Ceng499 THE2 Report

Seda Civelek - 2237147

May 2021

1 Part 1: K-Nearest Neighbor

1.1 K-fold Cross-validation

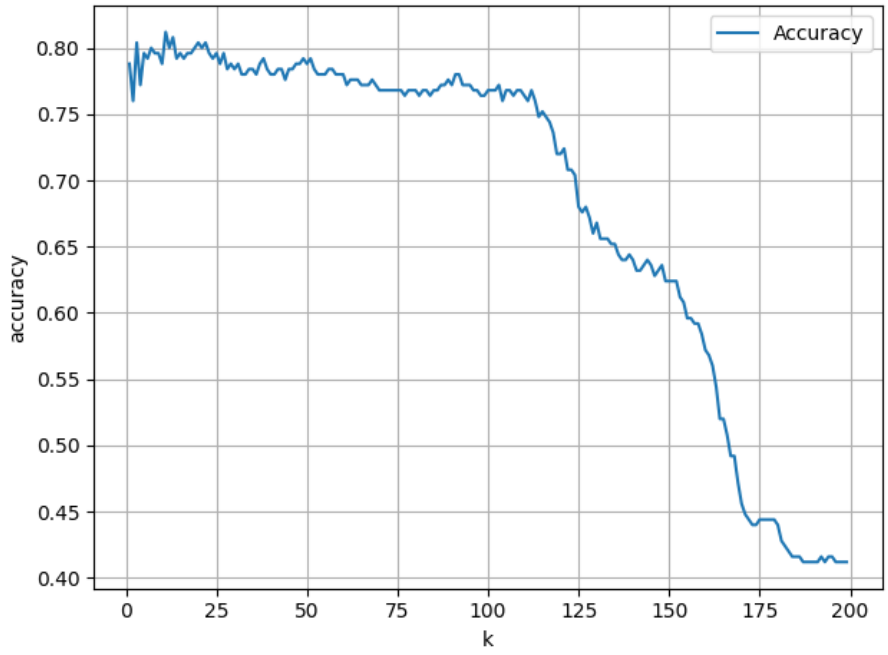


Figure 1: K-fold cross validation

This plot represents the average accuracies for each $k_{knn} = 1, 2, 3, ..., 199$ values from cross validation using training data only.

1.2 Accuracy drops with very large k values

If we choose k as large values, the k neighbors will be choosed more globally. So, the prediction will be made same for every instance. Thus, the accuracy will be dropped dramatically.

1.3 Accuracy on test set with the best k

The best k value is **11** since the best accuracy that is **0.82** is obtained with $k_{knn} = 11$.

2 Part 2: K-means Clustering

2.1 Elbow method

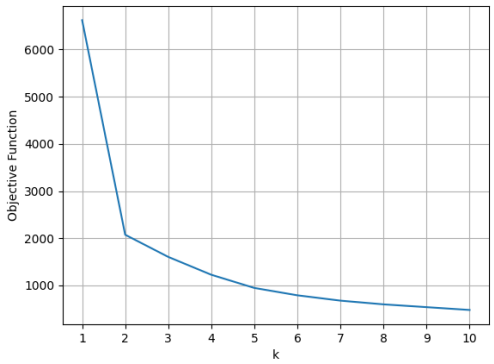


Figure 2: Elbow method for clustering data1
Best suitable k value is 2.

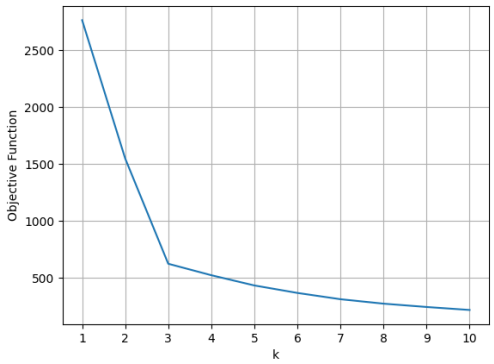


Figure 3: Elbow method for clustering data2
Best suitable k value is 3.

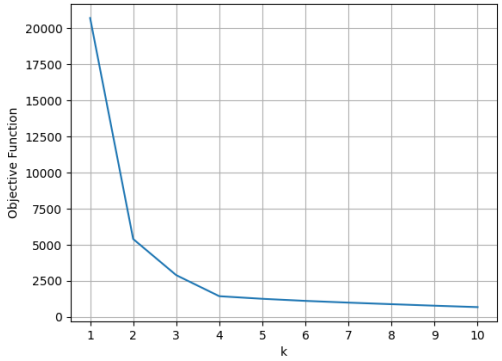


Figure 4: Elbow method for clustering data3
Best suitable k value is 4.

