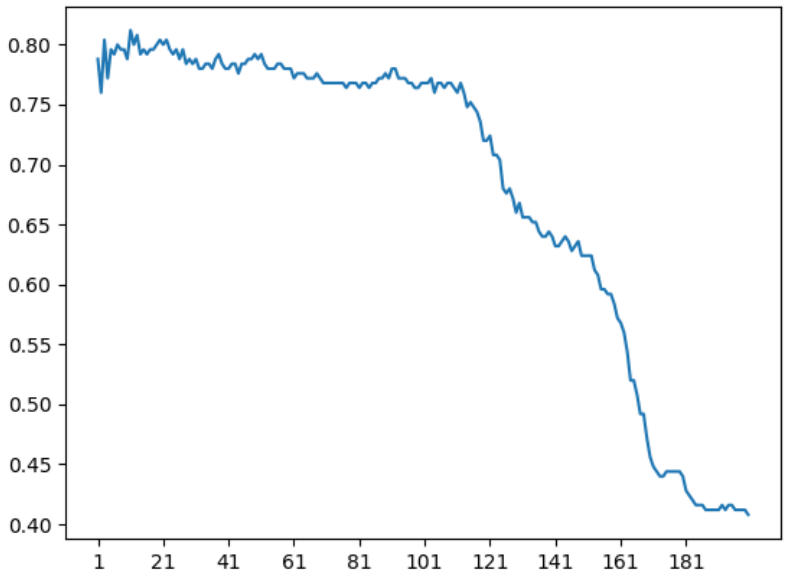
Report

Burak Akgül

18.05.2021

1 Part 1: K-Nearest Neighbor

1.1 K-fold Cross-validation



1.2 Accuracy drops with very large k values

Consider the extreme case where k is the whole data. Then we would get only the majority voted class for all our newly added data points. At very large k's, points that are added are started to be assigned by considering distant data too, so that reduces accuracy.

1.3 Accuracy on test set with the best k

Best k value is 11. The accuracy that I had was 84% on the test set.

