

Report

Yusif Aliyev

18.05.2021

1 Part 1: K-Nearest Neighbor

1.1 K-fold Cross-validation

The below graph represents k-fold cross validation with the average accuracy values on the y axis and k values on the x axis with $k=1,2,3,\dots,199$. Red dot represents maximum accuracy value.

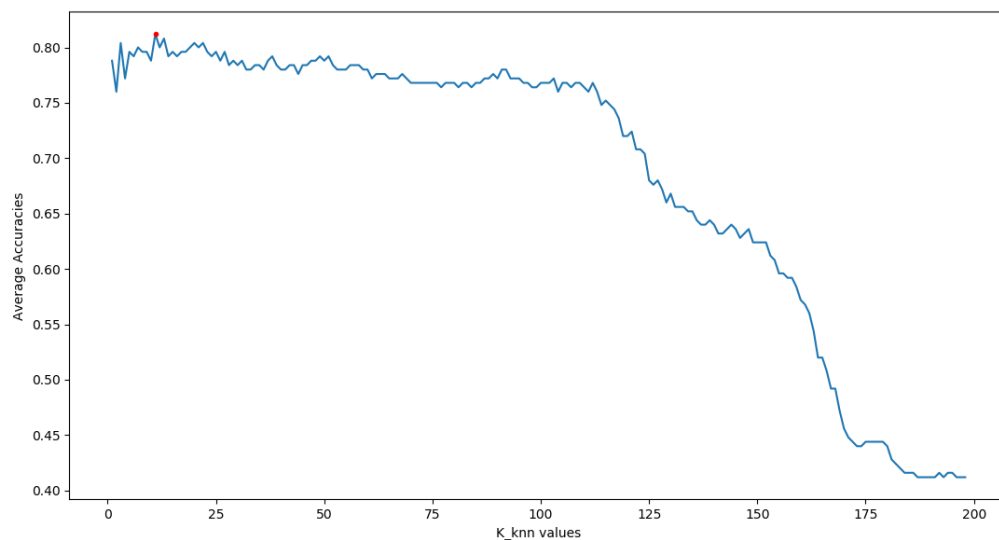


Figure 1: K-fold cross validation graph

1.2 Accuracy drops with very large k values

While increasing the k value to the very large point, overfitting will be happened, because with such high value of k , it can cover most of the data points, such

that decision boundary becomes smoother and more resilient to outliers and thus resulting in reduced variance and high bias.

1.3 Accuracy on test set with the best k

Maximum average validation accuracy is 0.79. The best test set accuracy value is 0.82 with the best value of $k_{KNN} = 11$.

2 Part 2: K-means Clustering

2.1 Elbow method

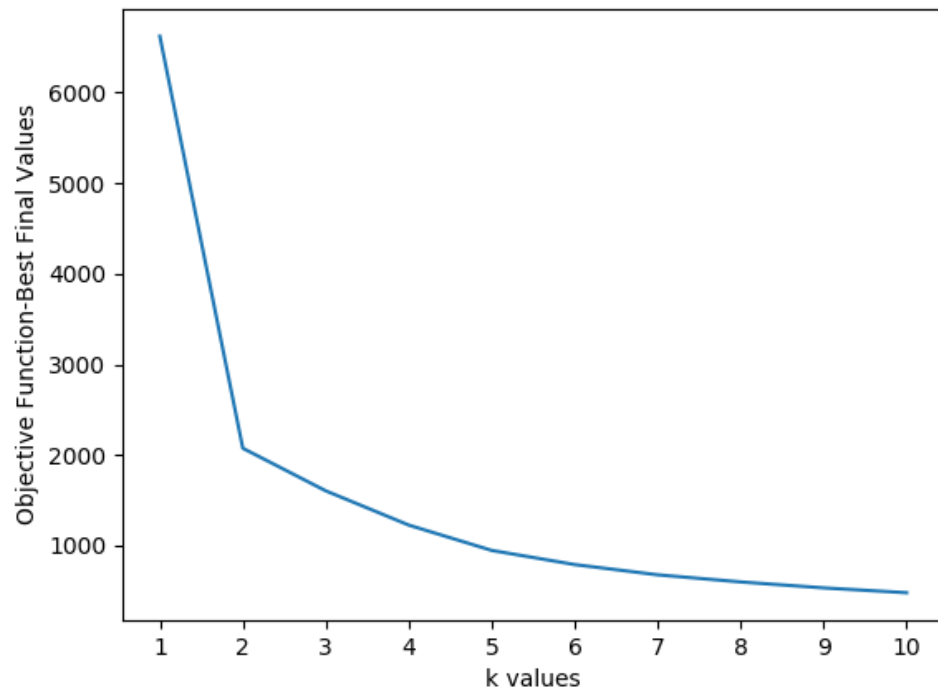


Figure 2: Elbow method graph for Clustering 1 (suitable $k = 2$)

