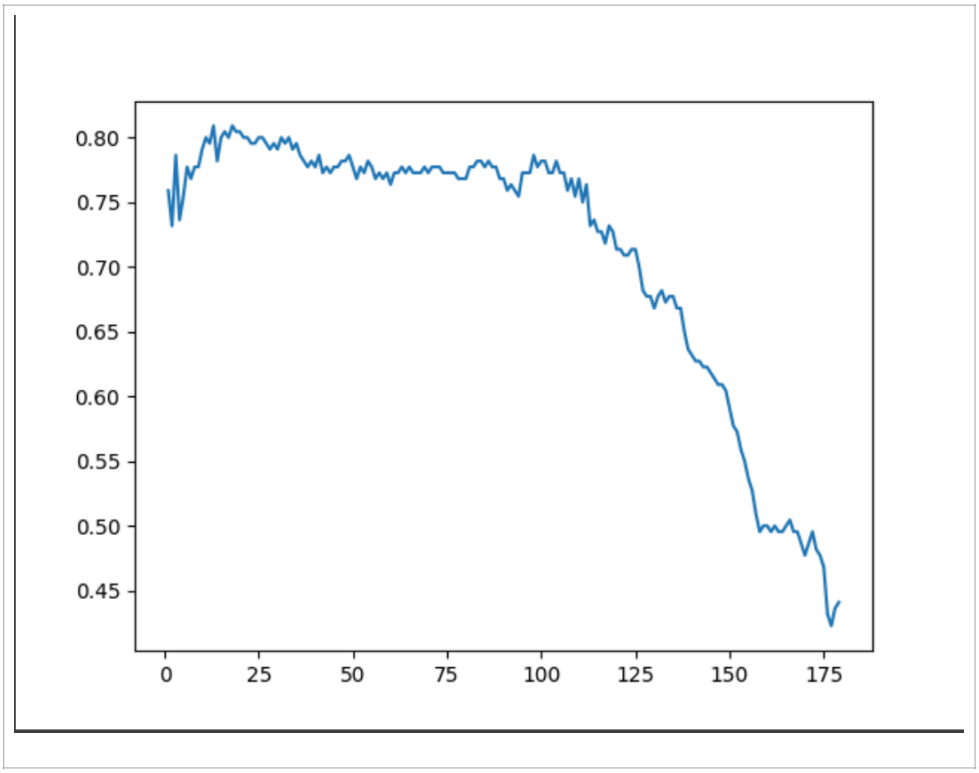
K-NEAREST NEIGHBOURS

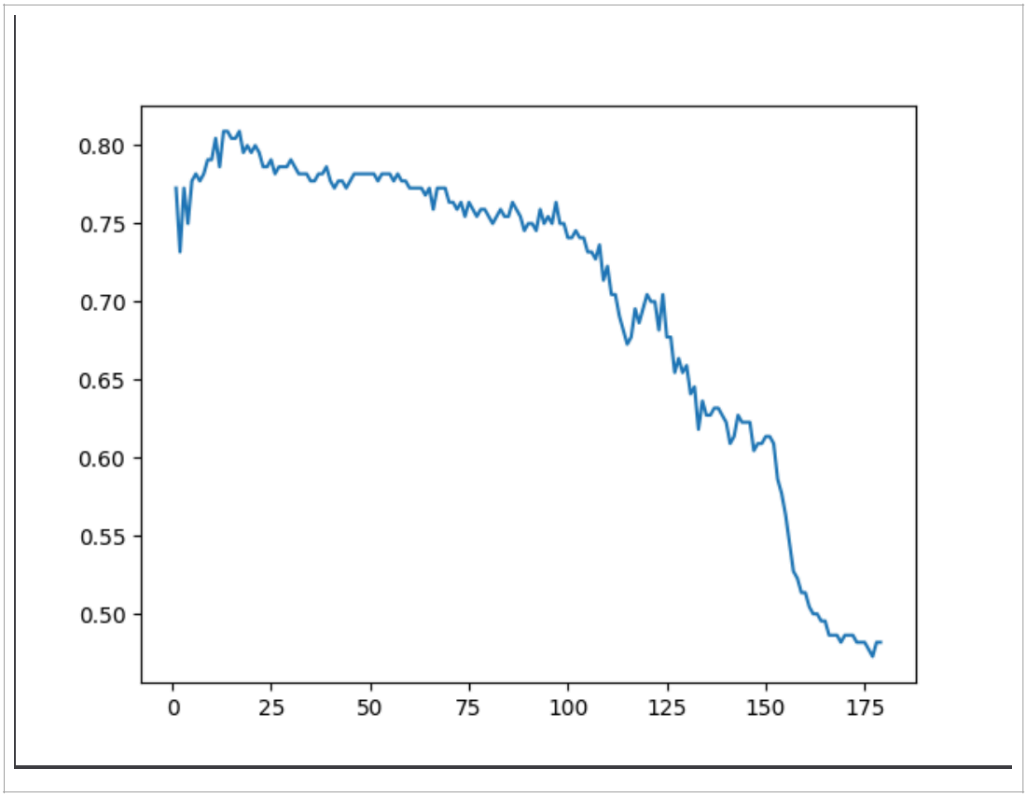
L1 DISTANCE

Best k Value = 13