2 Part 2: K-means Clustering

2.1 Elbow method

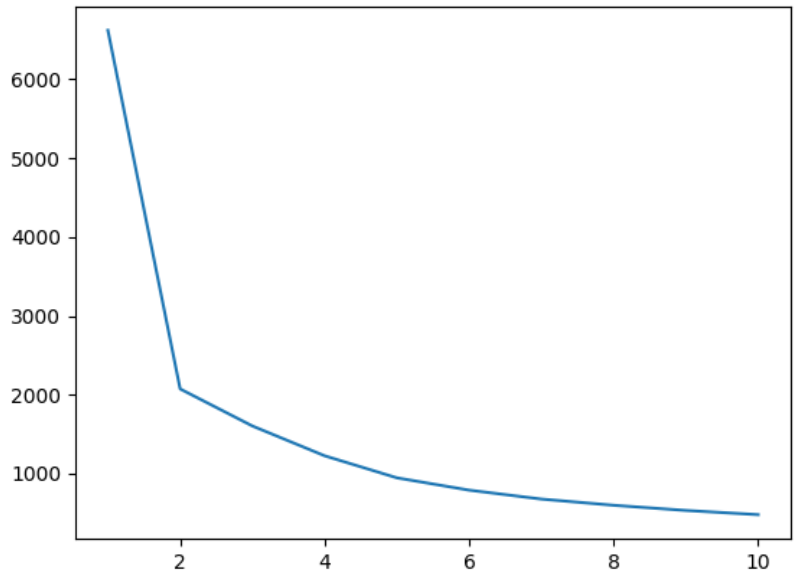


Figure 1: Data1 - k vs objective function graph

From the elbow method, k must be 2

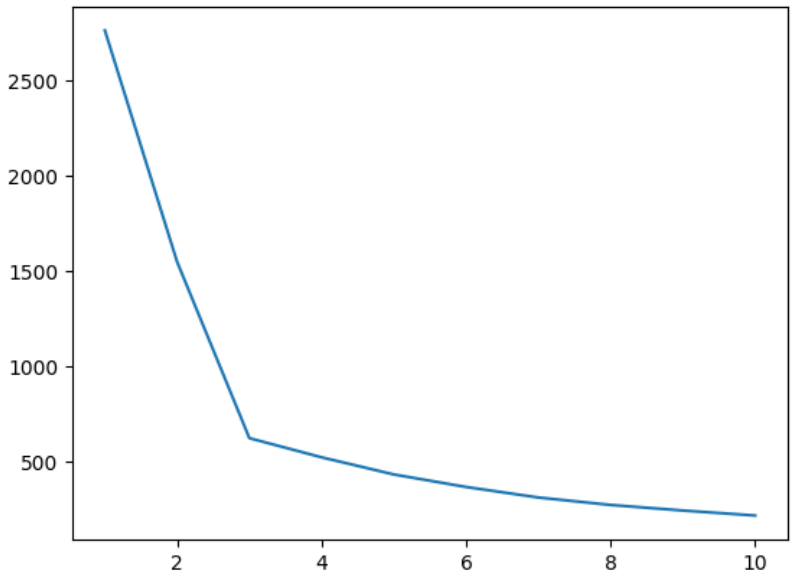


Figure 2: Data2 - k vs objective function graph

From the elbow method, k must be 3

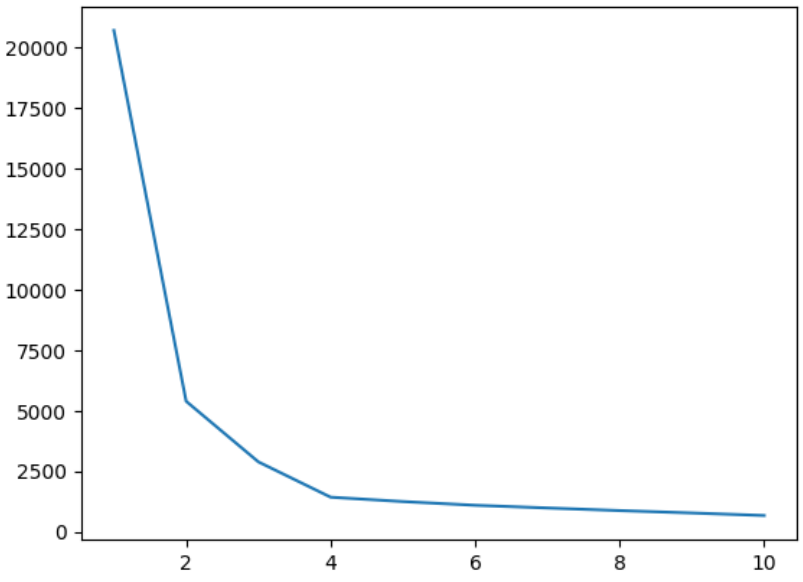


