

Machine Learning Clustering Types And Examining Results

Sahin Kasap

25.12.2020

1 Part 1: K-Nearest Neighbor

1.1 K-fold Cross-validation

Plot the result of you k-fold cross-validation according to the description in the homework.

1.2 Accuracy drops with very large k values

Explain why this happens.

1.3 Accuracy on test set with the best k

Report your best k value and test set accuracy using it.

2 Part 2: K-means Clustering

2.1 Elbow method

In the Part2 images we can see the cluster centers according to the number of cluster centers pre-given. And in the graph we can see the sums of the lengths to the cluster centers. It is clearly seen that the result is 2 in the first cluster by applying elbow method. 2 is the number of clusters.

2.2 Resultant Clusters

3 Part 3: Hierarchical Agglomerative Clustering

3.1 data1

The nearest clustering(Single-Linkage method) has done what it must supposed to but complete linkage didn't give the desired outputs. I think that is normal.

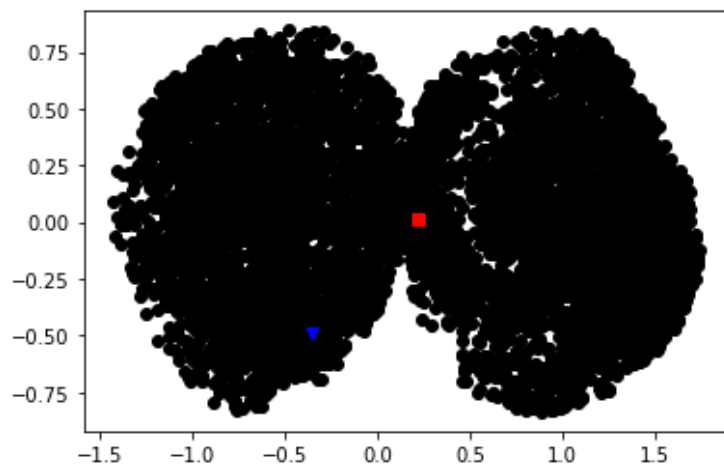


Figure 1: Part2-1 The cluster centers with one center

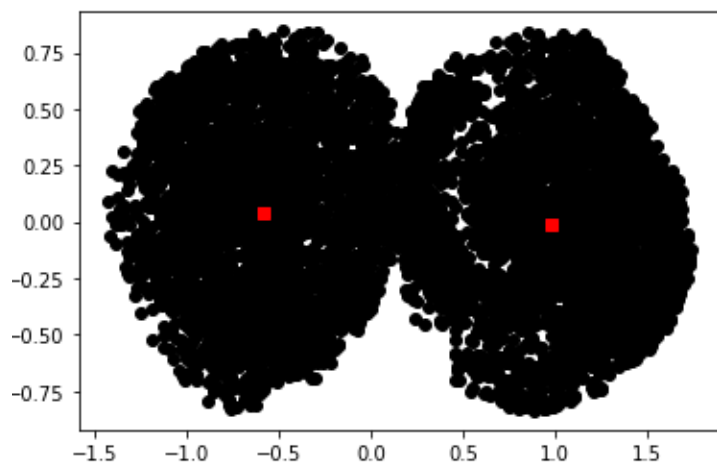


Figure 2: Part2-1 The cluster centers with two centers

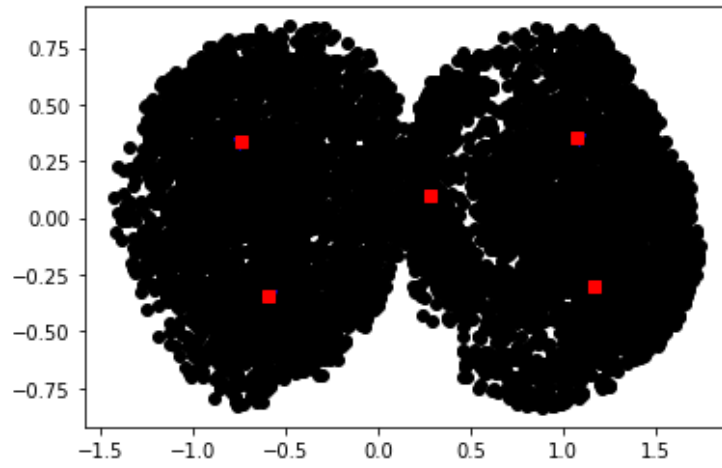


Figure 3: Part2-1 The cluster centers with five centers

