CENG499 HW-2
Report

Ufuk Cem Birbiri
e2171379

23/12/2020

1 Part 1: K-Nearest Neighbor

1.1 K-fold Cross-validation

Figure 1 shows the 10-fold cross validation KNN results with different k values.

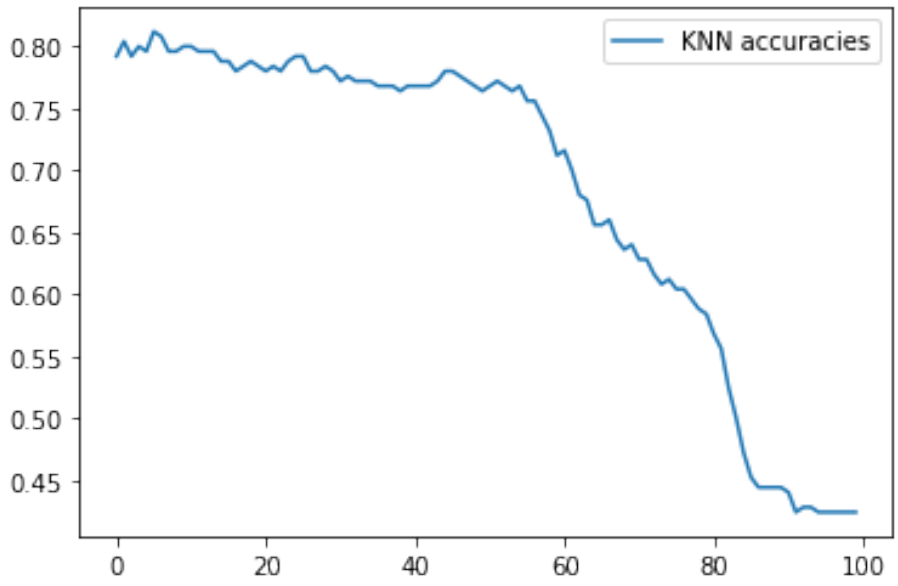

Figure 1: KNN accuracies with different k values. The x-axes is the k values in range [1,199]. y-axes is the accuracy results in range [0,1]

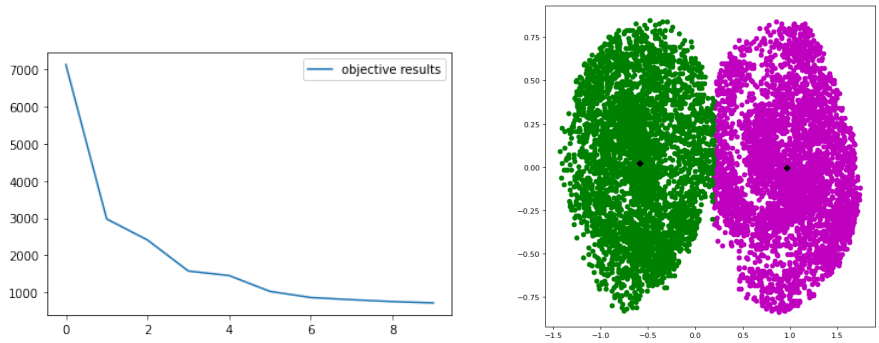

Figure 2: Objective function values of data1 and visualization of the data after k-means clustering where k=2. The black points represent cluster centers.

1.2 Accuracy drops with very large k values

With small k values, we are checking the nearer neighbors of the data point compared to larger k values. When the k gets large, the euclidean distance between the data points gets also larger, as a result we are considering wrong class labels. Small k values are better because distance between data points are also small and we are getting correct class labels

1.3 Accuracy on test set with the best k

The best k value is 11 with accuracy 0.81199.
The test set accuracy is 0.92

2 Part 2: K-means Clustering

2.1 Elbow method and Resultant Clusters

Figure 2 shows the data1's objective function results and data plot. According to elbow method k was chosen as 2.

Figure 3 shows the data2's objective function results and data plot. According to elbow method k was chosen as 3.

Figure 4 shows the data3's objective function results and data plot. According to elbow method k was chosen as 4.