Figure 3: Data3 - k vs objective function graph

From the elbow method, k must be 4

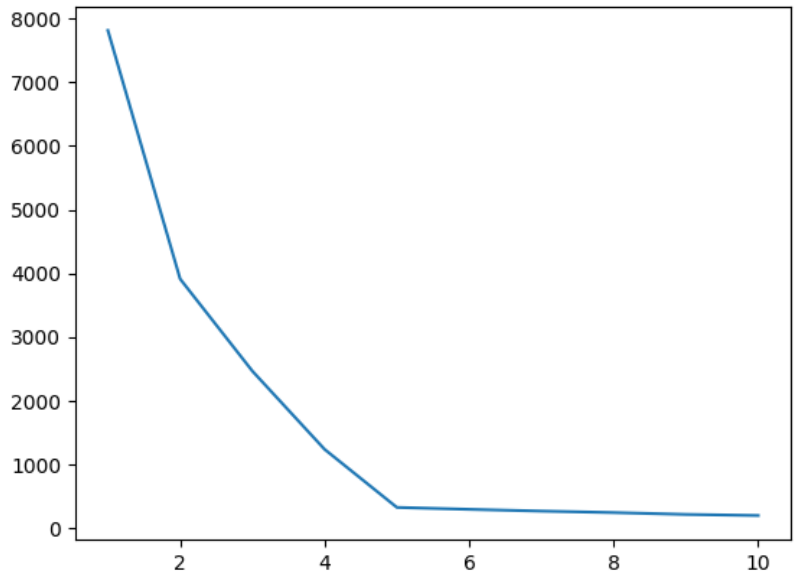


Figure 4: Data4 - k vs objective function graph

From the elbow method, k must be 5

2.2 Resultant Clusters

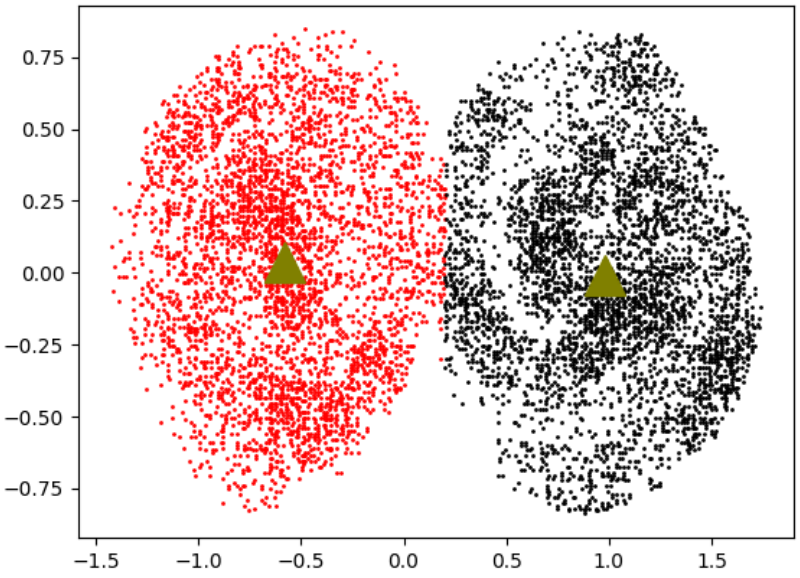


Figure 5: Data1 with k=2

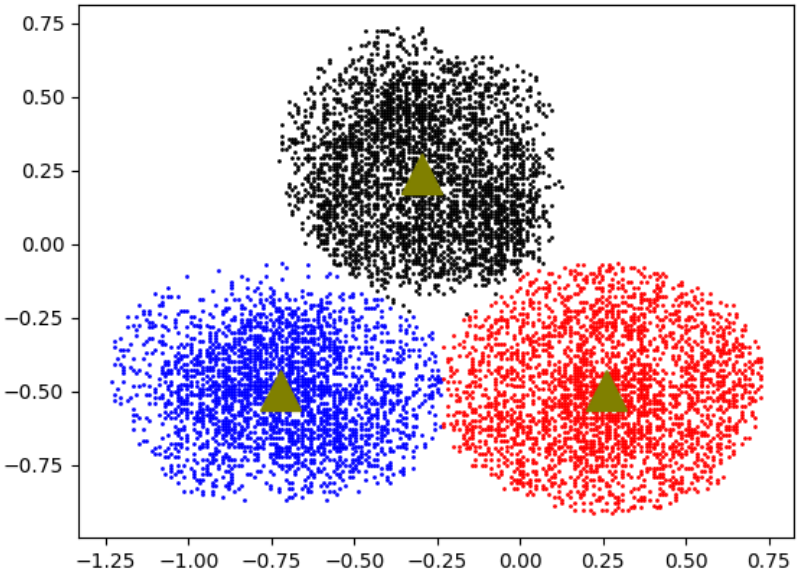


Figure 6: Data2 with k=3