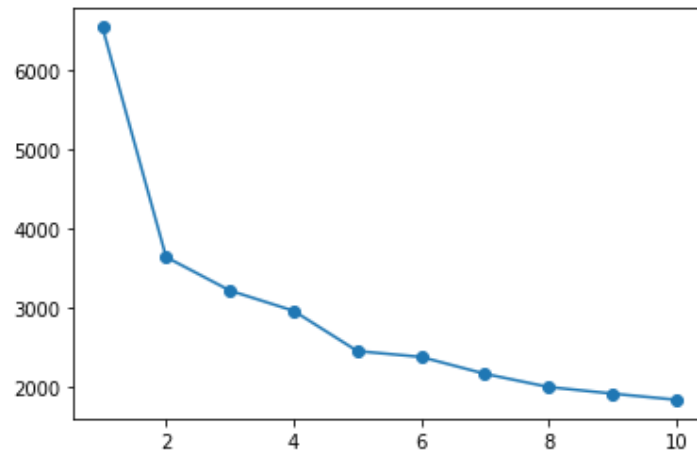


Figure 4: Part2-1 The sums of the lengths to the centers

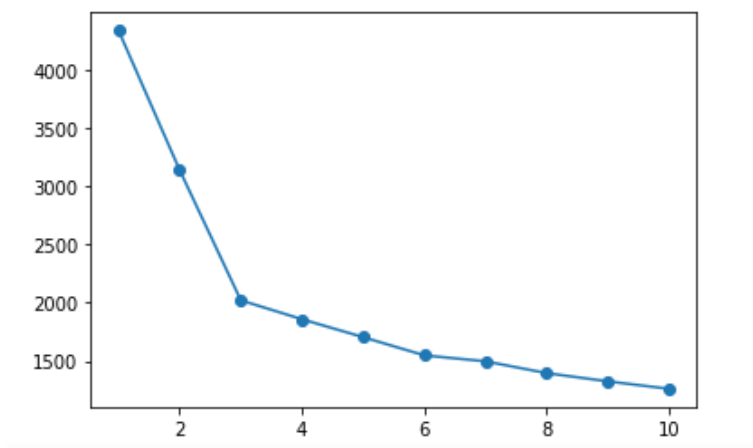


Figure 5: Result of objective function in cluster 2

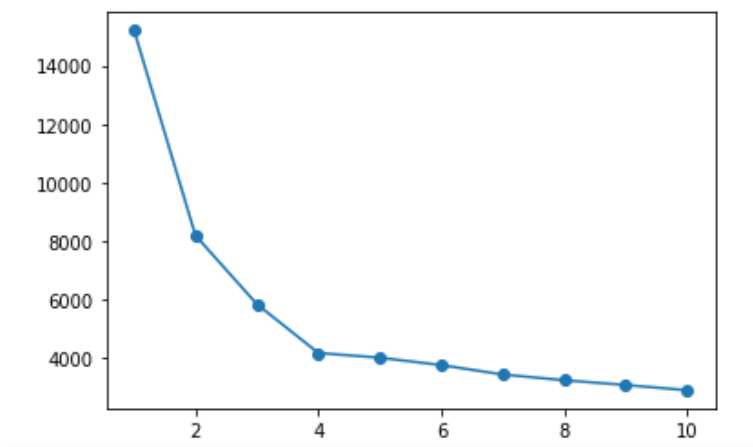


Figure 6: Result of objective function in cluster 3

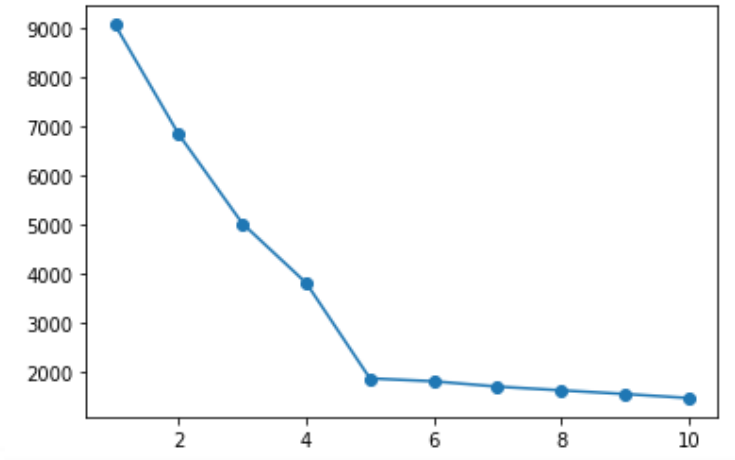


Figure 7: Result of objective function in cluster 4

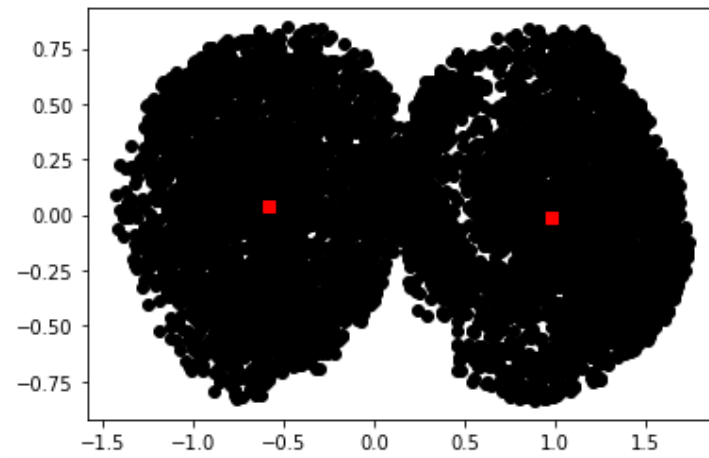


Figure 8: The cluster centers according to the result of the elbow method in cluster 1

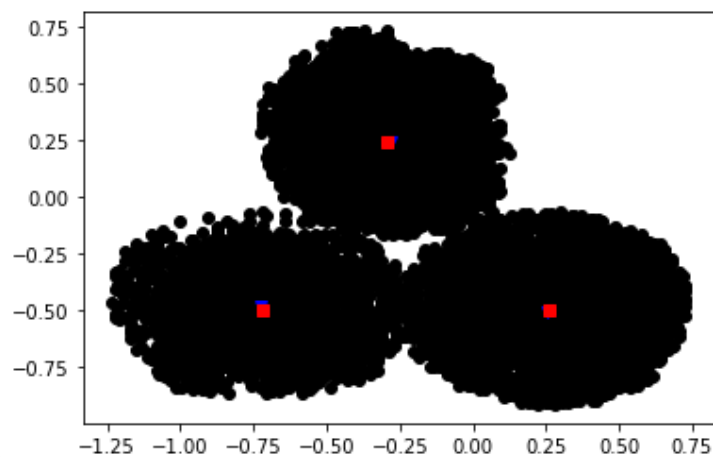


Figure 9: The cluster centers according to the result of the elbow method in cluster 2

