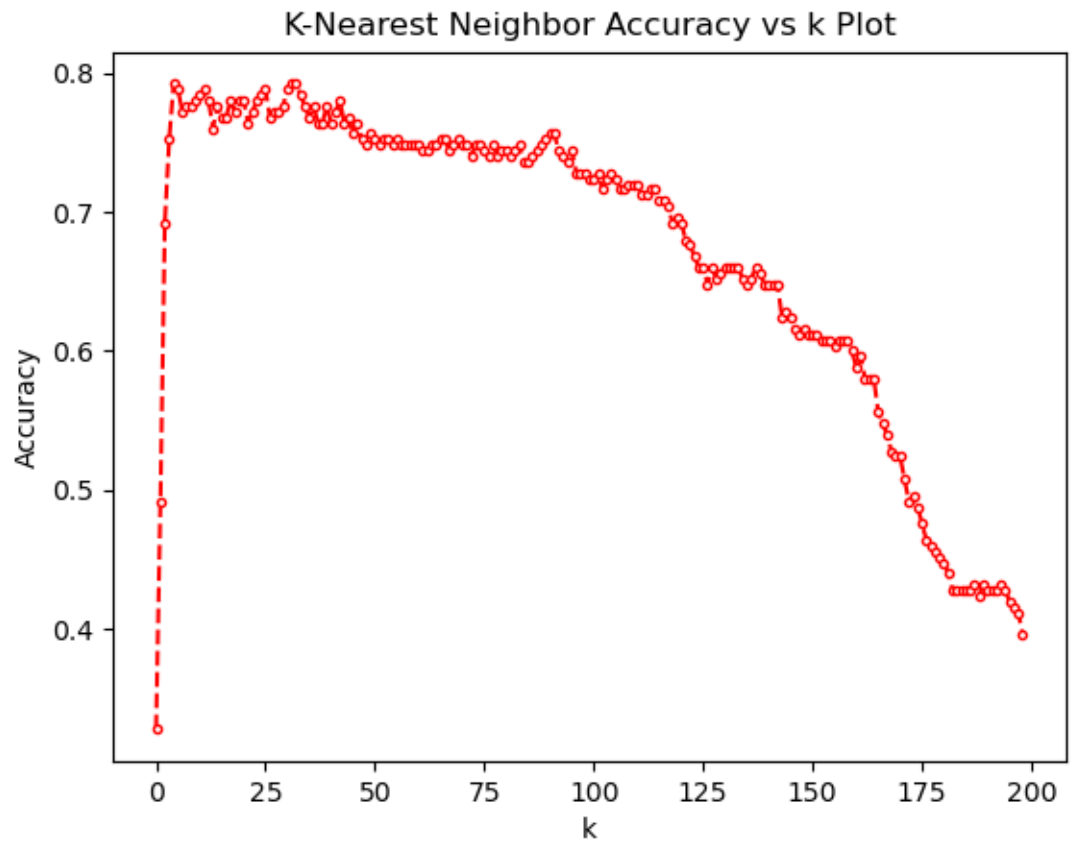# Report

Yusuf Eren Tunç

23/12/2020

1 Part 1: K-Nearest Neighbor

1.1 K-fold Cross-validation

1.2 Accuracy drops with very large k values

In large k values, system is too generalized for nearest neighbors. In other word, number of neighbors are too much for a healthy classification. K-Nearest Neighbor with large k value is bad in both train data and test data. Shortly, system would be under-fitting.
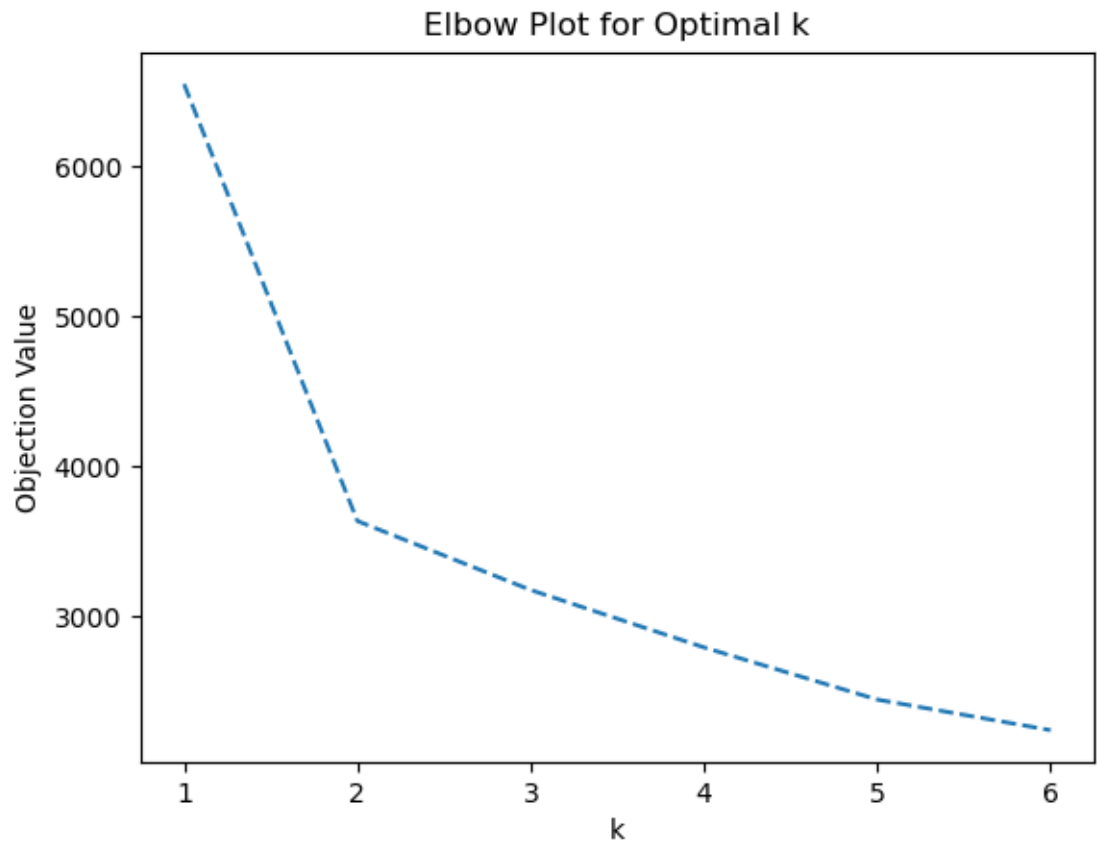
1.3 Accuracy on test set with the best k

When I start my script, I get 5 for k value. With that value, accuracy of testing data was %76
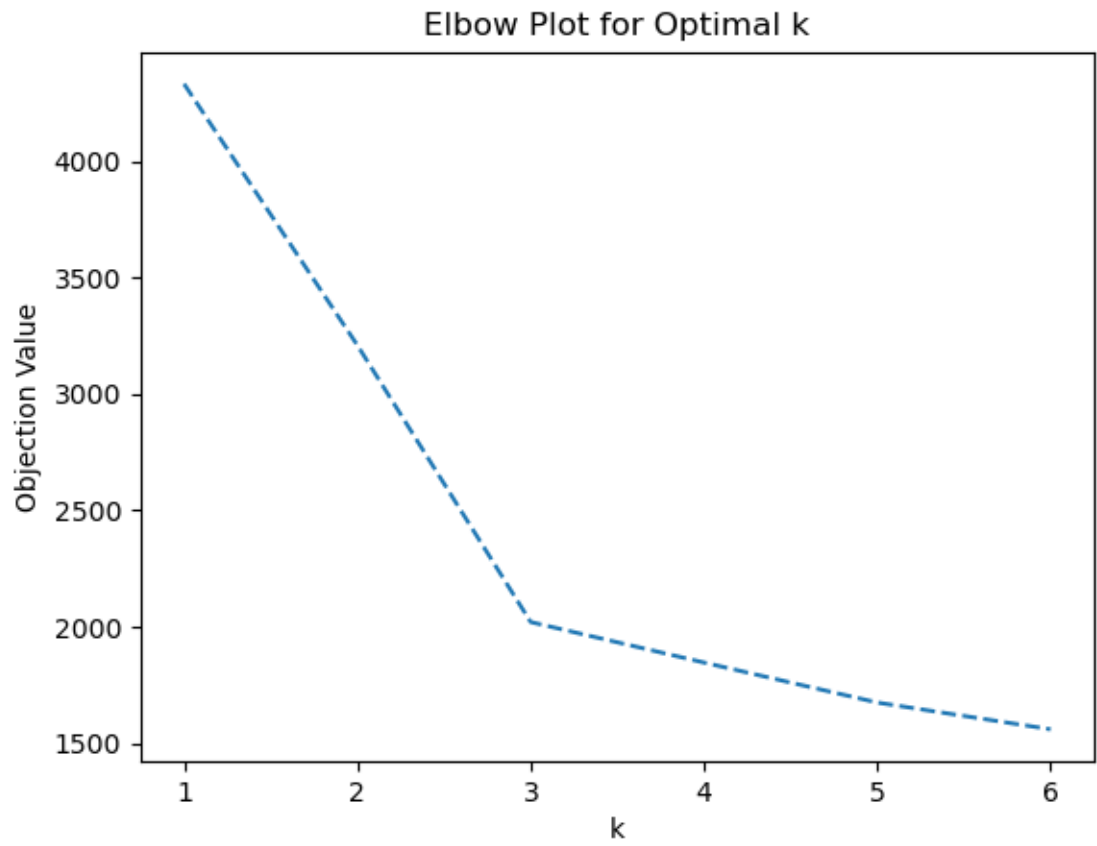
2   Part 2: K-means Clustering