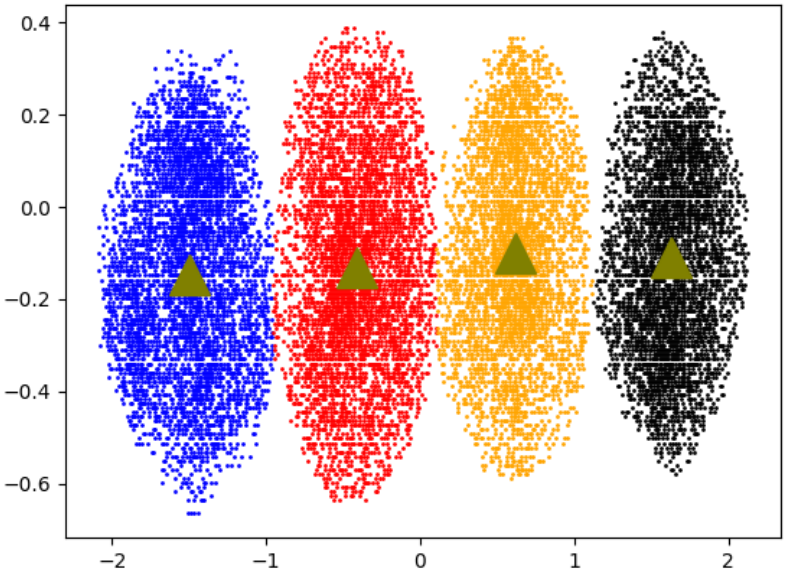


Figure 7: Data3 with k=4

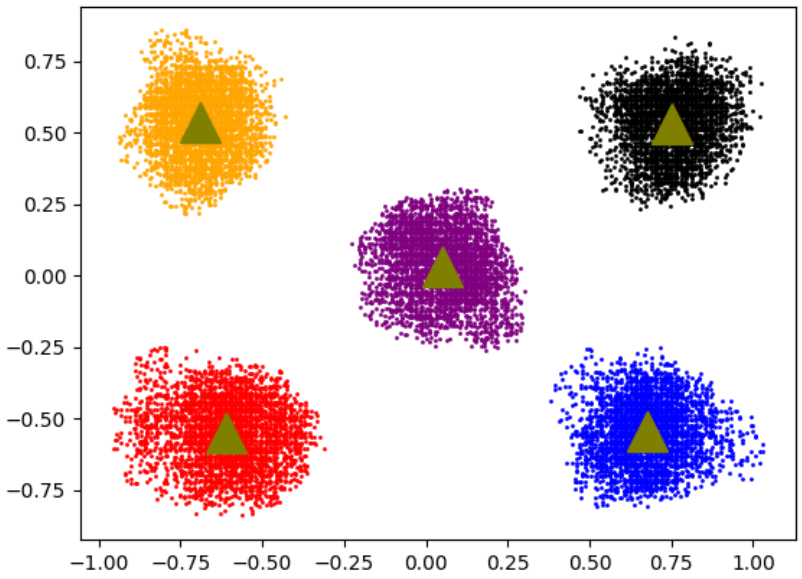


Figure 8: Data4 with k=5

3 Part 3: Hierarchical Agglomerative Clustering

3.1 data1

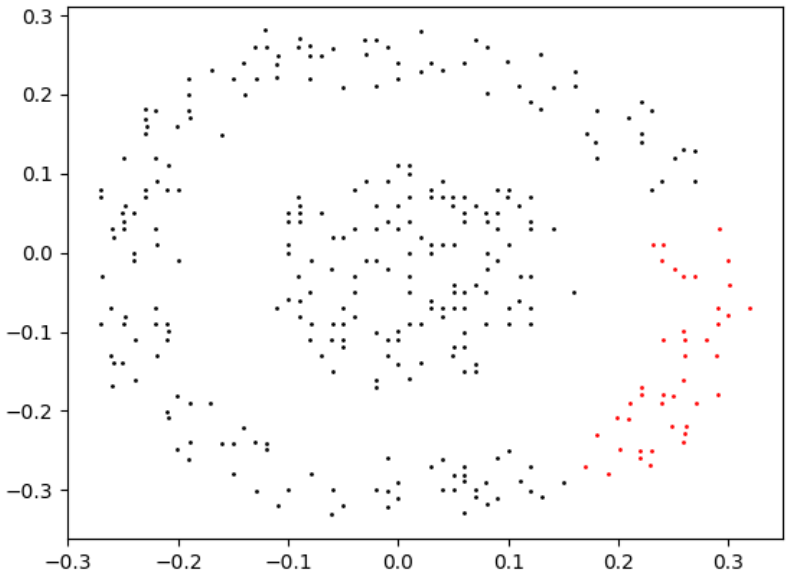


Figure 9: Data1 with average linkage

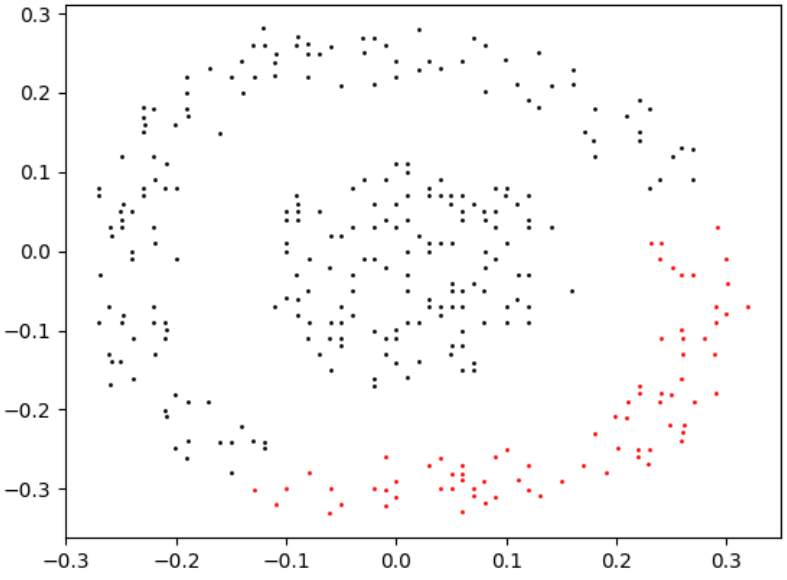


Figure 10: Data1 with centroid linkage