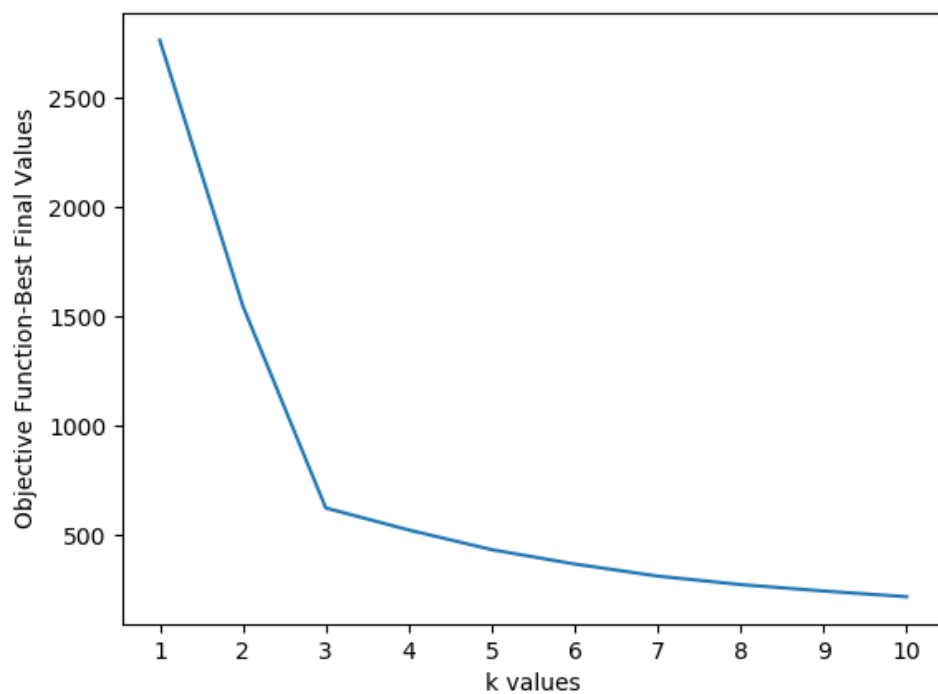


Figure 3: Elbow method graph for Clustering 2 (suitable $k = 3$)

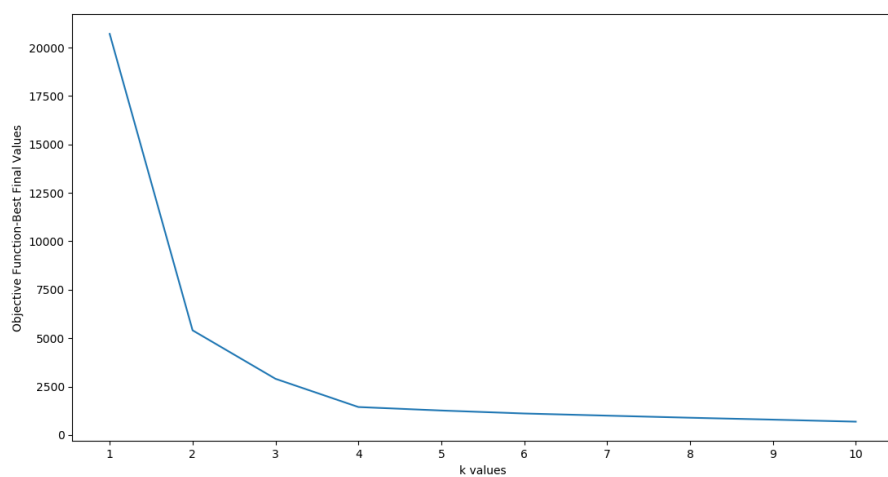


Figure 4: Elbow method graph for Clustering 3 (suitable $k = 4$)