2.1   Elbow method

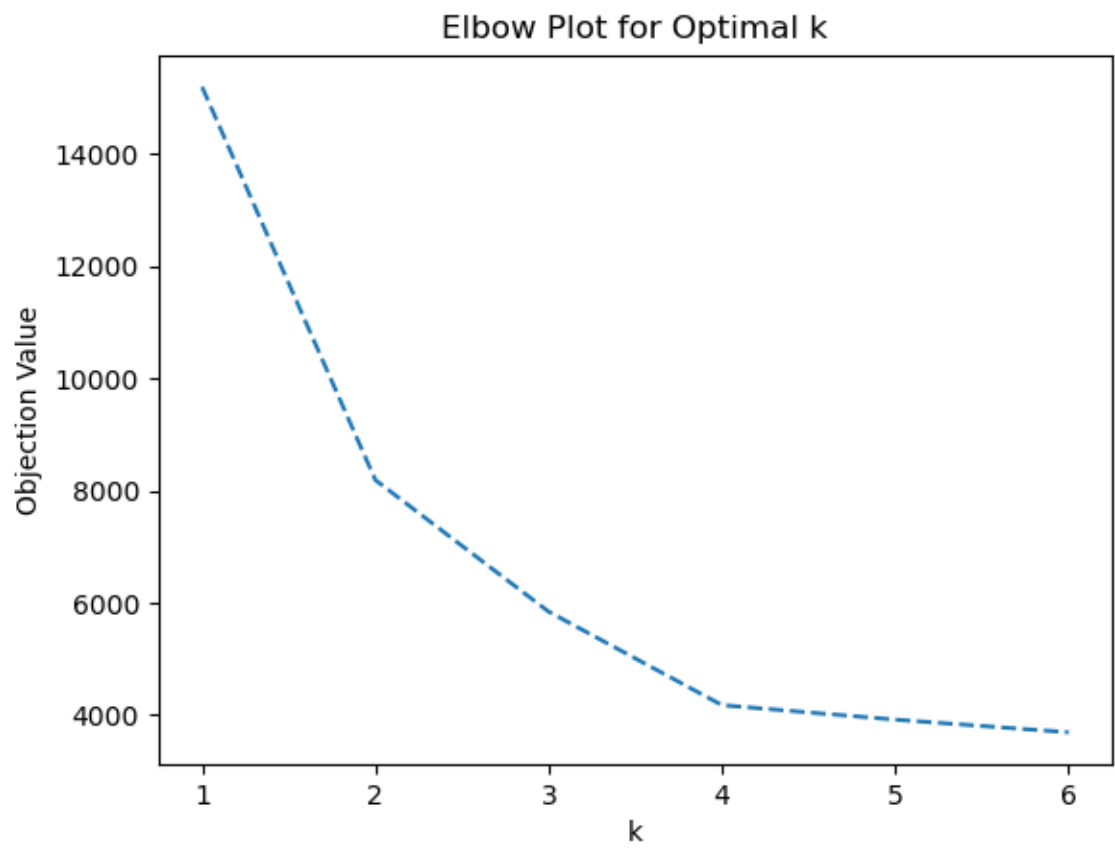### 2.1.1   Clustering 1

Elbow Plot for Optimal k

Selected k value is 2

2.1.2 Clustering 2

Selected k value is 3

2.1.3 Clustering 3

Selected k value is 4

2.1.4 Clustering 4

Elbow Plot for Optimal k
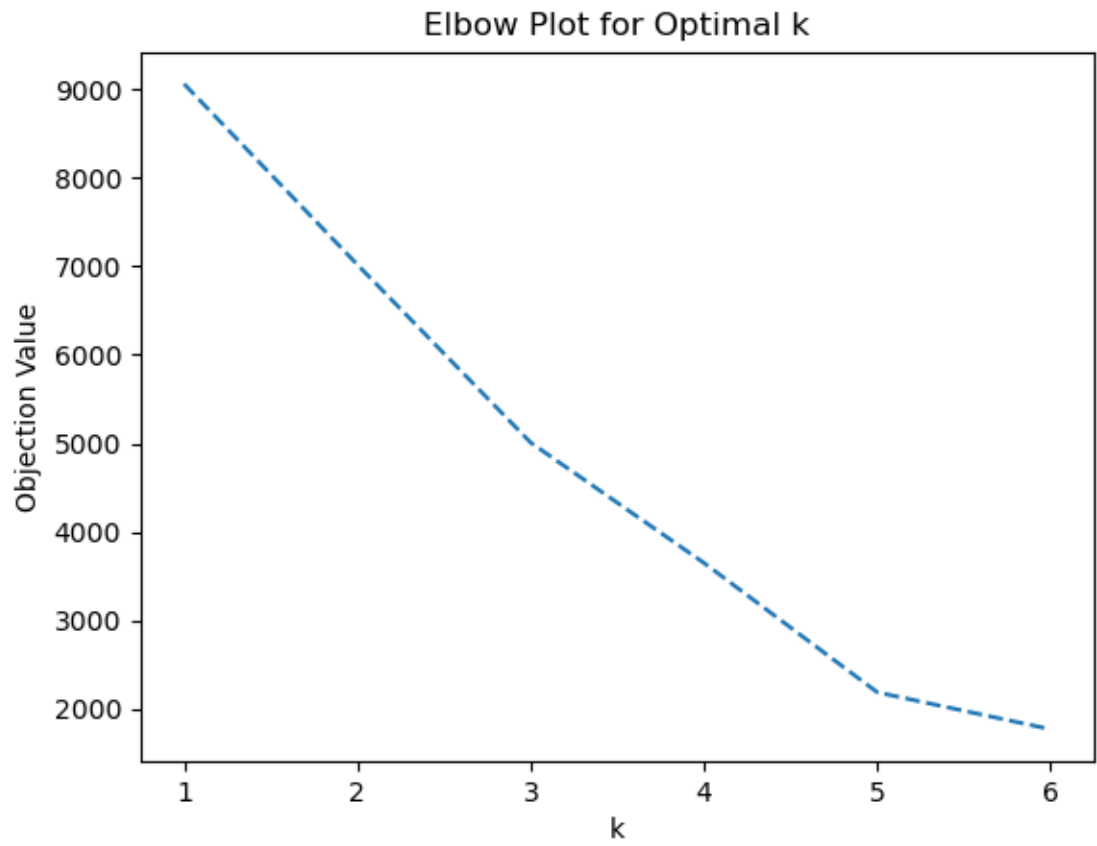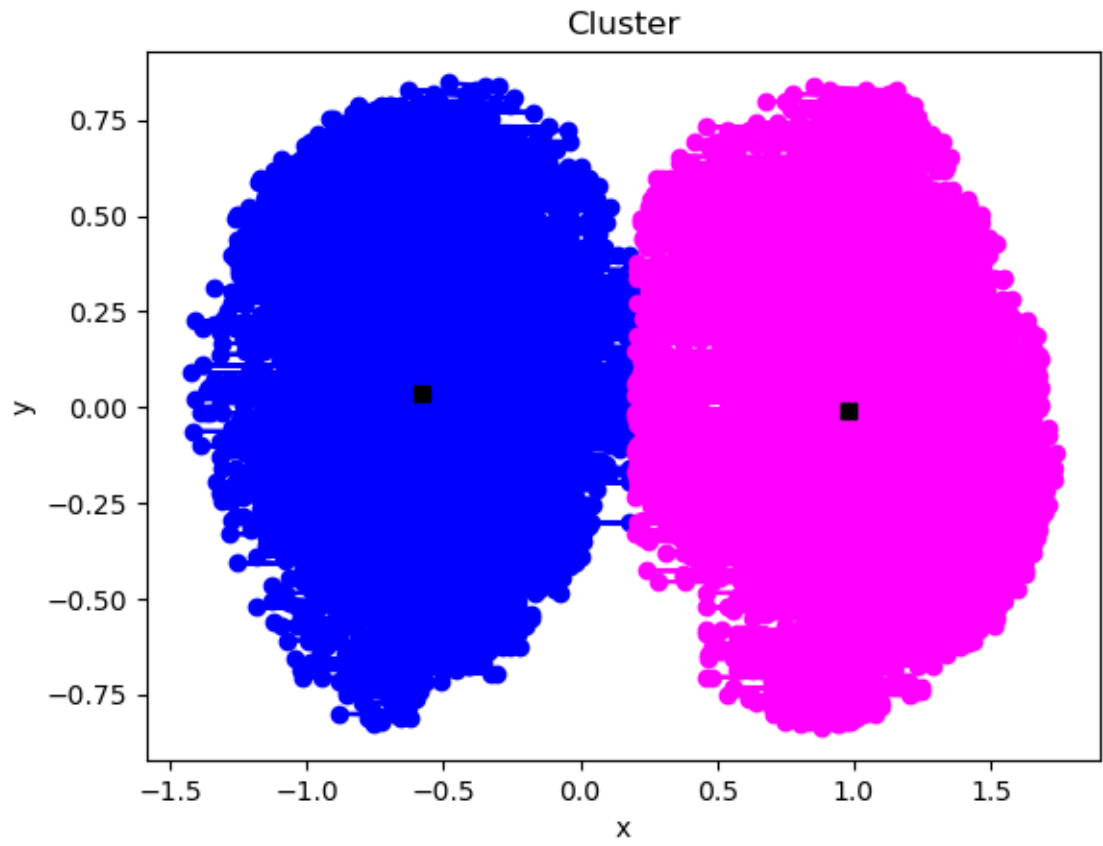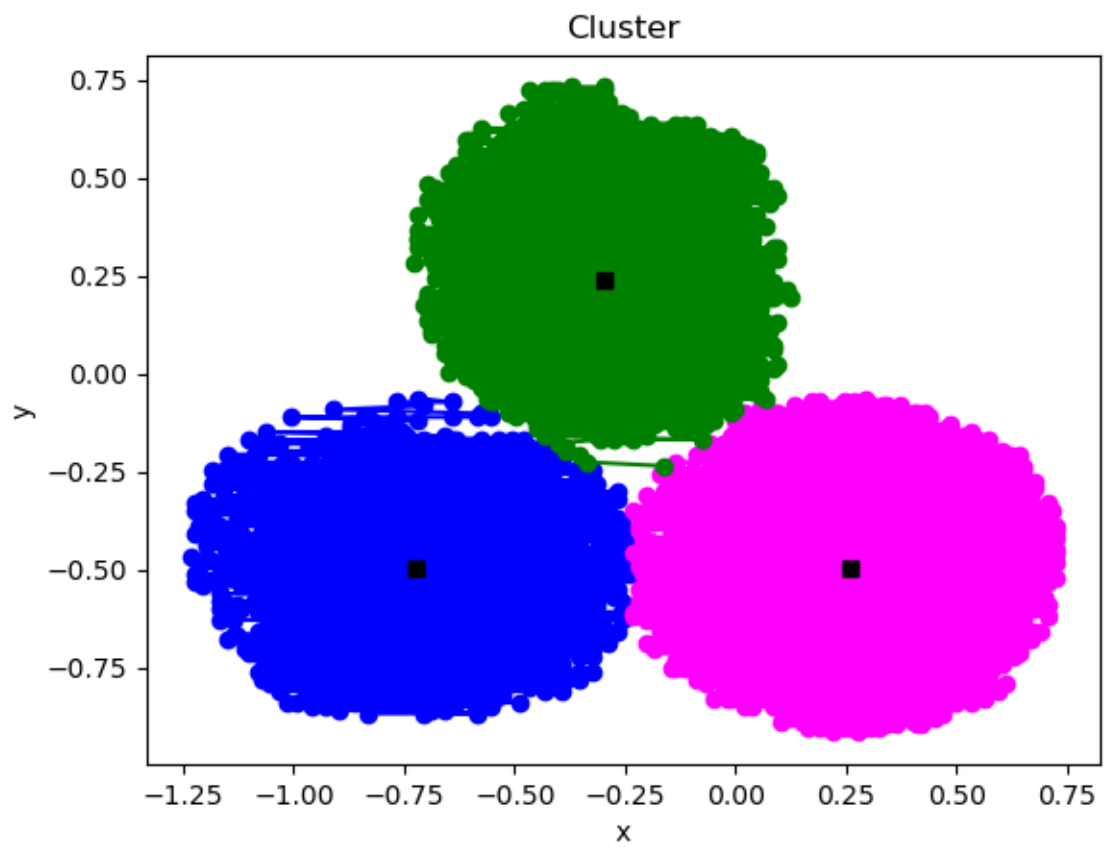
Selected k value is 5

