MA 2233 Data Structures and Applications Lab
Team Project 5 - Clustering using Minimum Spanning Tree
11.11.2021

Clustering is the task of assigning a set of objects into groups so that the objects in the same cluster are more similar to each other than to those in other clusters.

Given a set of N data points, a minimum spanning tree is a spanning tree that connects all the data points either by a direct edge or by a path and has minimum total weight. The total weight is the sum of the weights of all the edges of the spanning tree.

In MST based clustering, the weight for each edge is usually considered as the distance between the end points forming the edge. Removal of the longest edge results in 2 clusters. Removal of the next longest edge results in 3 clusters, and so on.

Different termination criteria can be used, for instance,

1) Number of pre-determined clusters
2) Equal-density clusters
3) Thresholded by average distance within clusters
4) Setting a minimum threshold below which edges will not be removed to make new clusters.
5) Thresholding based on the ratio between the intra-to-inter cluster distances.

To do:

- Generate N points in 2-D that are from k number of clusters.
- Find the MST of the above dataset using the Euclidean distance.
- Implement the above algorithm for any 3 of the above 5 termination criteria.
- Implement a menu-driven MST-based clustering where a user can choose from the 3 termination criteria you have implemented.