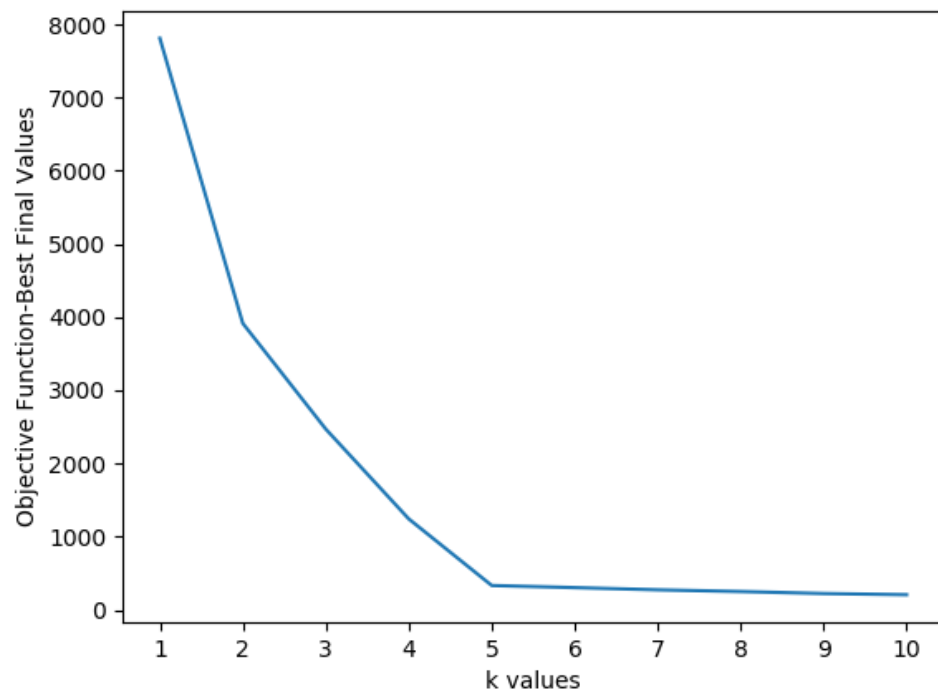


Figure 5: Elbow method graph for Clustering 4 (suitable $k = 5$)

2.2 Resultant Clusters

There are resultant clusters plots for each clustering data below:

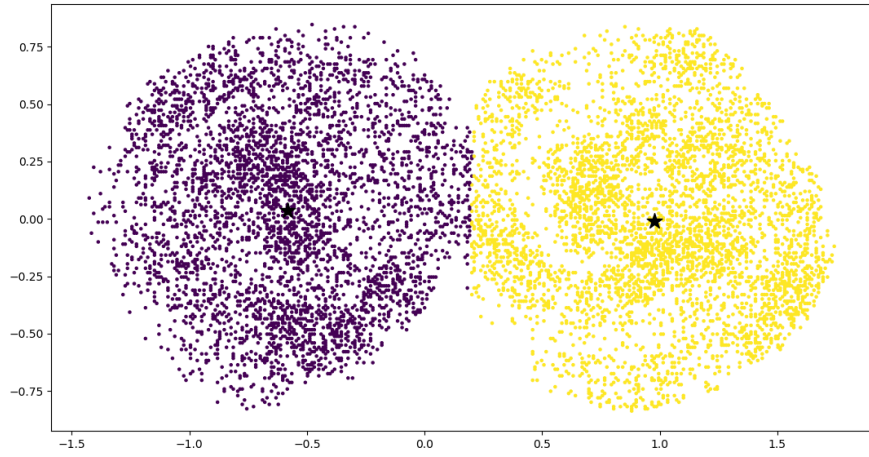


Figure 6: Resultant Cluster 1 (with the value of $k = 2$)