2.2 Resultant Clusters

2.2.1 Clustering 1

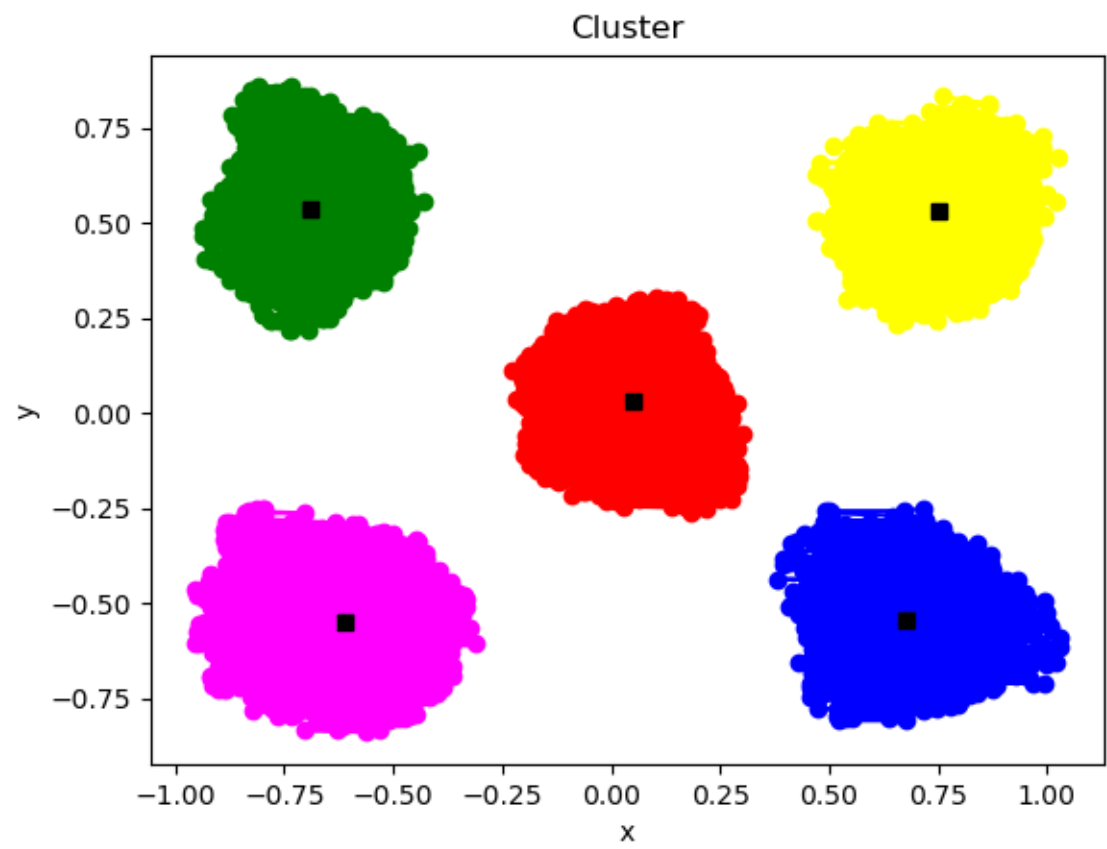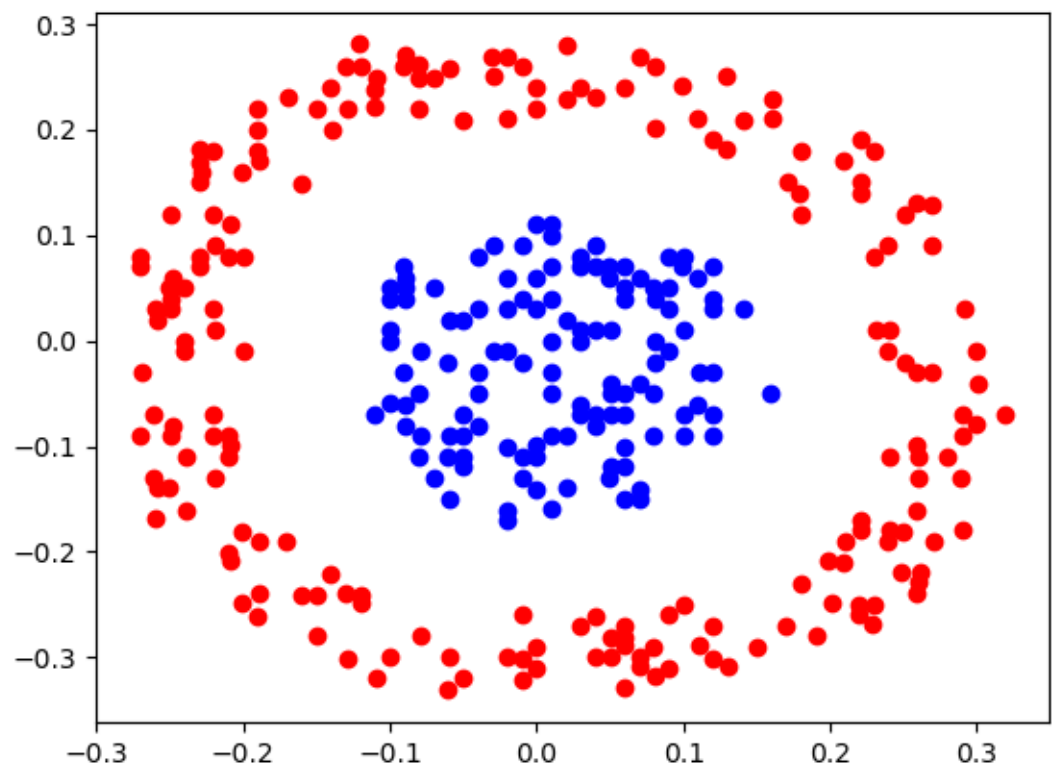### 2.2.2   Clustering 2