Test accuracy for k = 13: 0.8388888888888889



L2 DISTANCE

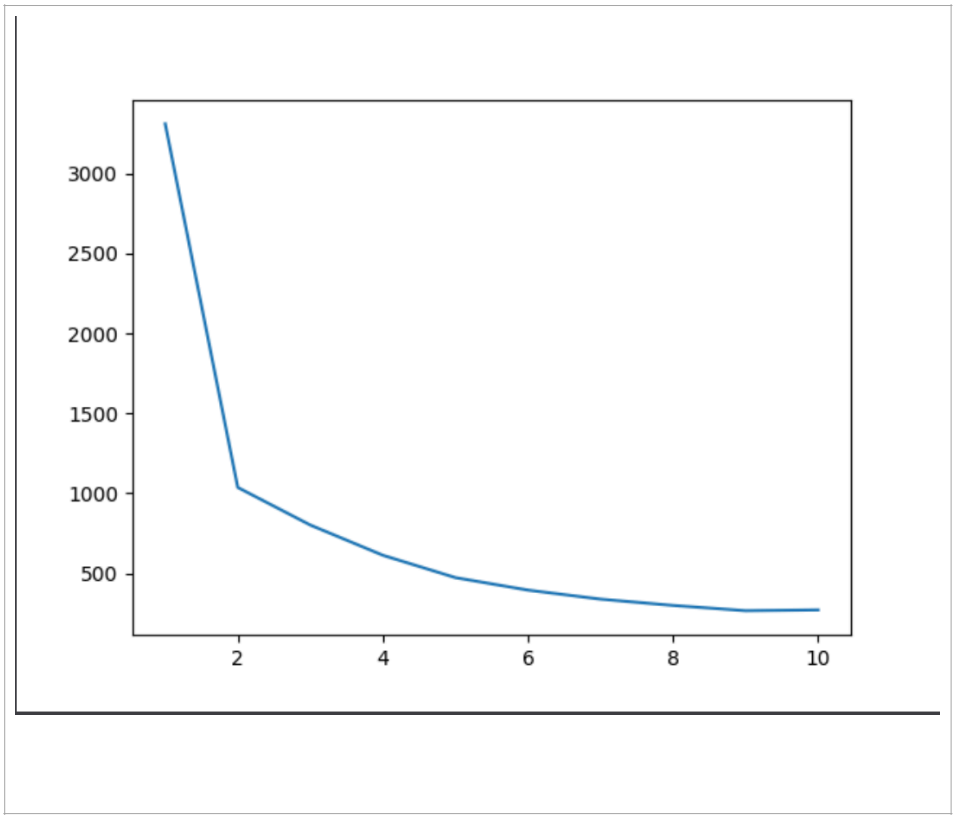
Best k value = 14
Test accuracy for k = 13: 0.8055555555555556