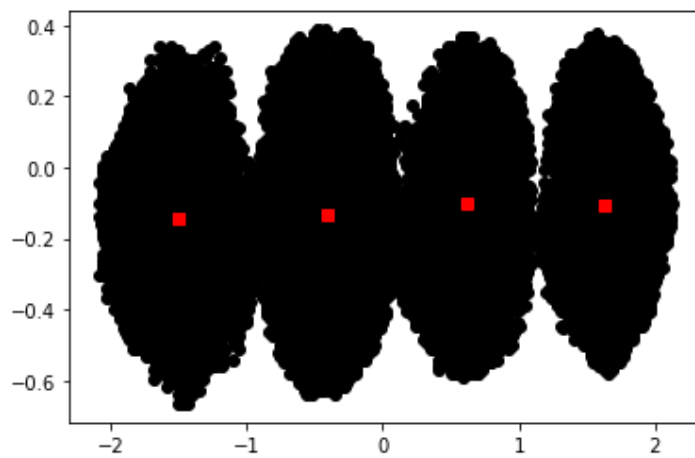


Figure 10: The cluster centers according to the result of the elbow method in cluster 3

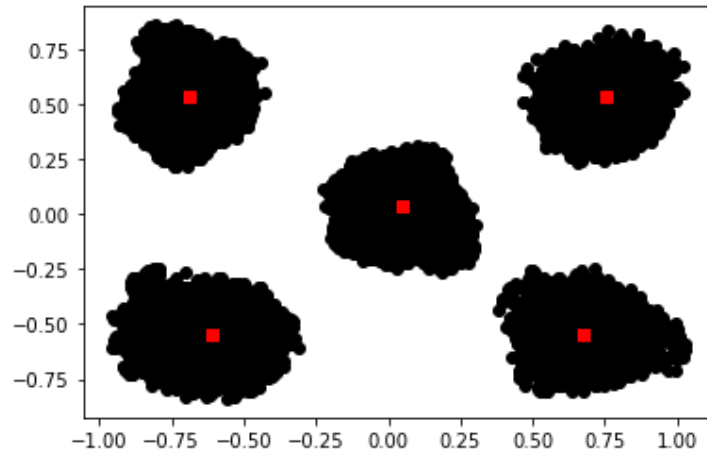


Figure 11: The cluster centers according to the result of the elbow method in cluster 4

3.2 data2

The nearest(single linkage) method is the best one for this, because the data consists of two clusters that are far apart

3.3 data3

The complete linkage is not good. This is because complete linkage is not good for together data.

3.4 data4

The plotted results are here: The results are not desired ones in single and complete linkage but in average linkage, the results are the best one. This is because in average linkage, these kind of kinky data are better resulted.