Cluster
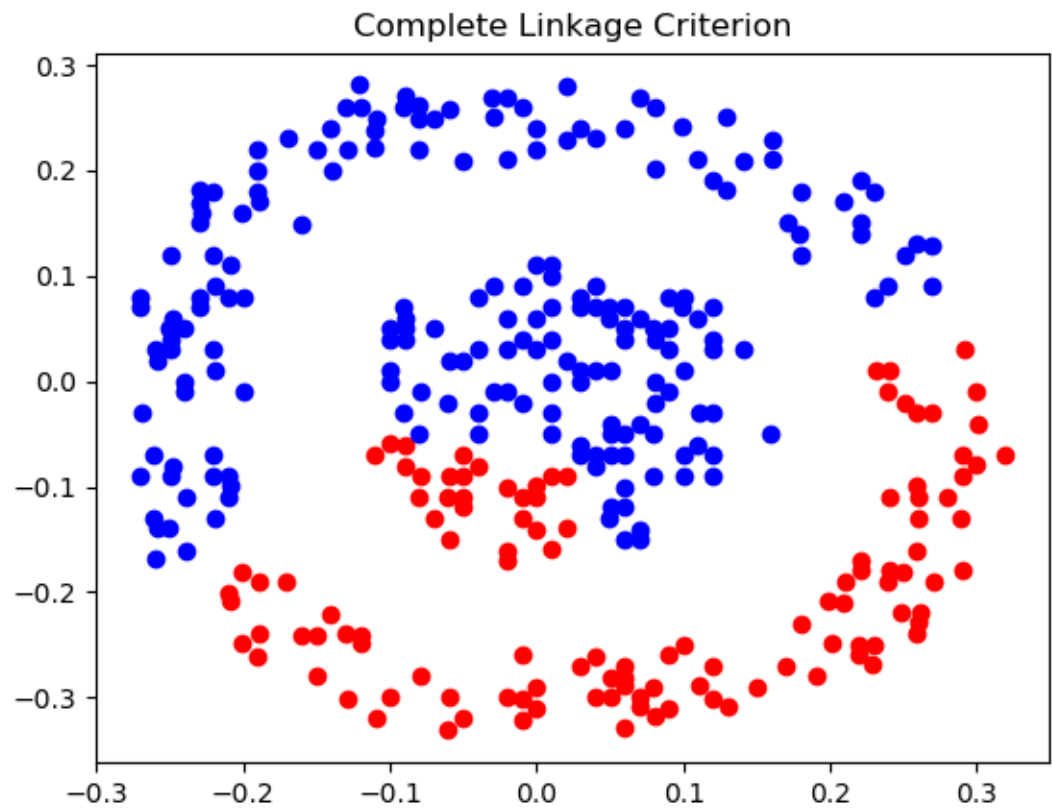
2.2.3 Clustering 3
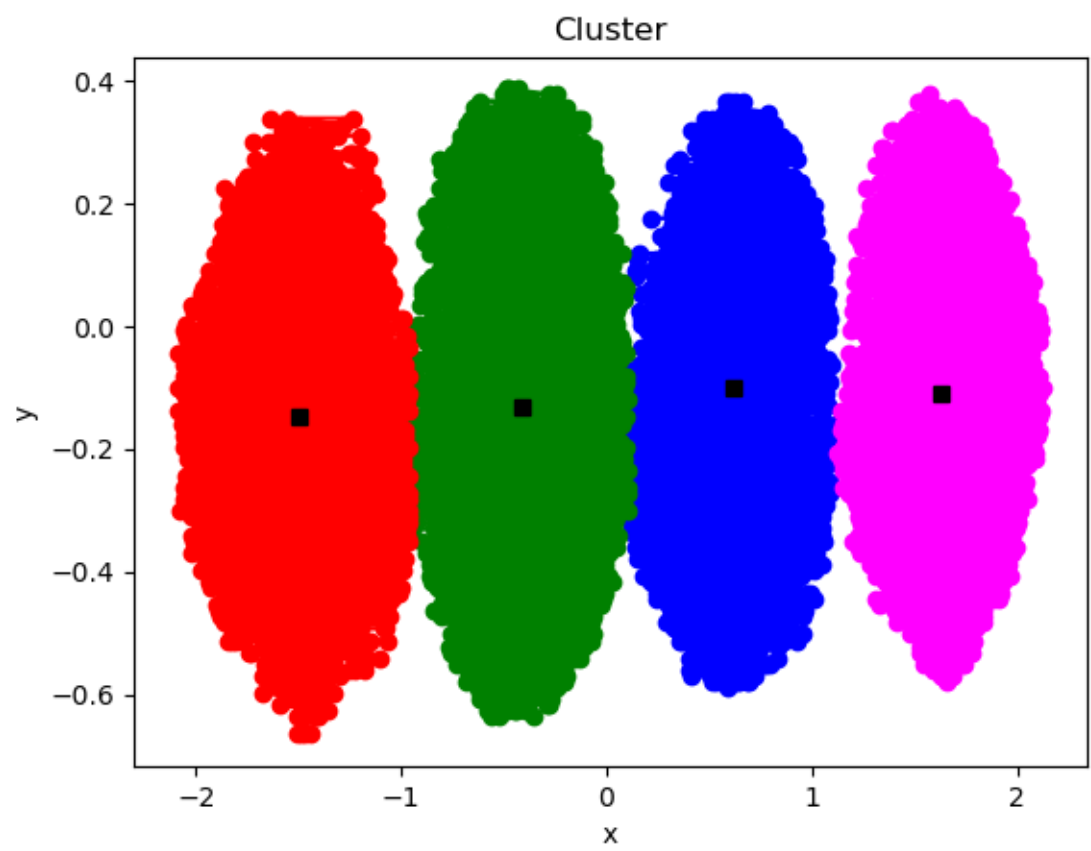
Cluster

