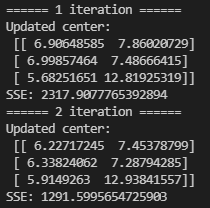
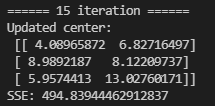
MSBD5002 Assignment 4 Report

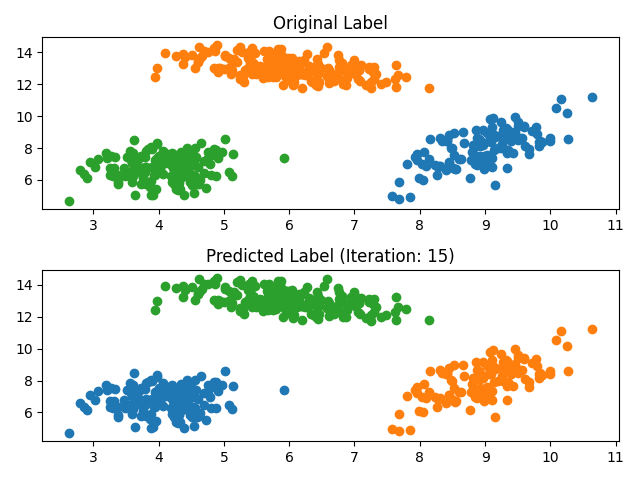
Fuzzy Clustering

SSE and the center for the first two iteration. No hyperparameters has been set for SSE calculation. Initial center was set by the first N (N be number of cluster) input data.

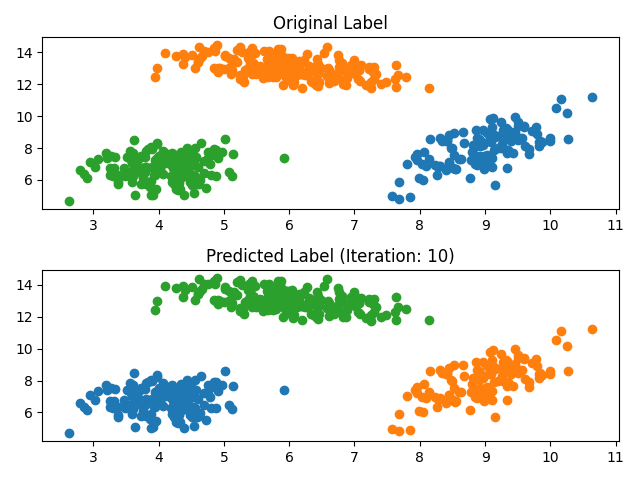
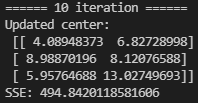


Final center coverage



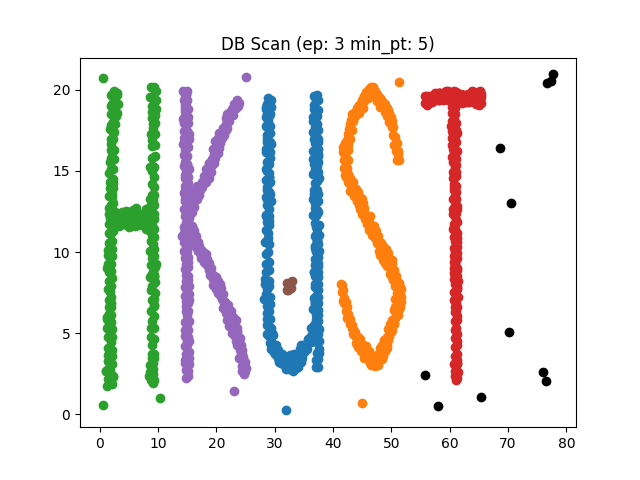


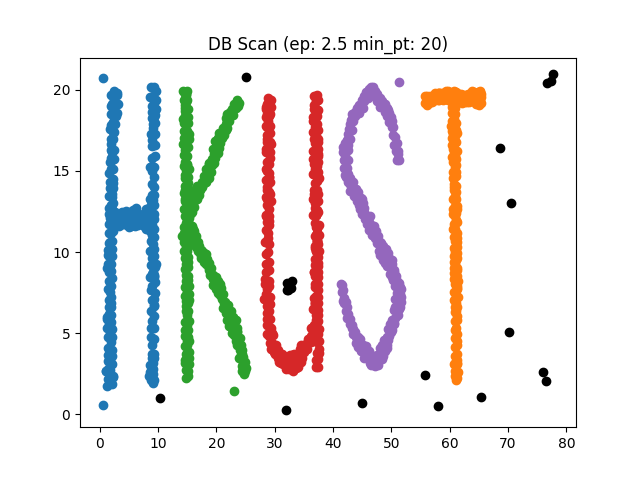
The result is the same with the original label. I have taken 15 iteration for moving, but for the small change in SSE from 10 iteration. As we can see in the graph below, the result is still the same. But we can see for the SSE, and the center, it still have modification for a more precise classification.