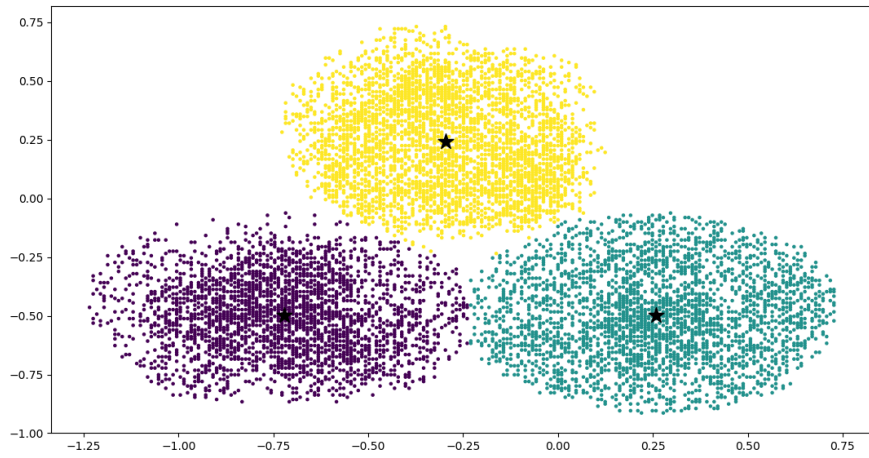


Figure 7: Resultant Cluster 2 (with the value of $k = 3$)

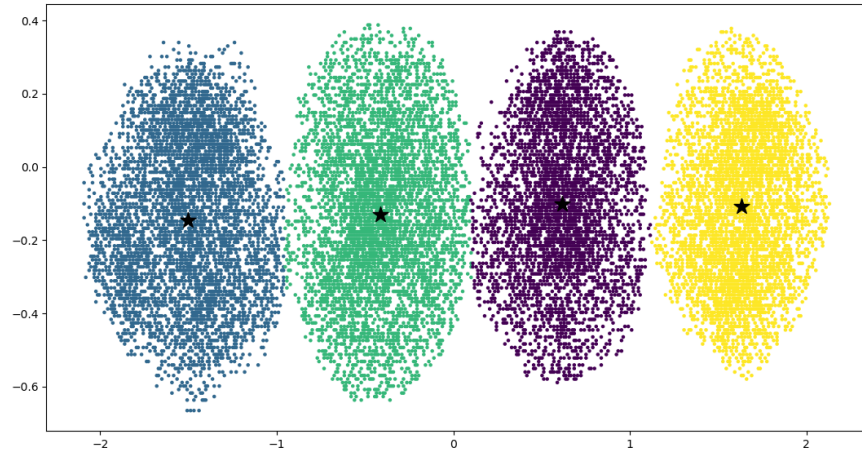


Figure 8: Resultant Cluster 3 (with the value of $k = 4$)

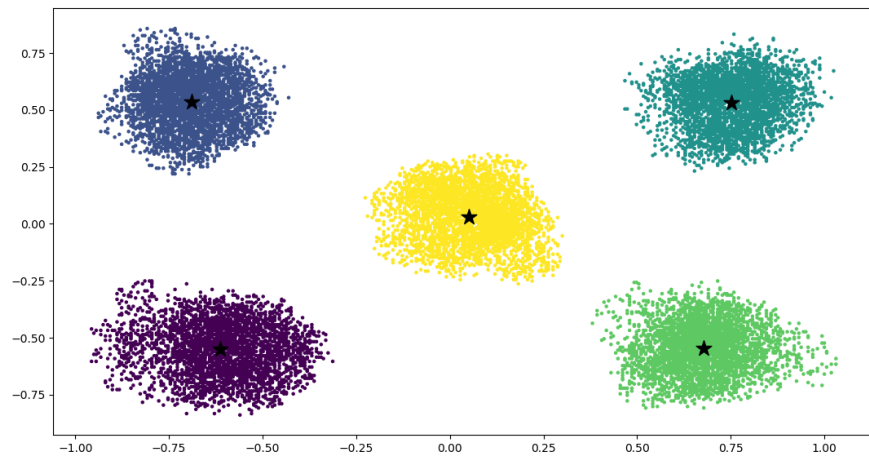


Figure 9: Resultant Cluster 4 (with the value of $k = 5$)

3 Part 3: Hierarchical Agglomerative Clustering

3.1 data1

Plot the resultant clusters using each criterion and **shortly comment on their behaviour and why they work in that way.**

3.2 data2

Plot the resultant clusters using each criterion and **shortly comment on their behaviour and why they work in that way.**

3.3 data3

Plot the resultant clusters using each criterion and **shortly comment on their behaviour and why they work in that way.**

3.4 data4

Plot the resultant clusters using each criterion and **shortly comment on their behaviour and why they work in that way.**