2.2.4 Clustering 4

3 Part 3: Hierarchical Agglomerative Clustering
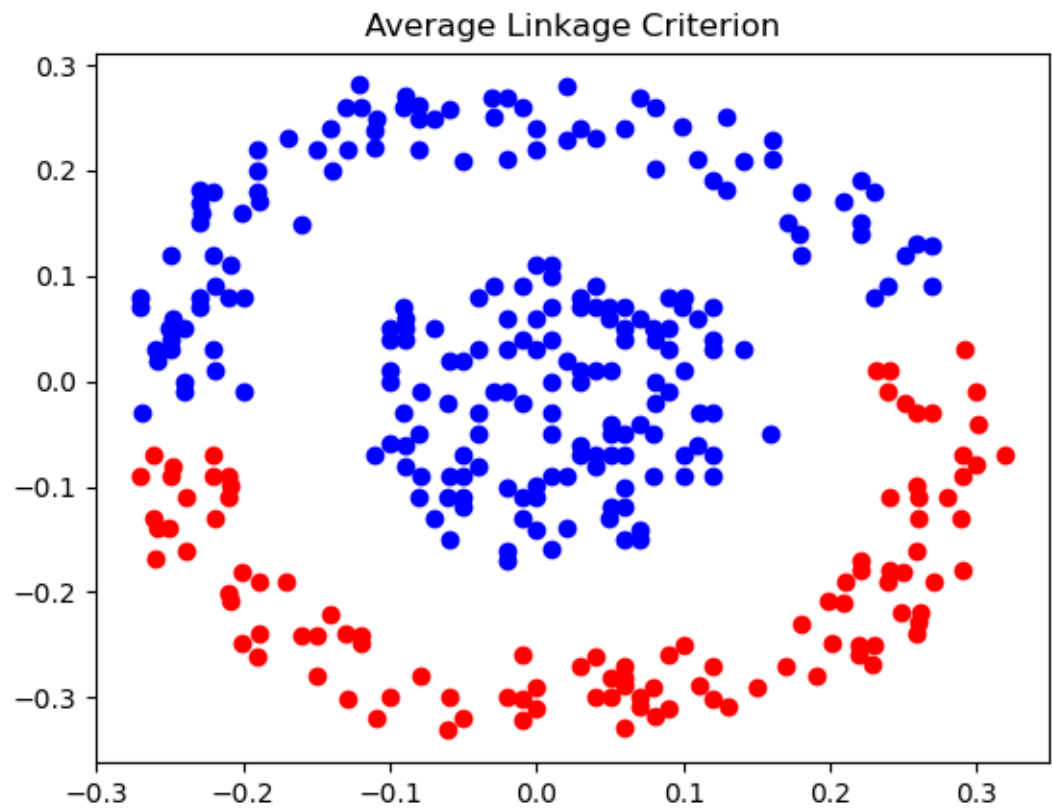
3.1 data1

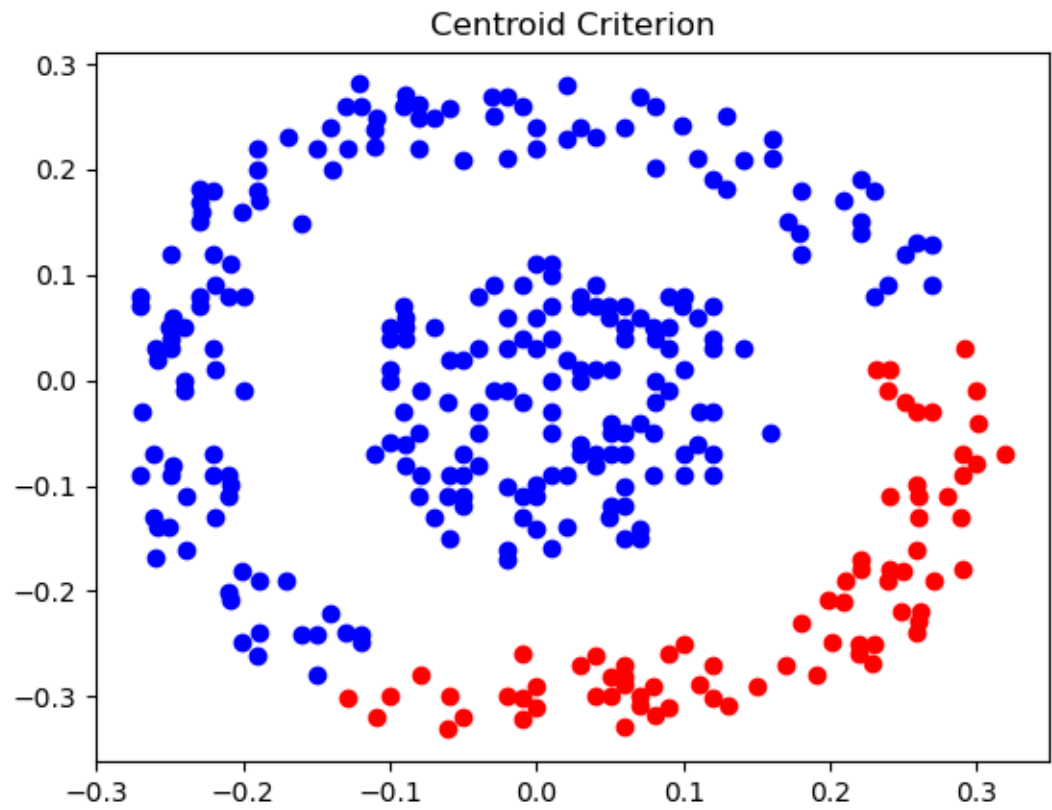Since there is a reasonable gap between two final cluster, single linkage works perfectly.

Since our data's shape has similarity to circle shape, complete linkage divided all data into two small part almost from center of graph.

With same reason of complete linkage, average linkage divides data set into
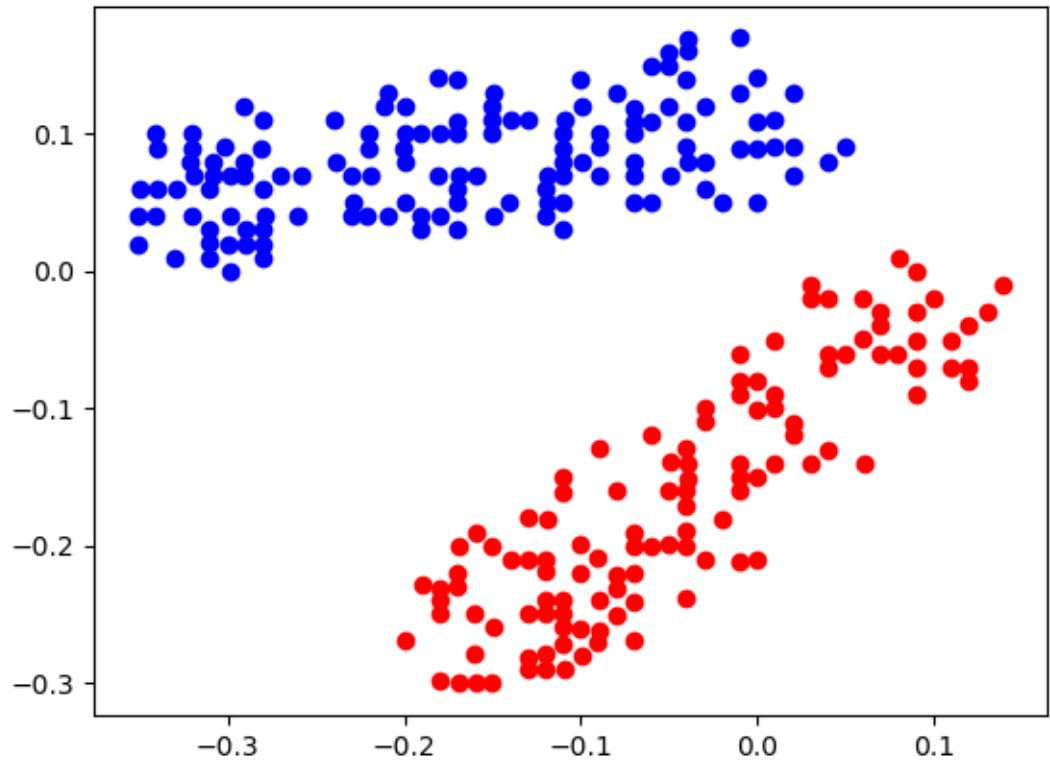two part from almost center.

Average Linkage Criterion

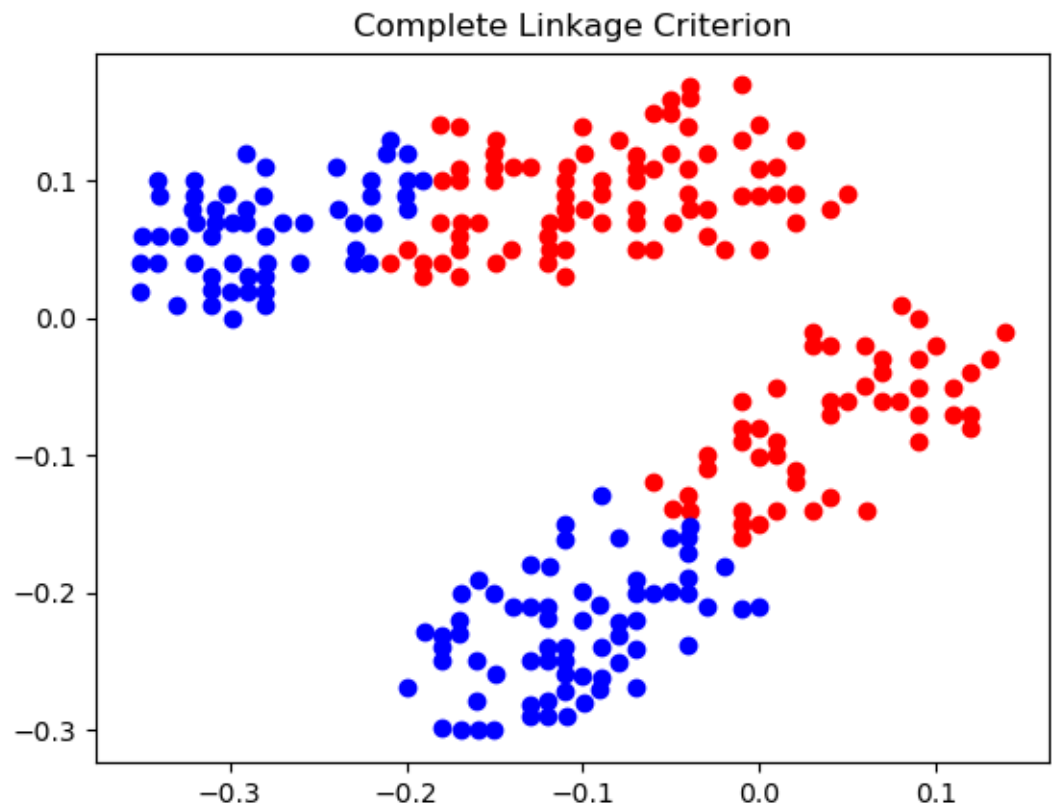Centroids of clusters are not perfectly share data between each other.