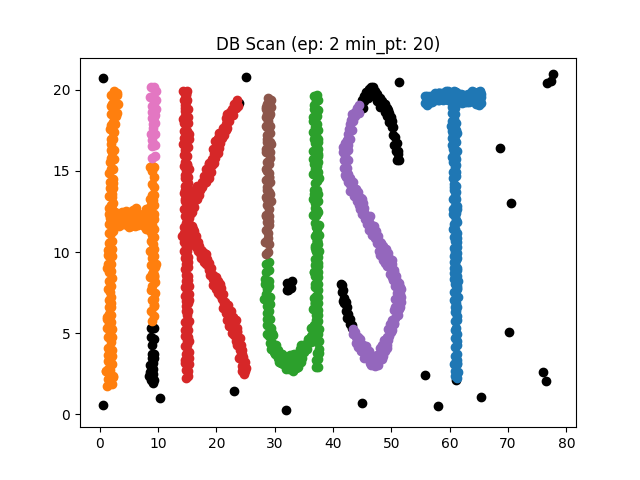
DBScan

Black will be the outlier points. The parameters are written in the title. All the result list in below.

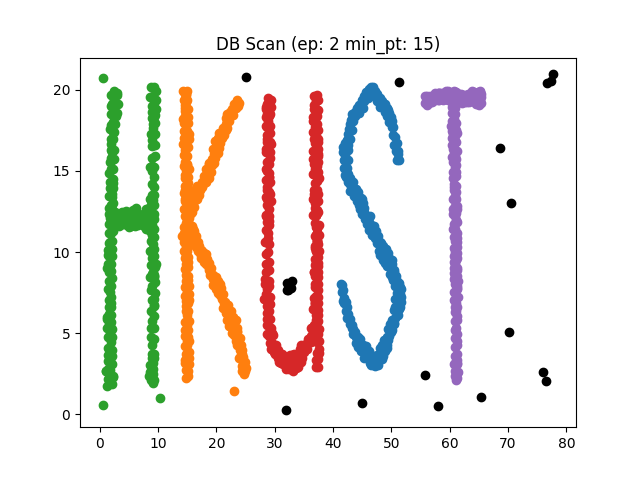






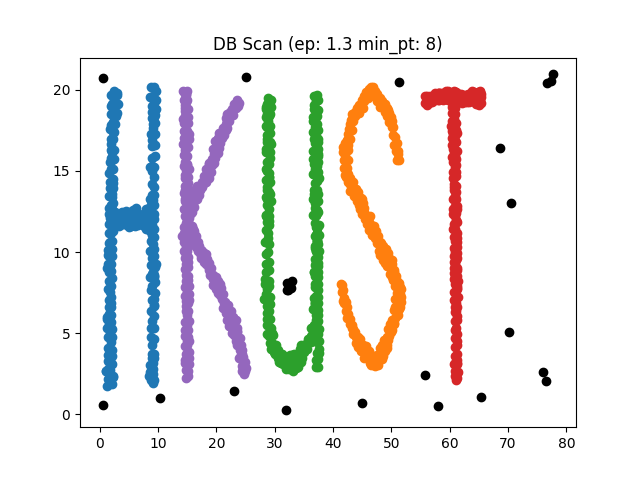








The best result is the below, as all the points do not contain in UST be the outlier. Epsilon 1.3, with minimum point 8.

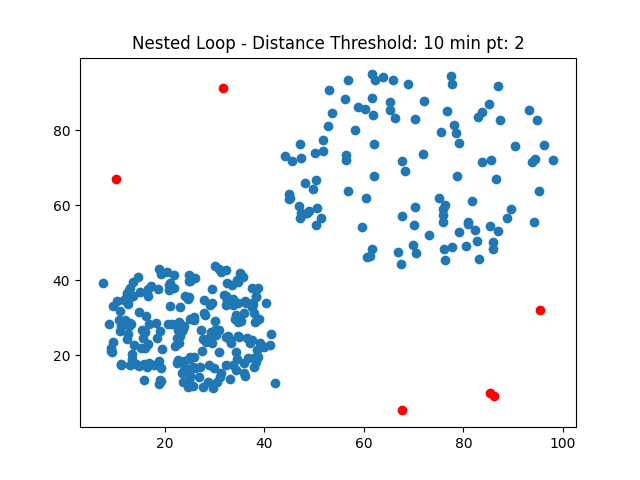




To conclude that, as most of the above setting (list in the specification), outlier point in H and K has been clustered into the cluster, therefore epsilon has been decreased such that it although it is near, it will not belongs to cluster. And we can see large min\_pt will affect original cluster (see ep:2, min\_pt: 20), and with min\_pt 5 has a better result without affecting original cluster. Therefore, min\_pt has been kept as small as possible. Therefore, setting ep:1.3, and min\_pt: 8 comes after a numerous round of testing.

Nested Loop Outlier Detection

There is two parameter, distance threshold, and minimum numbers of neighbors of the point should have. As we have to count the numbers of points within the specified distance, such that the data does not fulfill the constraint will treat as outlier. It set to 10, and 2 respectively.



Numbers of outliers and its coordinates as follow.