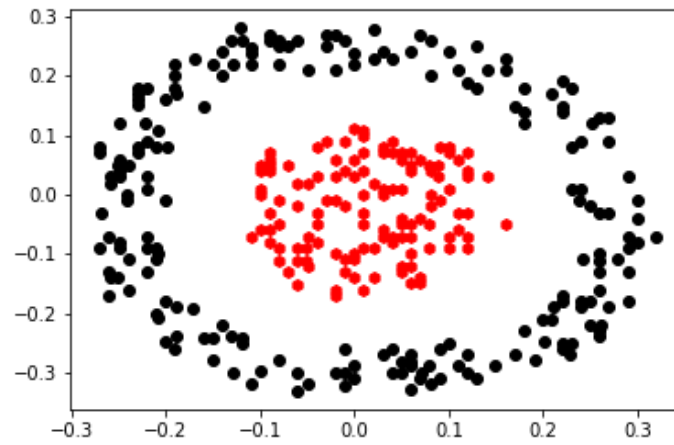


Figure 12: Nearest for cluster 1

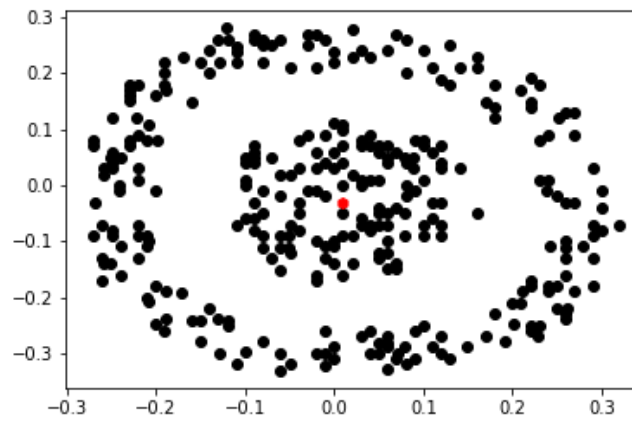


Figure 13: Complete linkage for data 1

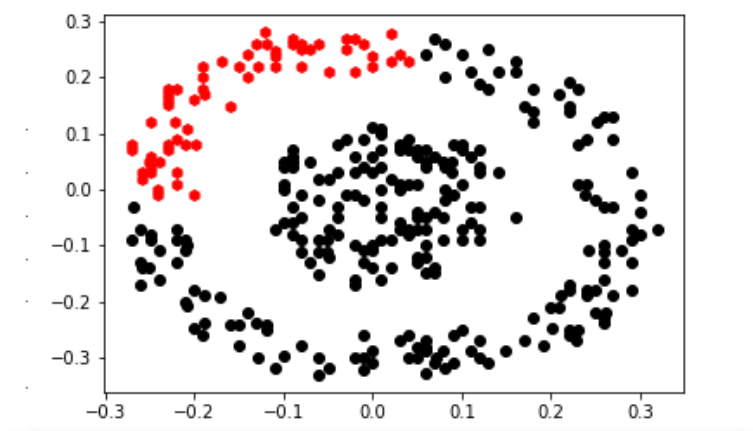


Figure 14: Average-Linkage for data 1

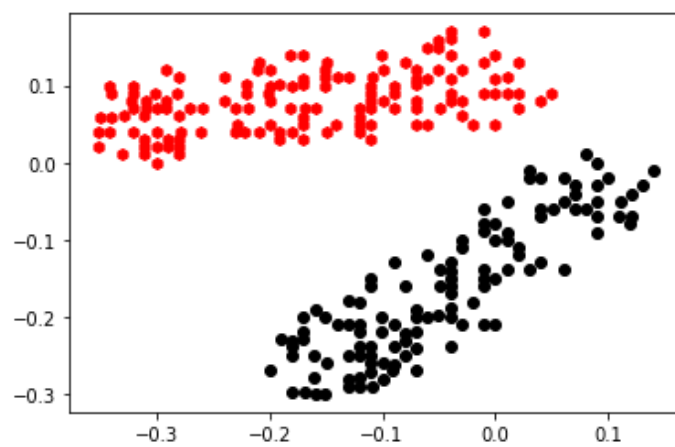


Figure 15: Nearest for cluster 2