14

3.2 data2

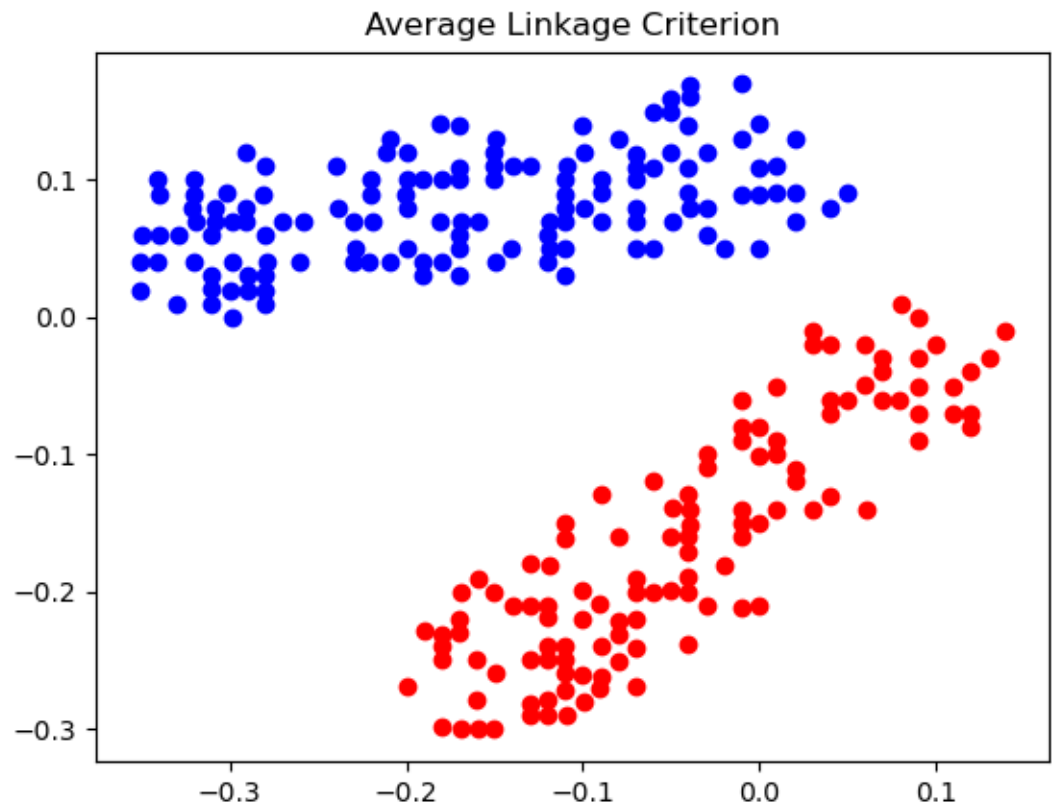Since there is a reasonable gap between two final cluster, single linkage works perfectly.

Complete linkage didn't work well since perfect cluster's inside long distance is longer that other cluster.
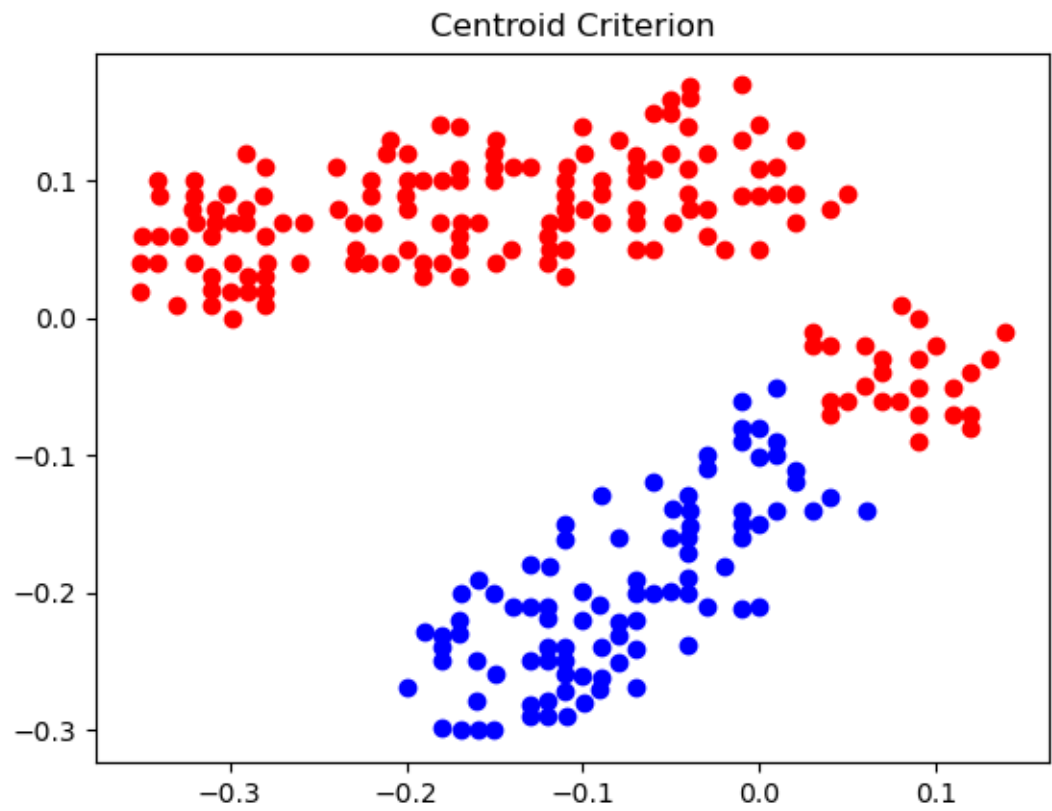
Complete Linkage Criterion

Average linkage works perfectly since clusters distribute to sides uniformly.
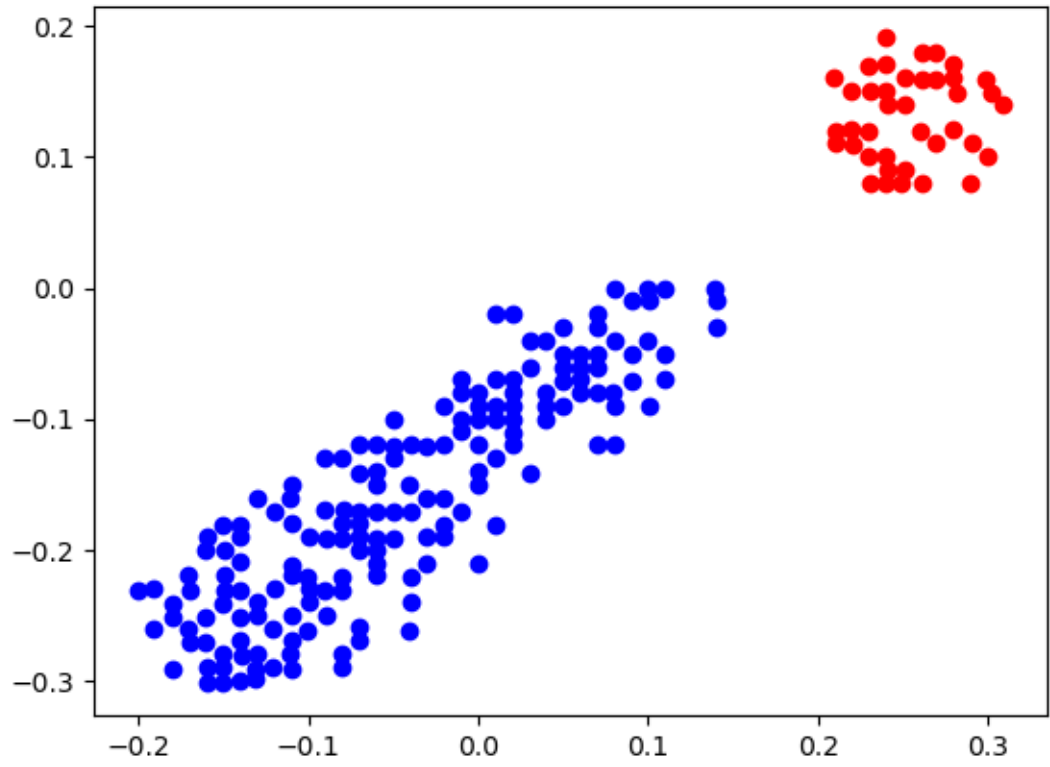