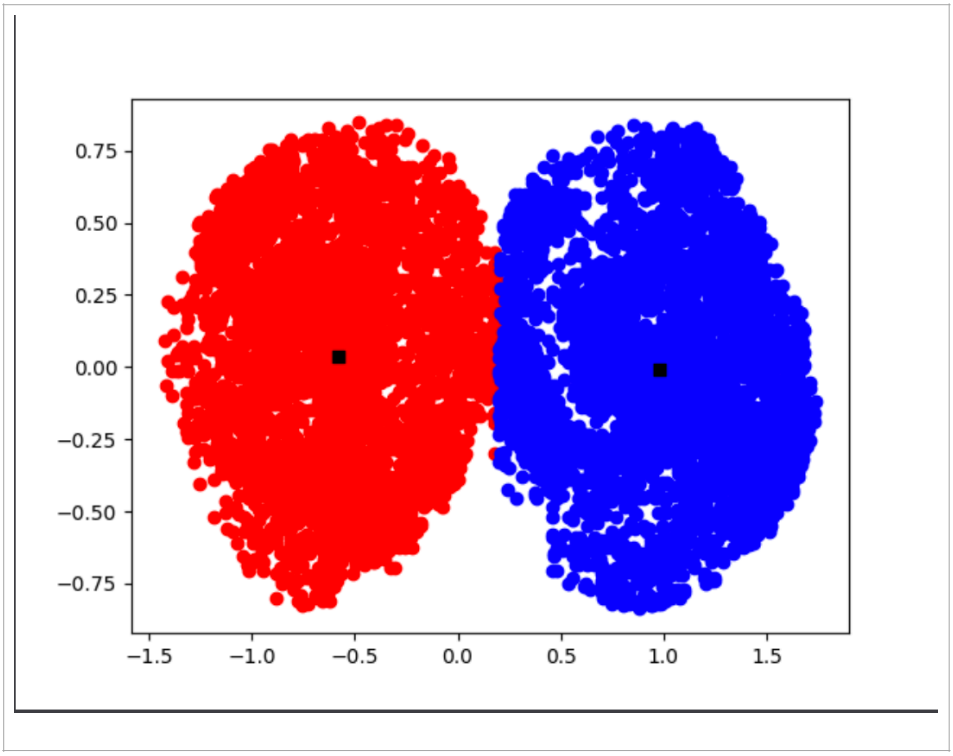
If we increase k value, obtained labels tend to be assigned to majority label in the dataset.Therefore, target labels which are in minority are labeled wrongly.After some k value, accuracy decreases dramatically.