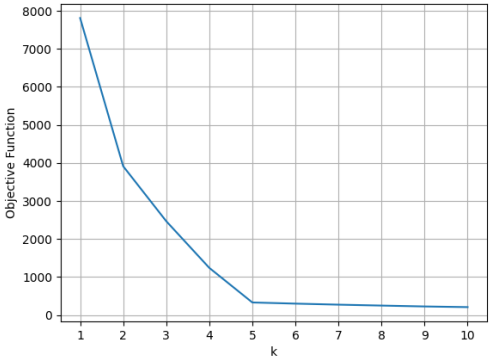


Figure 5: Elbow method for clustering data4
Best suitable k value is 5.

2.2 Resultant Clusters

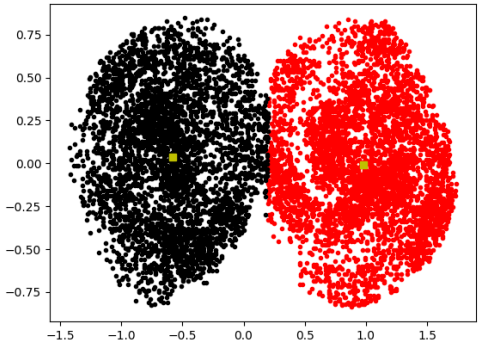


Figure 6: Resultant clusters for data1 with k=2

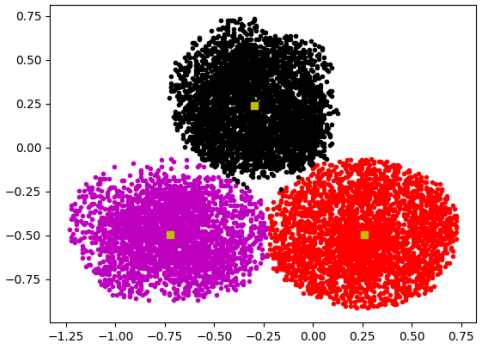


Figure 7: Resultant clusters for data2 with k=3

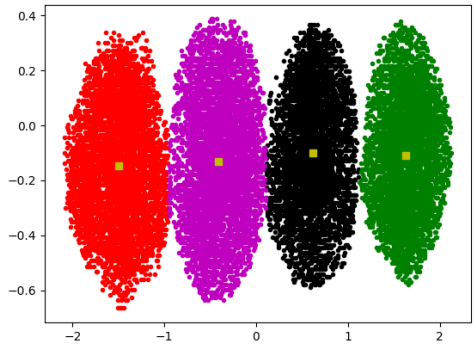


Figure 8: Resultant clusters for data3 with k=4

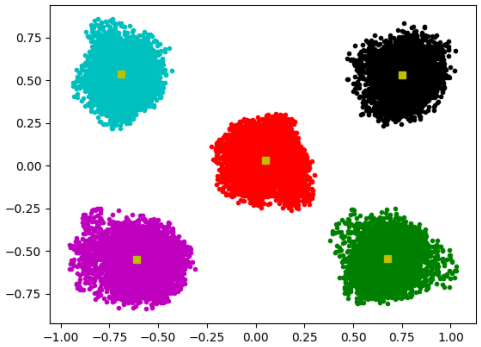
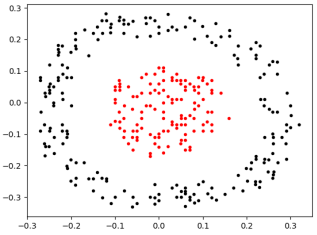


Figure 9: Resultant clusters for data4 with k=5

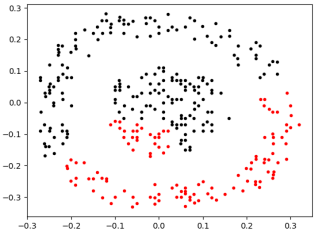
3 Part 3: Hierarchical Agglomerative Clustering

3.1 data1

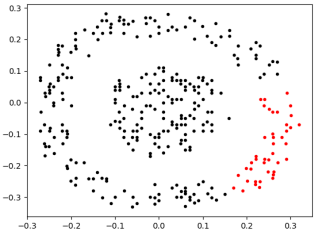
Single Linkage: It gives the best result since the nearest points in two clusters that are the round shape in the center and ring are far enough to distinguish.
Complete Linkage: It doesn't give a correct result since some points in the round shaped cluster in the center are closer to the points in ring shaped cluster than the points in cluster that it belongs.
Average Linkage: It doesn't give a correct result due to the same reason explained in Complete Linkage.
Centroid Linkage: It doesn't give a correct result since the center point of the round shaped cluster and the center point of the ring shaped cluster are so close.