Average Linkage Criterion

Centroid linkage doesn't work well. Probably centers of cluster's doesn't distribute uniformly.
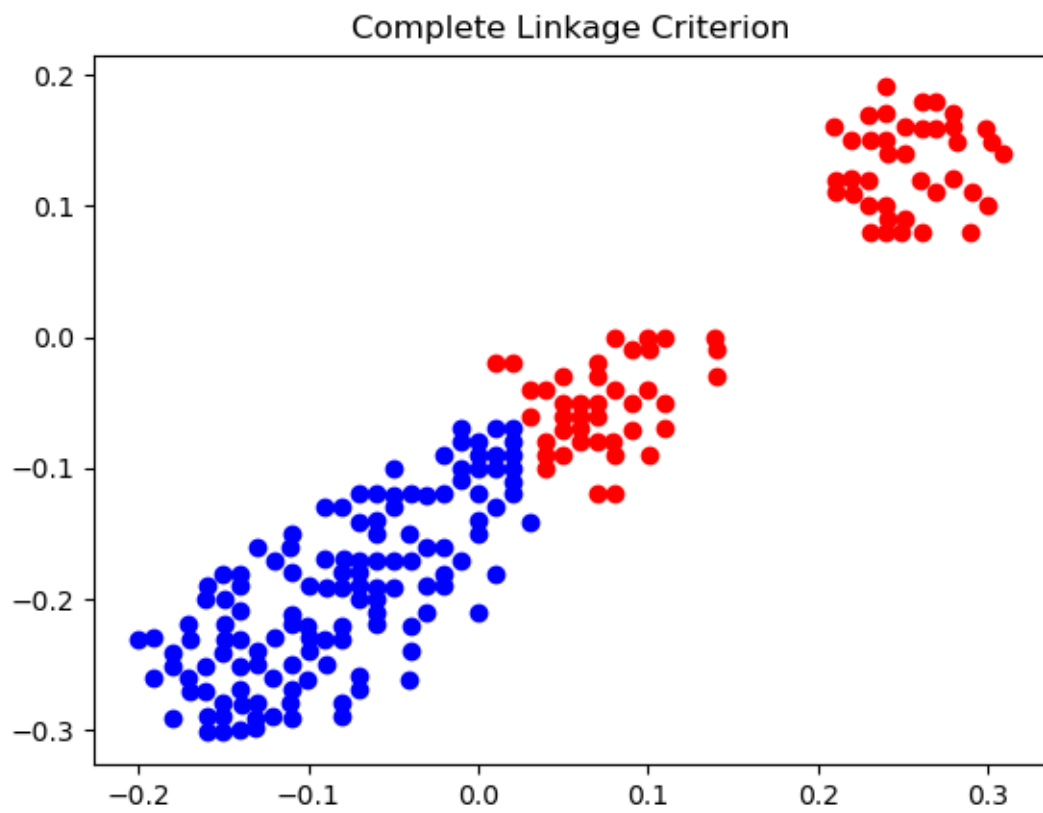


Centroid Criterion

3.3 data3

Since there is a reasonable gap between two final cluster, single linkage works
perfectly.

Since points at left down corner is farther than right up corner to middle points. Because of that, complete linkage didn't work perfectly.

Complete Linkage Criterion

Average linkage works perfectly since right up cluster is too far from other cluster.
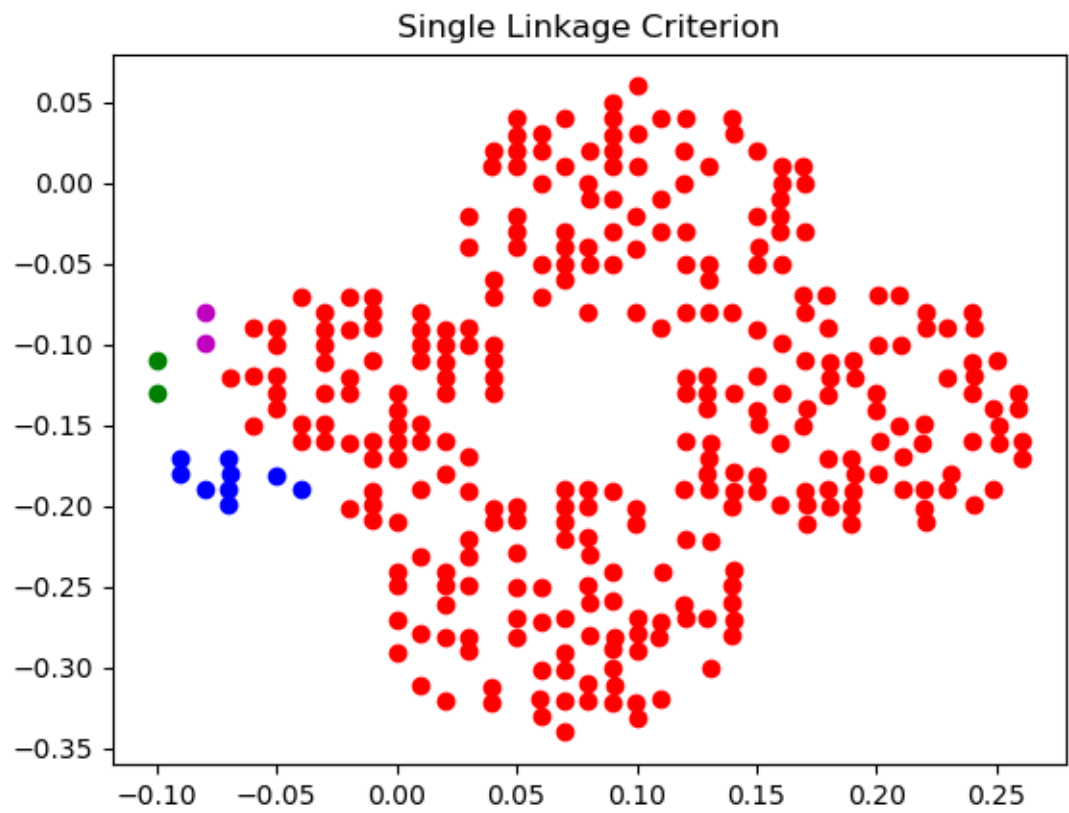

Average Linkage Criterion
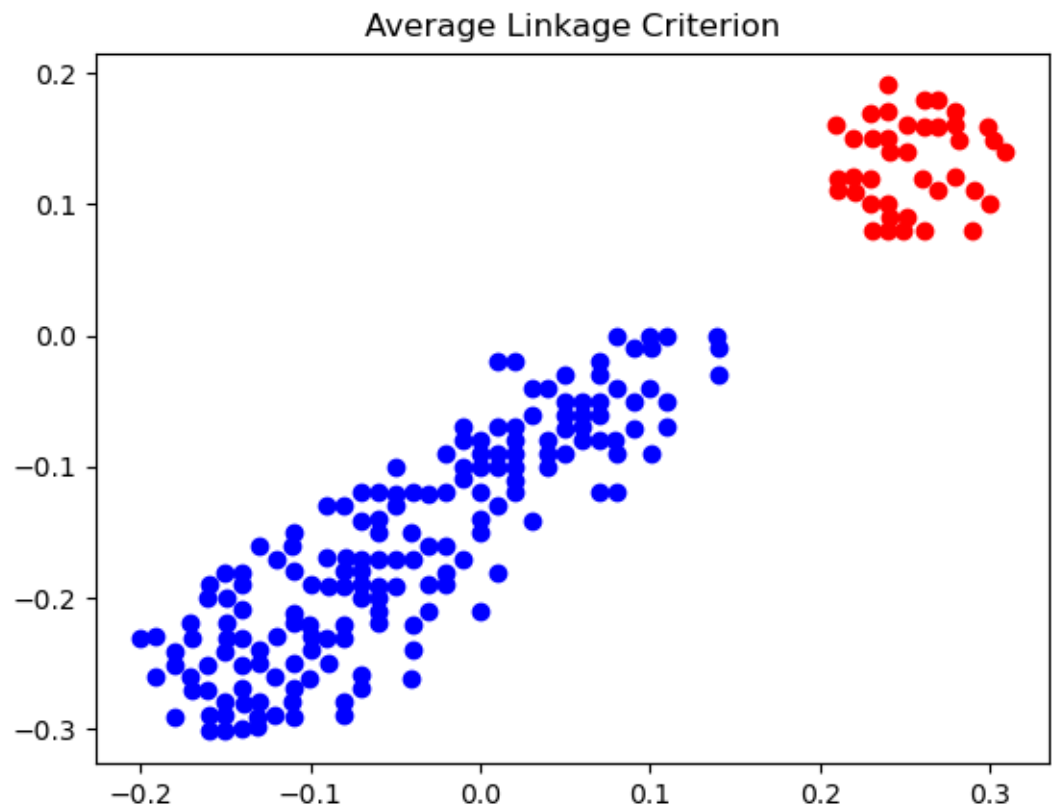
Centroids of clusters are perfectly positioning in clusters.Therefore, method works very well at data.