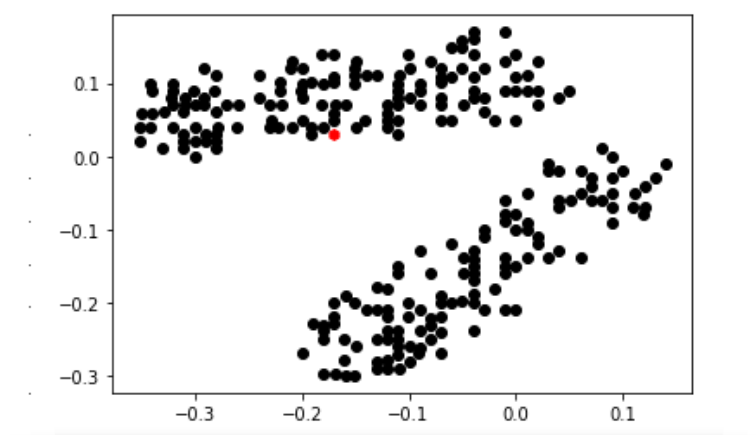


Figure 16: Complete linkage for data 2

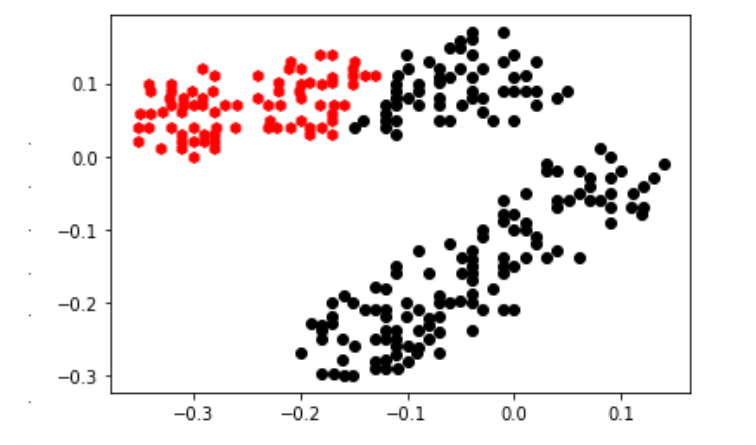


Figure 17: Average-Linkage for data 2

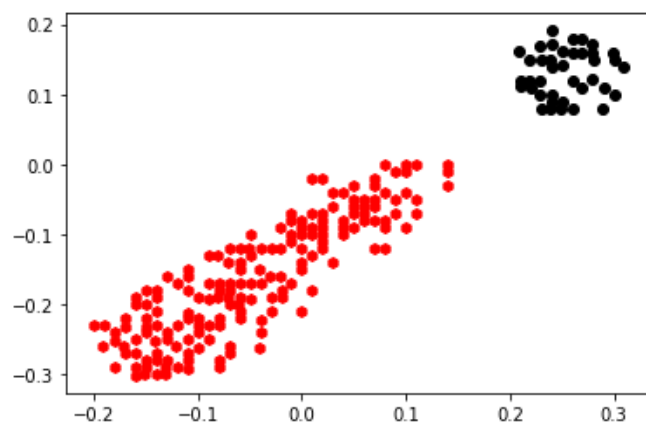


Figure 18: Nearest for cluster 3

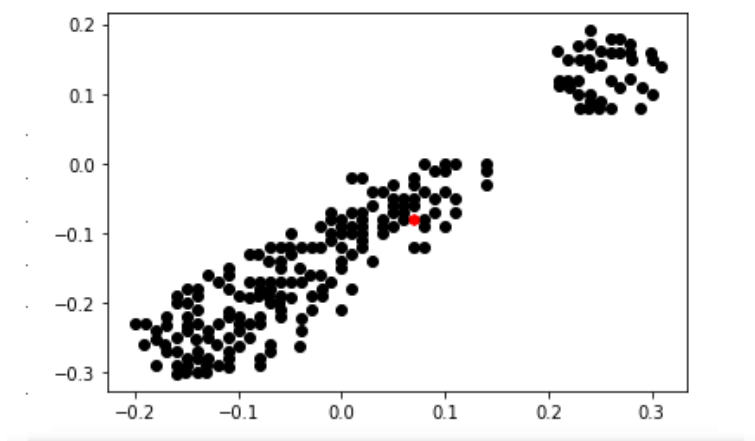


Figure 19: Complete linkage for data 3