Figure 5 shows the data4's objective function results and data plot. According to elbow method k was chosen as 5.

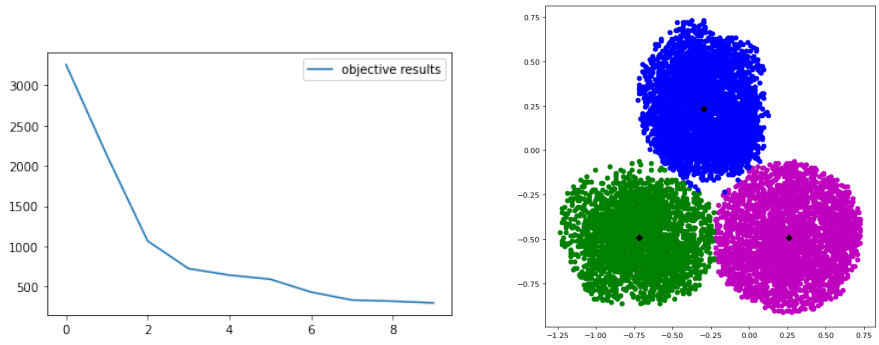

Figure 3: Objective function values of data2and visualization of the data after k-means clustering where k=3. The black points represent cluster centers.

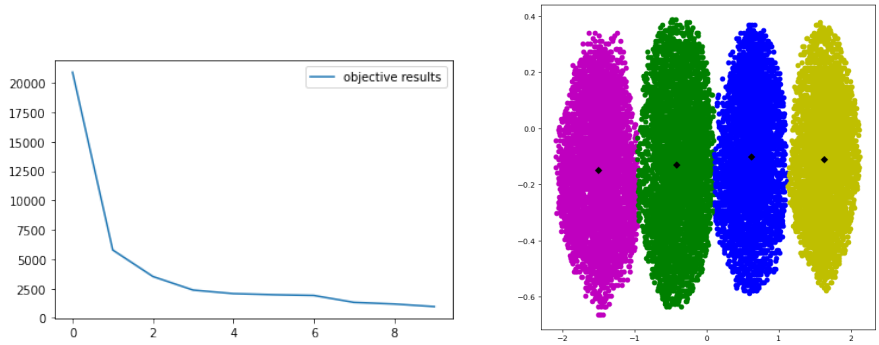

Figure 4: Objective function values of data3 and visualization of the data after k-means clustering where k=4. The black points represent cluster centers.

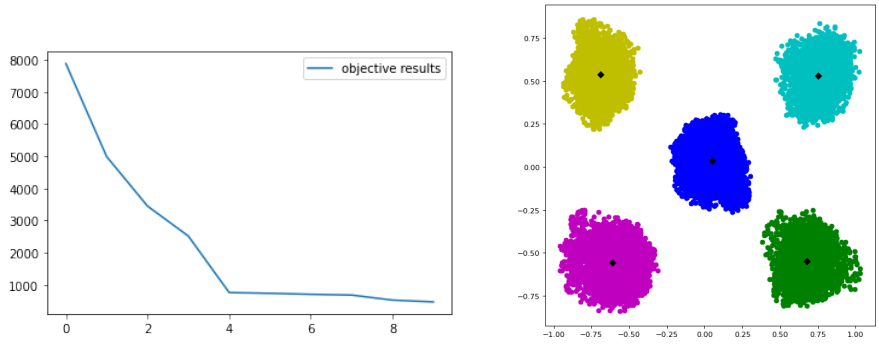

Figure 5: Objective function values of data4 and visualization of the data after k-means clustering where k=5. The black points represent cluster centers.

3 Part 3: Hierarchical Agglomerative Clustering

3.1 data1

Plot the resultant clusters using each criterion and shortly comment on their behaviour.

3.2 data2

Plot the resultant clusters using each criterion and shortly comment on their behaviour.

3.3 data3

Plot the resultant clusters using each criterion and shortly comment on their behaviour.

3.4 data4

Plot the resultant clusters using each criterion and shortly comment on their behaviour.