K-MEANS CLUSTERING

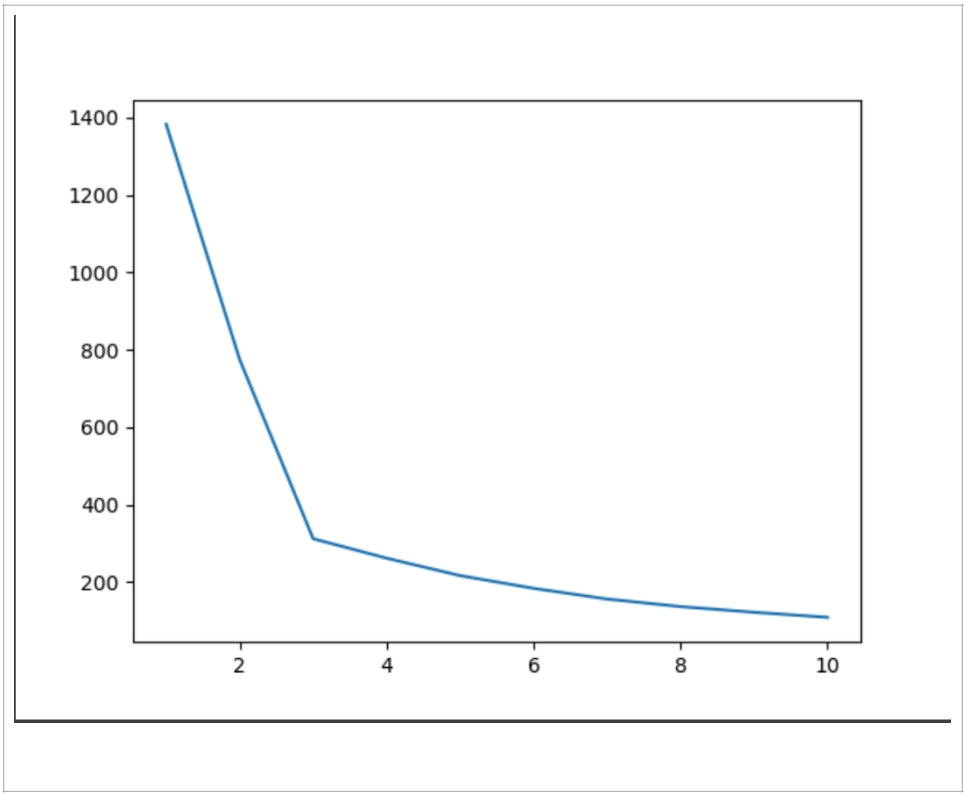
Dataset 1



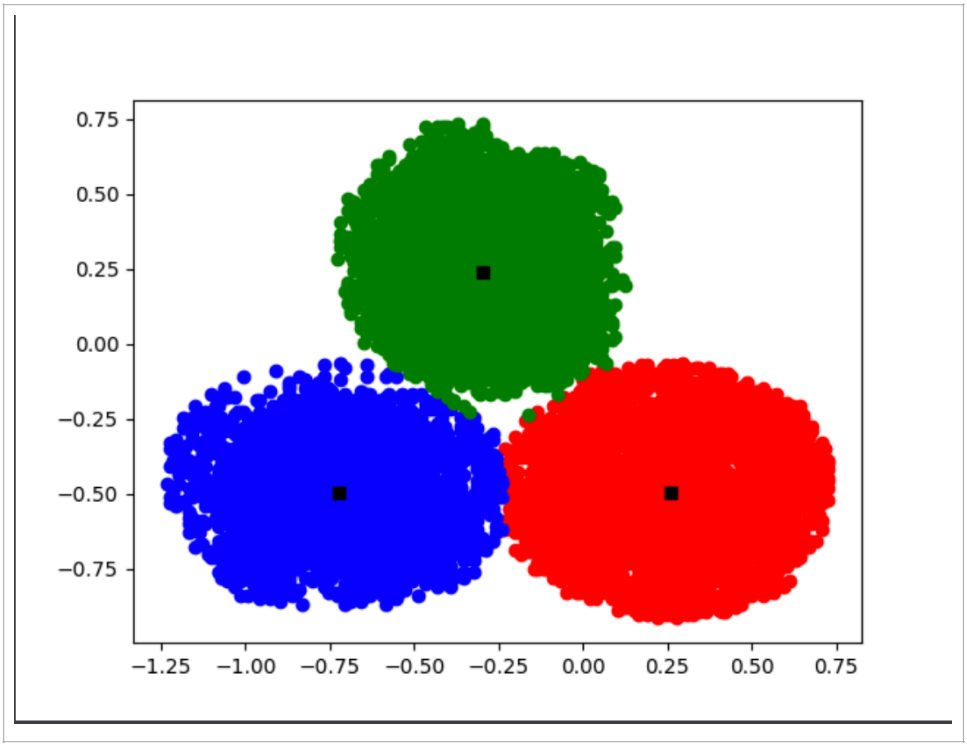
Best k is 2 in this dataset.So, if we apply K-means clustering with k = 2, we obtain:
Black squares show cluster centers.