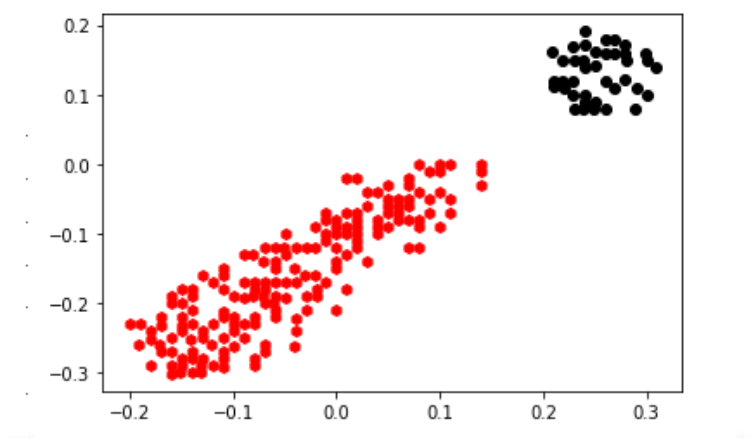


Figure 20: Average-Linkage for data 3

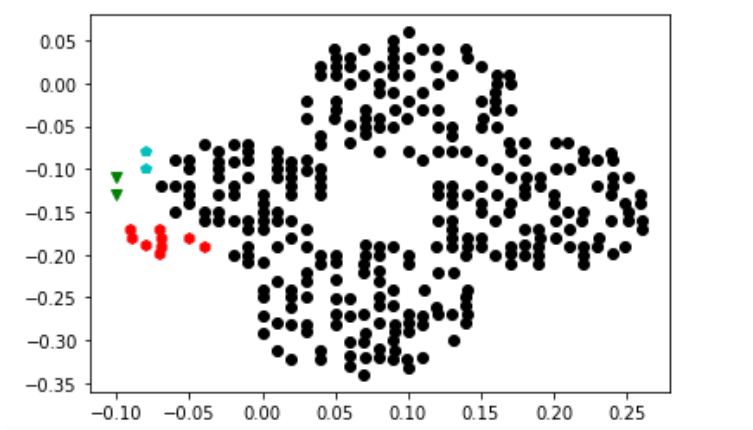


Figure 21: Nearest for cluster 4

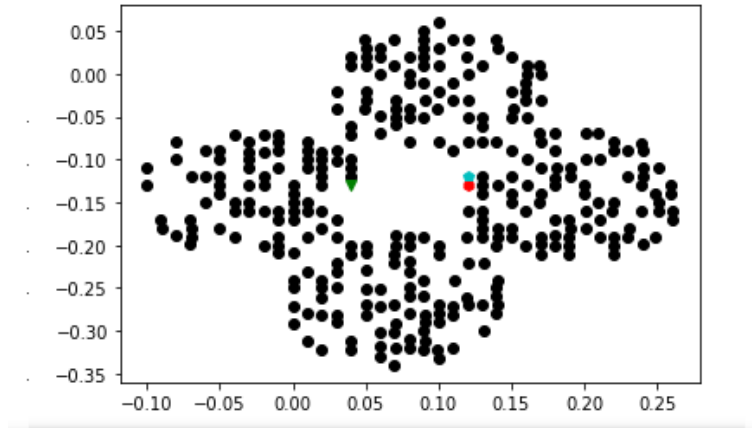


Figure 22: Complete linkage for data 4

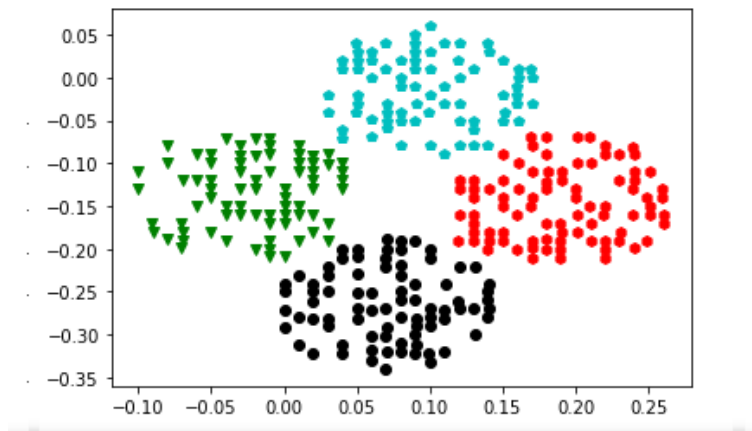


Figure 23: Average-Linkage for data 4