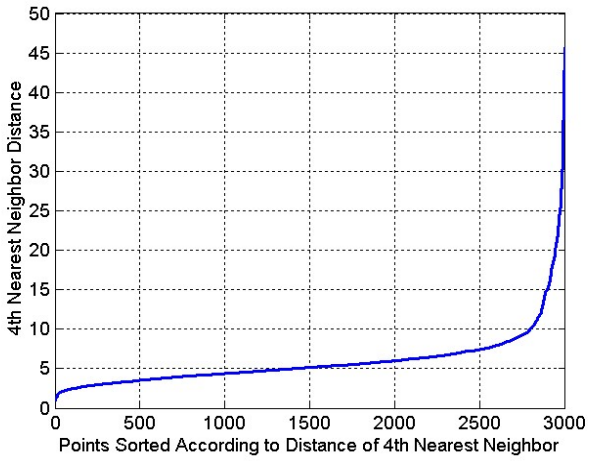
At first, you will have to implement DBSCAN. DBSCAN has two important parameters, EPS and MINPTS. These are calculated from the plot of (say,) 4-th Nearest Neighbour Distance vs Points Sorted According to 4th Nearest Neighbour. For the given dataset, you will have to make the plot and infer appropriate parameters, and use those to run DBSCAN.

You will have to code the DBSCAN with flexible parameters. For some datasets, you may have to use different EPS and MINPTS. So code rigorously. Using the parameters calculated as above, run DBSCAN.

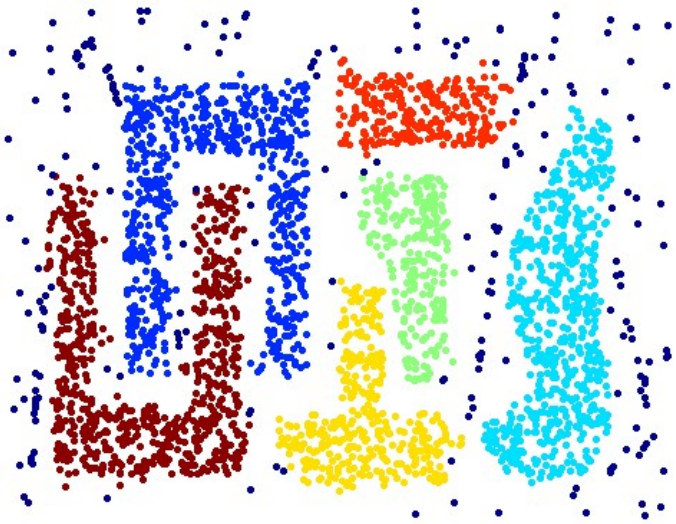
Next, you will have to implement the *k*-means clustering algorithm. For *k*-means, we will have to provide the number of clusters to the algorithm (we do not have to do this for DBSCAN). This number of clusters will be determined from the results of the DBSCAN algorithm.

So in a nutshell, you have to do the following:

1. **For each dataset provided, make a plot similar to the following figure, determine DBSCAN parameters for all the dataset.**



2. **Run DBSCAN to determine the clusters and number of clusters, show all the clusters in different colors. The following figure is an example.**



Clusters in different colors

3. **Run *k*-means. Set the value of *k* from the outcome of DBSCAN. If necessary, you can use bisecting *k*-means algorithm with proper justification provided in a separate doc file. Graphically show all the clusters in different colors.**
4. **Zip source codes, all plot graphs, figures (clustering outcomes) and upload in moodle by *11 am on Saturday, December 12, 2020*.**