Centroid Criterion
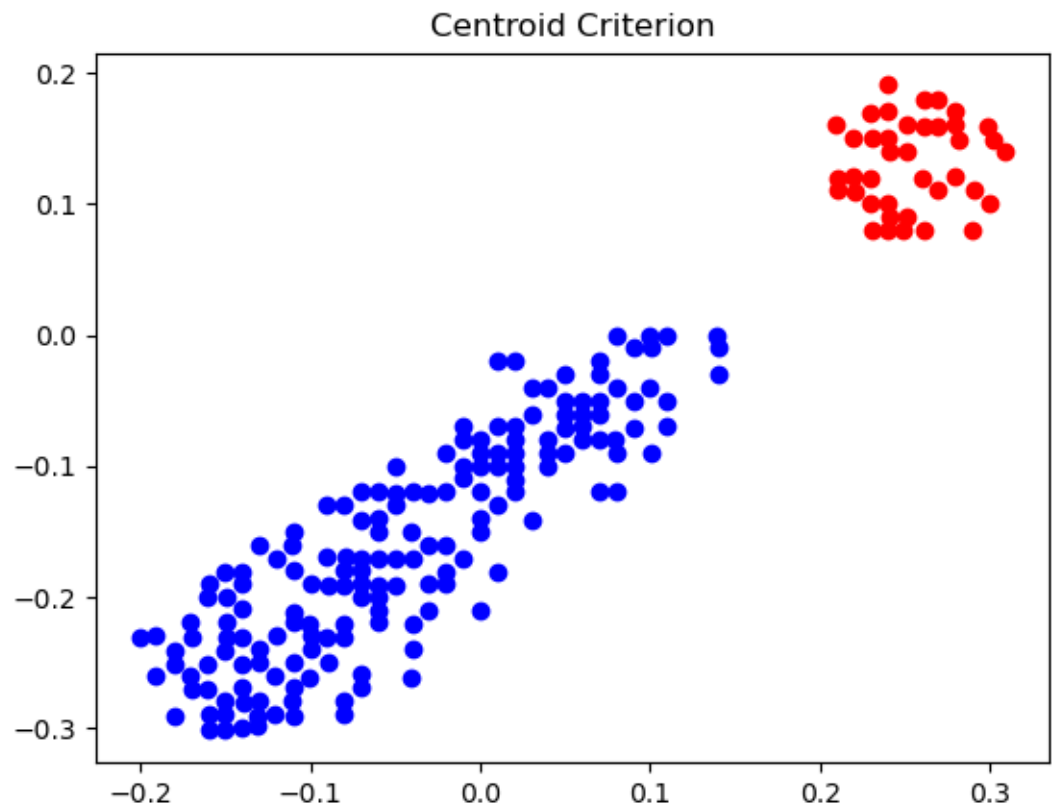
3.4 data4

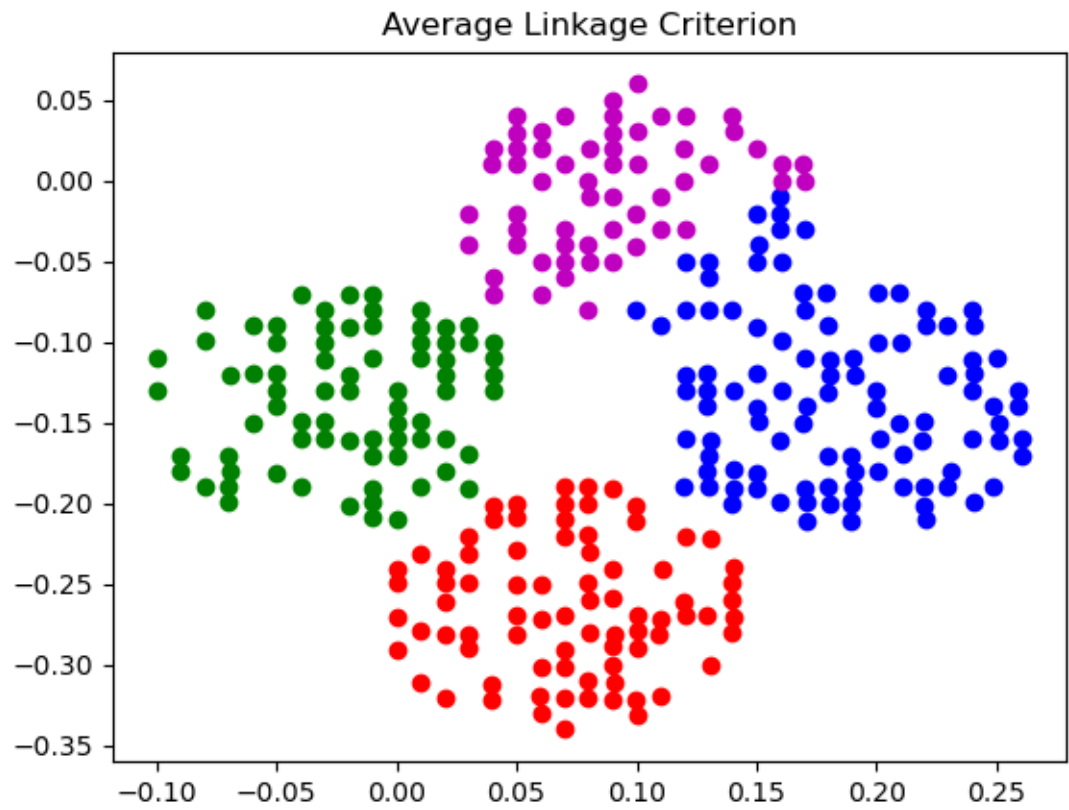Single linkage didn't work well since most of points are too close to each other between clusters.

Single Linkage Criterion

Complete linkage almost works well because farthest points of clusters are distributed well for system.

Complete Linkage Criterion
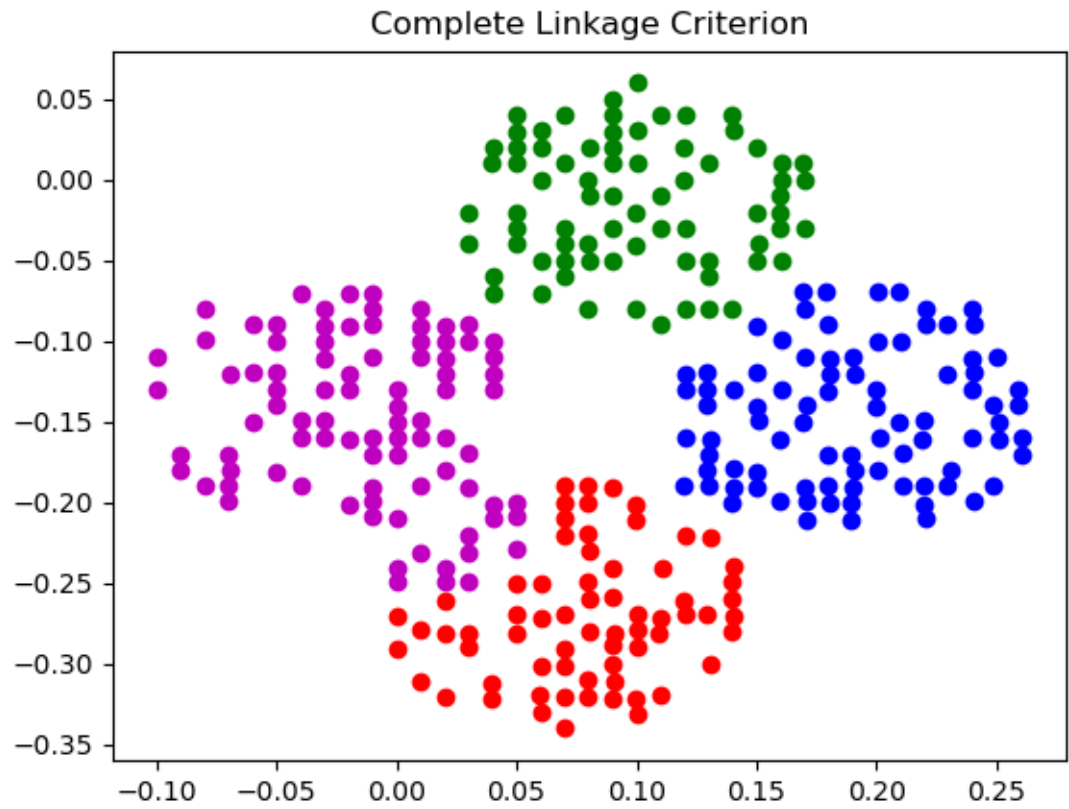
Average linkage almost works since clusters are 4 different circle.

Average Linkage Criterion

Centroid linkage works really well since clusters are 4 different circle.

Centroid Criterion