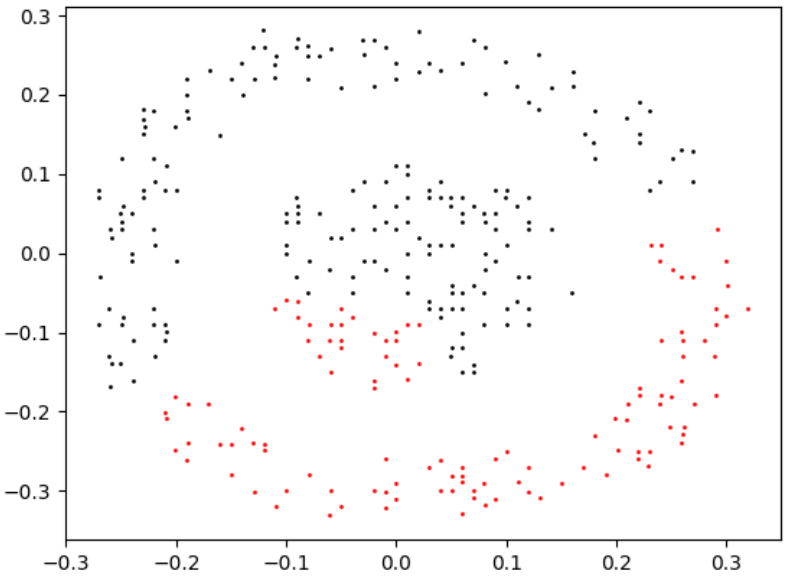


Figure 11: Data1 with complete linkage

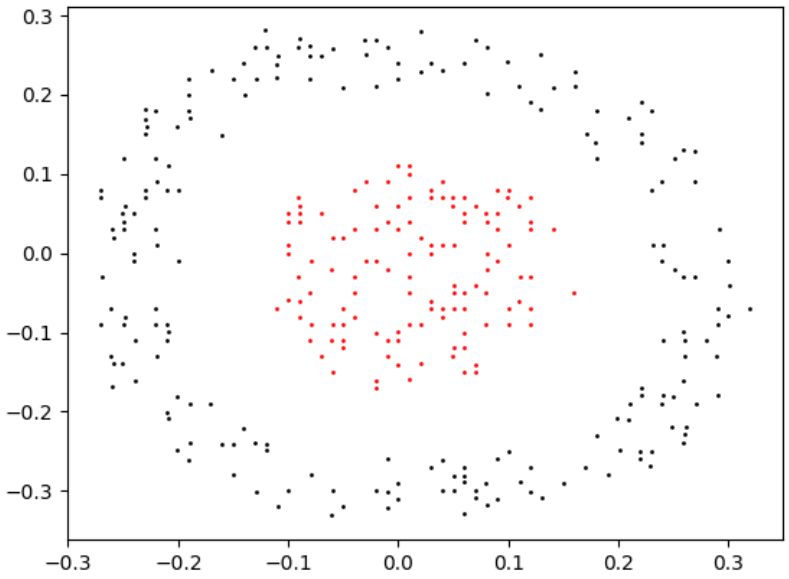


Figure 12: Data1 with single linkage

It only works correctly for single linkage because the data outside the center cluster encloses the inner cluster, so we need to connect the immediate near point by using single linkage.