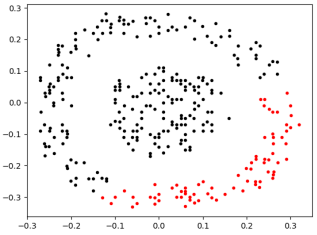
(a) Single linkage



(b) Complete linkage



(c) Average Linkage

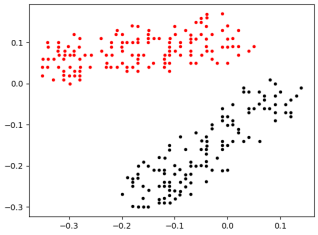


(d) Centroid Linkage

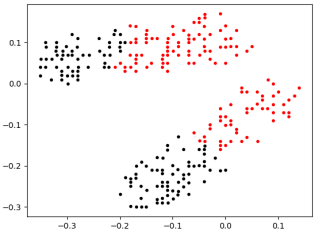
Figure 10: Plots of clustered data1 for each criterion

3.2 data2

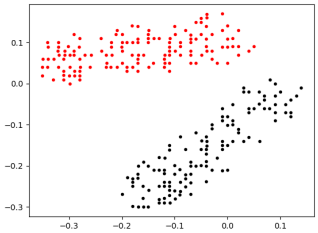
Single Linkage: It gives one of the best results since the nearest points in two clusters are far enough to distinguish.
Complete Linkage: It doesn't give the correct result since some points in the cluster above are not far enough to points in the cluster below.
Average Linkage: It gives one of the best results since cluster boundaries of the one above and the one below are seperated from each other.
Centroid Linkage: It doesn't give the correct result since black points in the cluster below are closer to the center of the cluster above than the center of the cluster below.



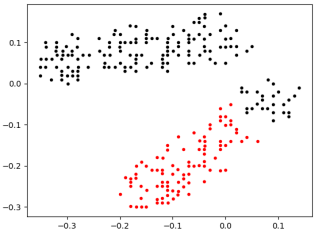
(a) Single linkage



(b) Complete linkage



(c) Average Linkage

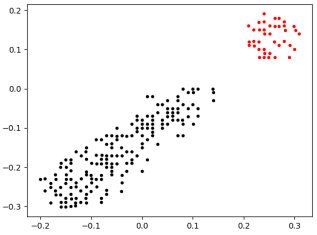


(d) Centroid Linkage

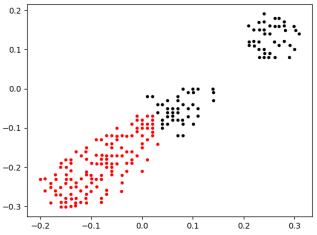
Figure 11: Plots of clustered data2 for each criterion

3.3 data3

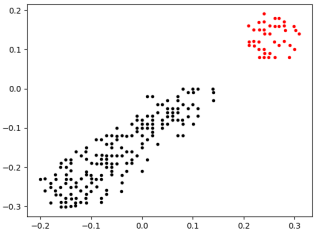
Single Linkage: It gives one of the best results since the nearest points in two clusters are far enough to distinguish.
Complete Linkage: It doesn't give the correct result since the black points in the big clusters are closer to the points in round shaped cluster than the red points in below.
Average Linkage: It gives one of the best results since the mean points of the two clusters are separable enough.
Centroid Linkage: It gives one of the best results since the center points of the two clusters are far from each other.



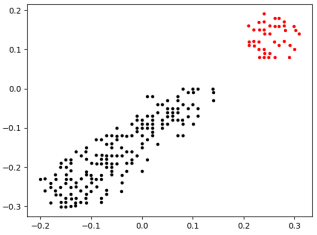
(a) Single linkage



(b) Complete linkage



(c) Average Linkage

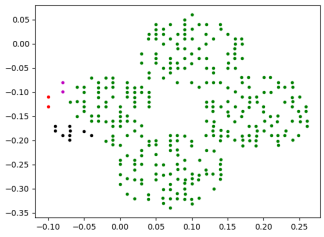


(d) Centroid Linkage

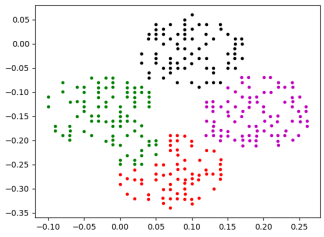
Figure 12: Plots of clustered data3 for each criterion

3.4 data4

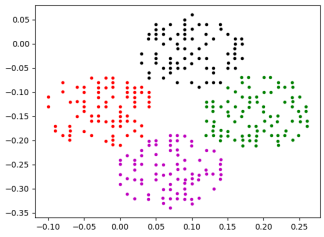
Single Linkage: It doesn't give the correct result since all points are too close to each other respectively.
Complete Linkage: It gives a better result but as you can see red and green parts are not seperated fully correct.
Average Linkage: It gives one of the best results since all clusters have a round shape.
Centroid Linkage: It gives one of the best results since all clusters have a round shape and their center point are far enough from each other and the center points are the center of the round shapes.



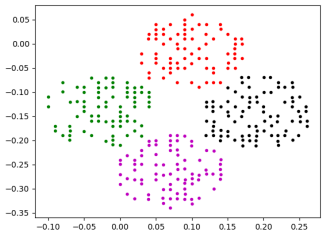
(a) Single linkage



(b) Complete linkage



(c) Average Linkage



(d) Centroid Linkage

Figure 13: Plots of clustered data4 for each criterion