3.2 data2

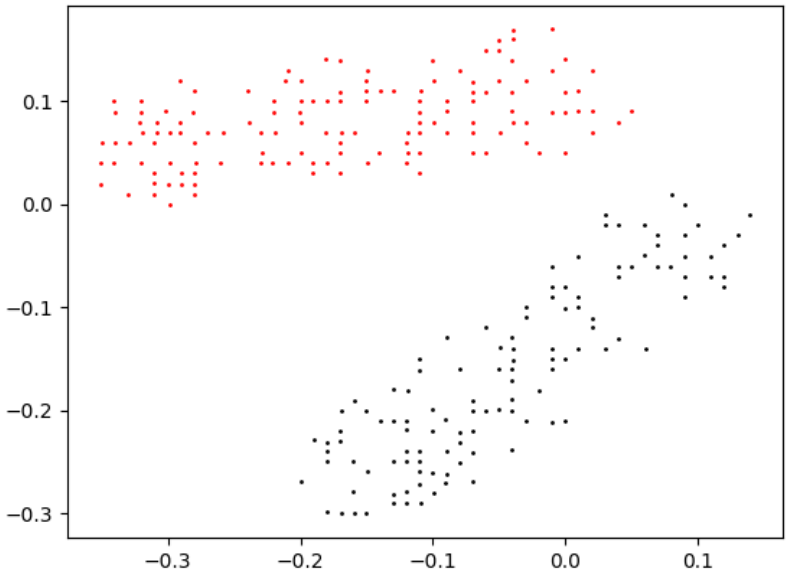


Figure 13: Data2 with average linkage

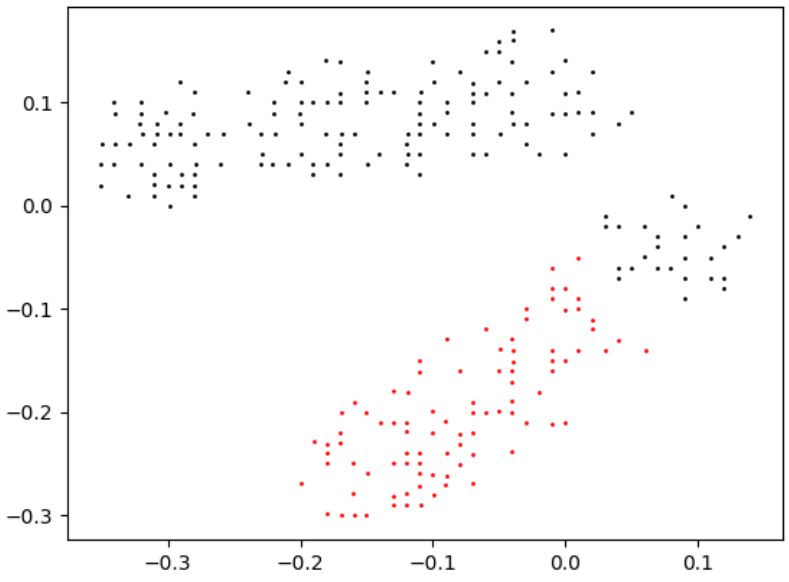


Figure 14: Data2 with centroid linkage

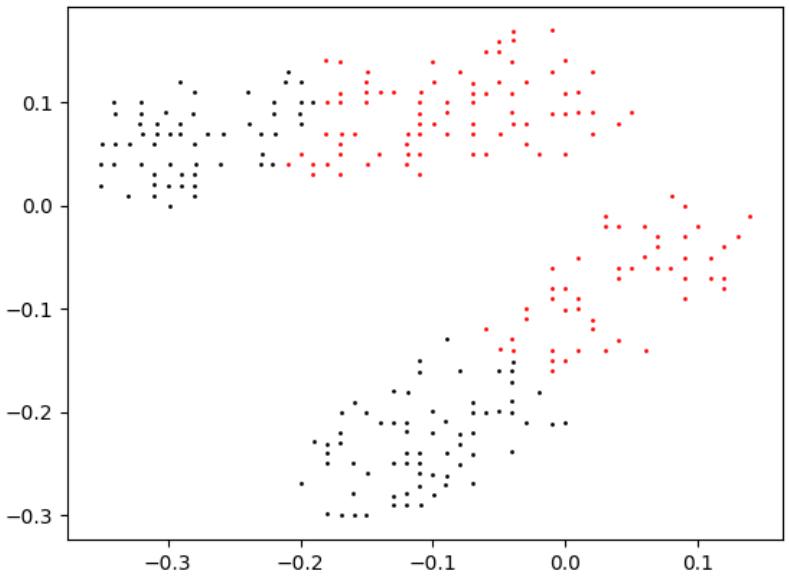


Figure 15: Data2 with complete linkage

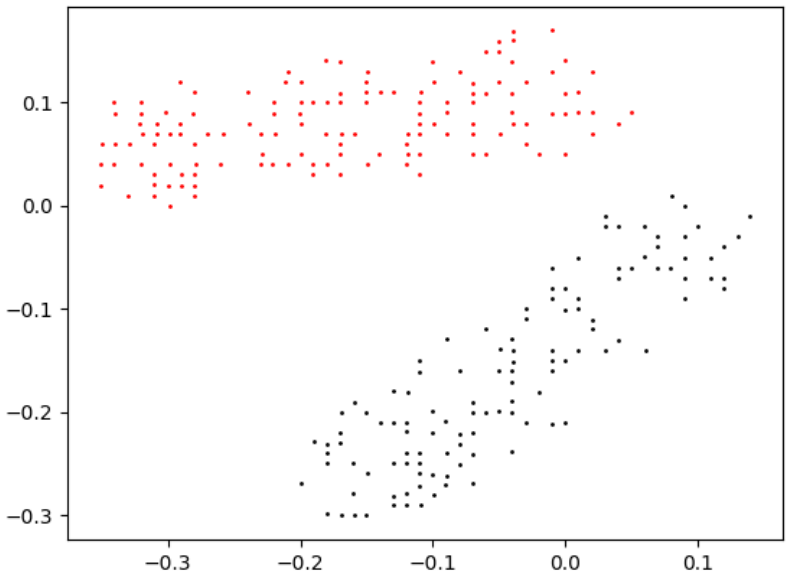


Figure 16: Data2 with single linkage

Single linkage and average linkage works because the clusters are far enough for them and connected, complete linkage and centroid linkage do not work because the points start to get closer to each other at some point.

3.3 data3

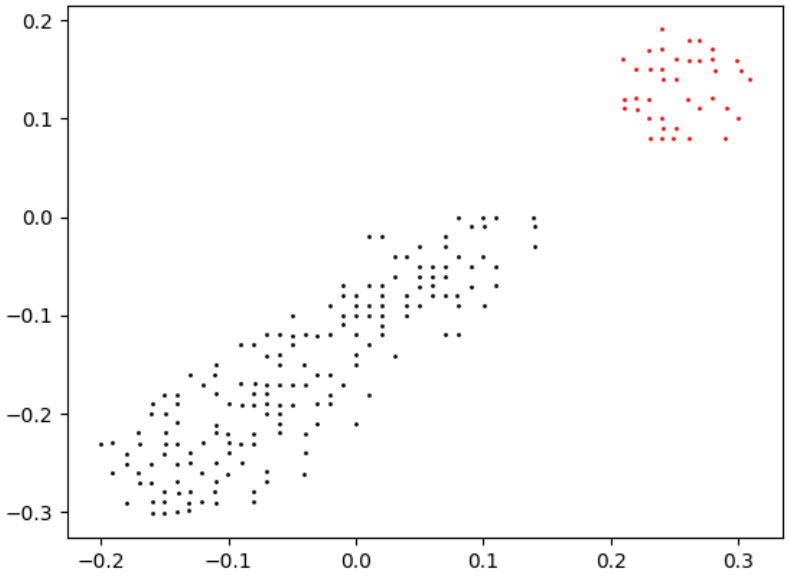


Figure 17: Data3 with average linkage

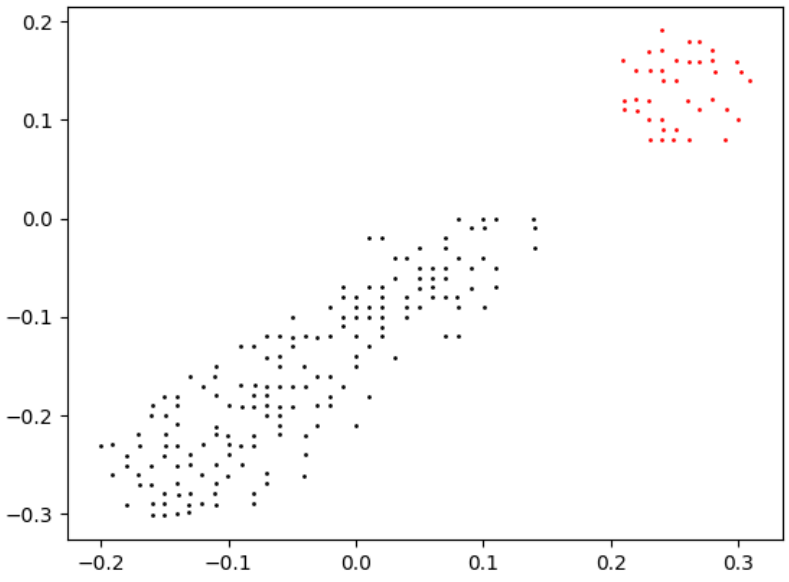


Figure 18: Data3 with centroid linkage

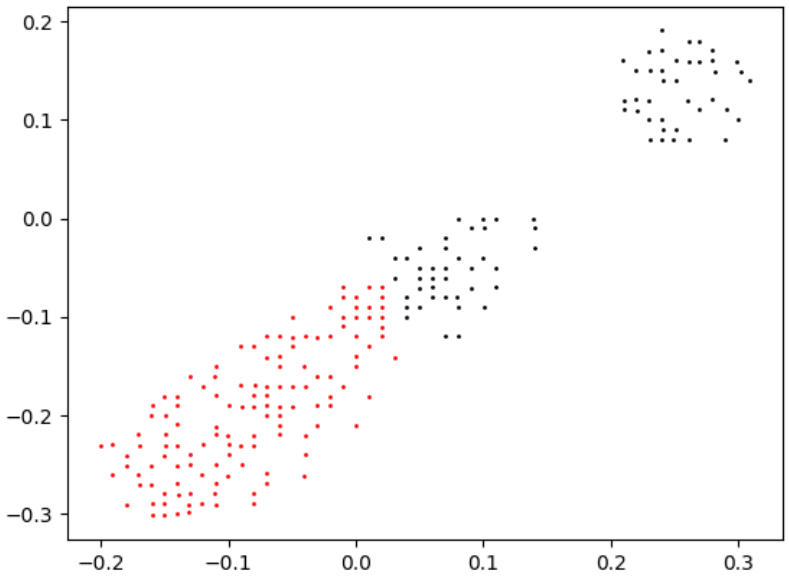


Figure 19: Data3 with complete linkage

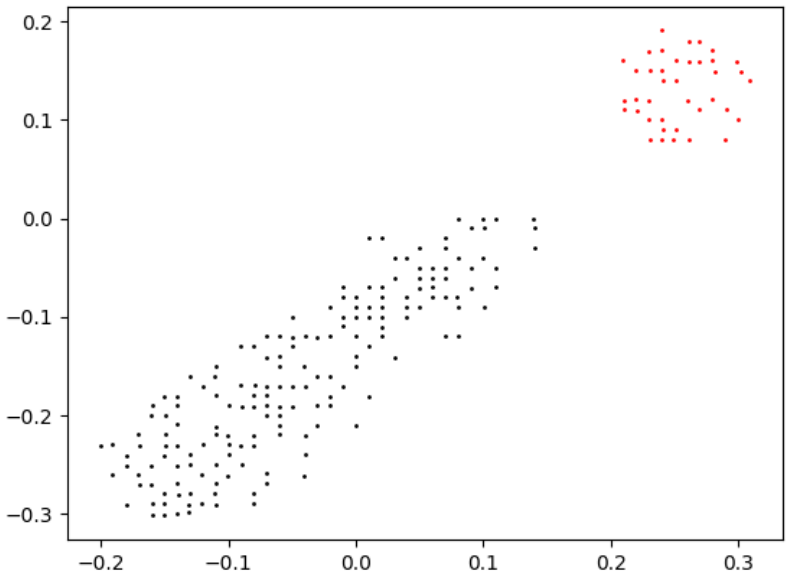


Figure 20: Data3 with single linkage

Complete linkage does not work because when calculation distance, we look for the longest distance of points inside each clusters, and the cluster's center that is far top-right is closer to the middle portion of the points than the cluster on the far bottom-left side.

3.4 data4

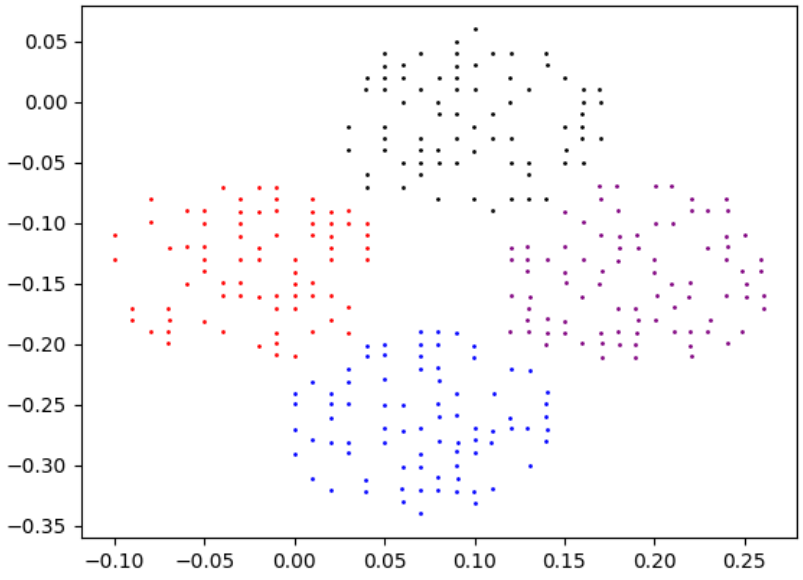


Figure 21: Data4 with average linkage

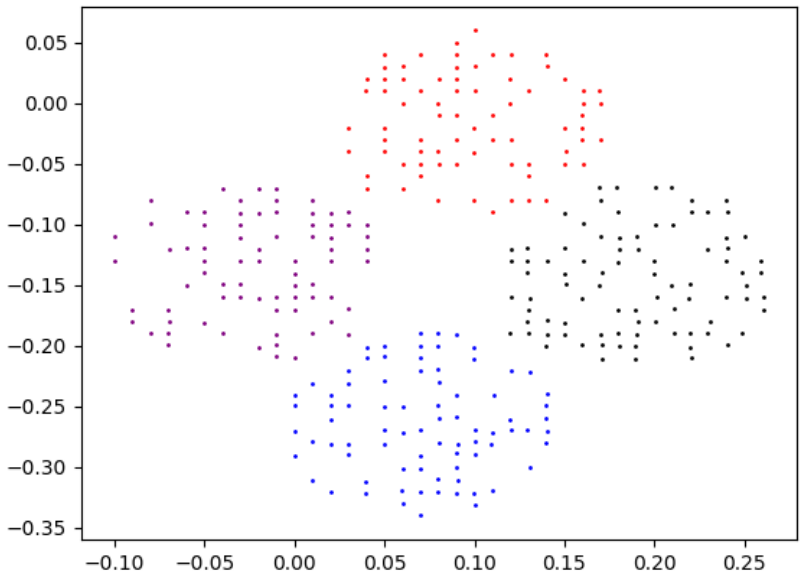


Figure 22: Data4 with centroid linkage

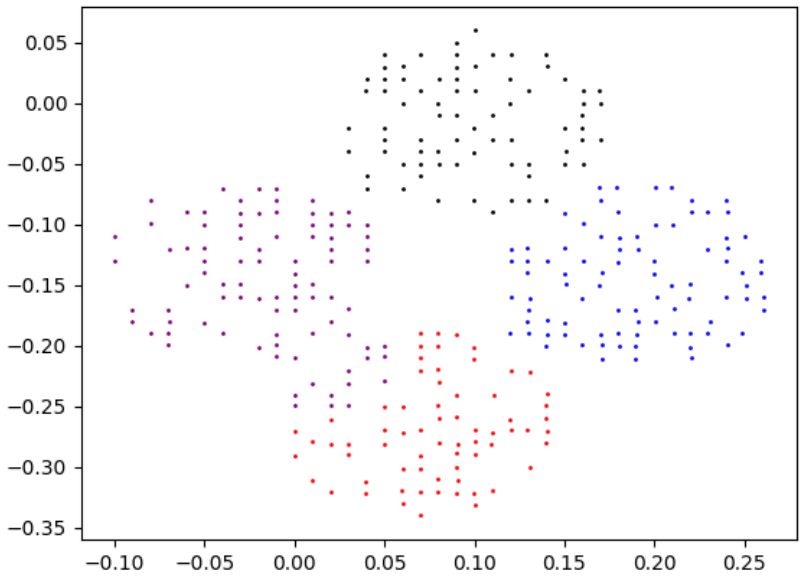


Figure 23: Data4 with complete linkage

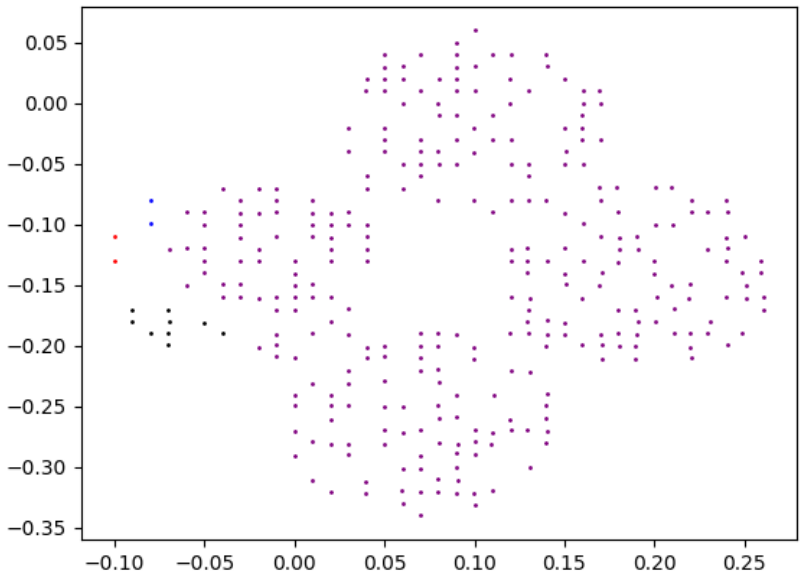


Figure 24: Data4 with single linkage

Single linkage does not work because all the points are nearly clustered together and that makes 1 whole cluster, complete linkage partially does not work because of the similar reason to Data3: Red cluster's farthest point is farther than the farthest distance from purple cluster.