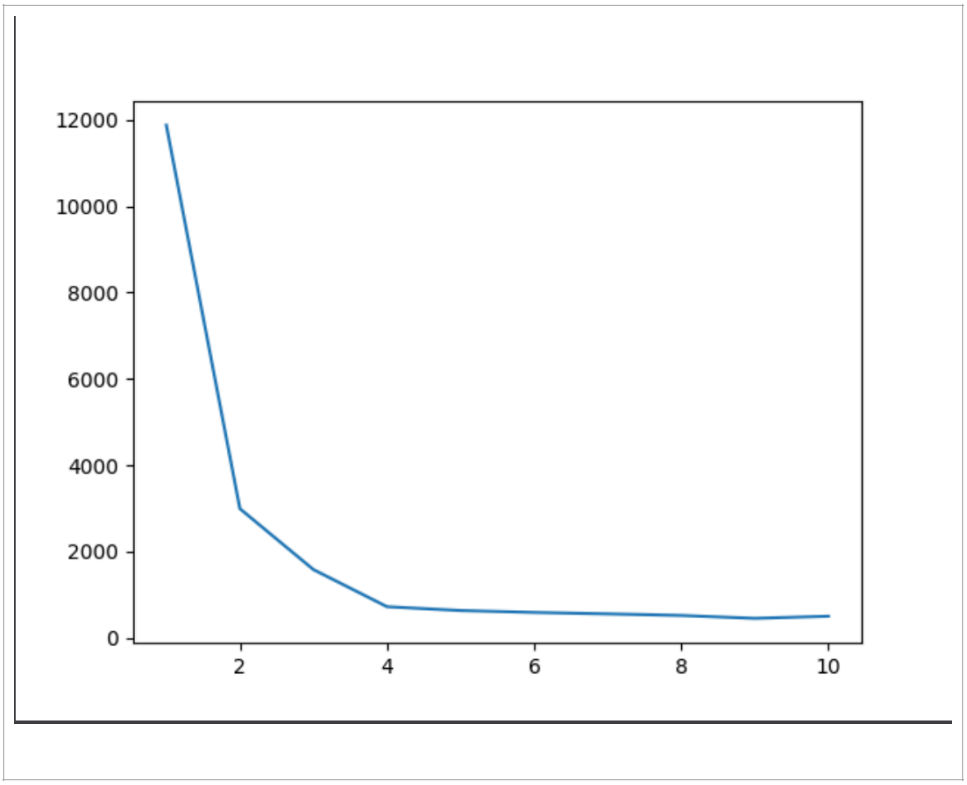
Dataset 2



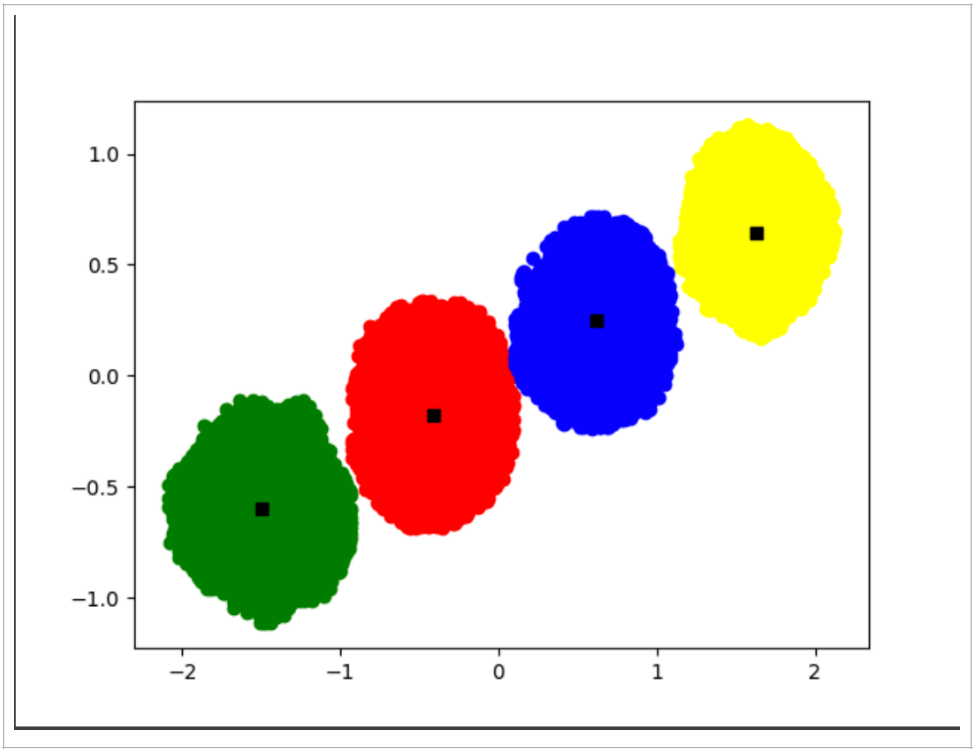
Best k is 3 in this dataset.So, if we apply K-means clustering with k = 3, we obtain:



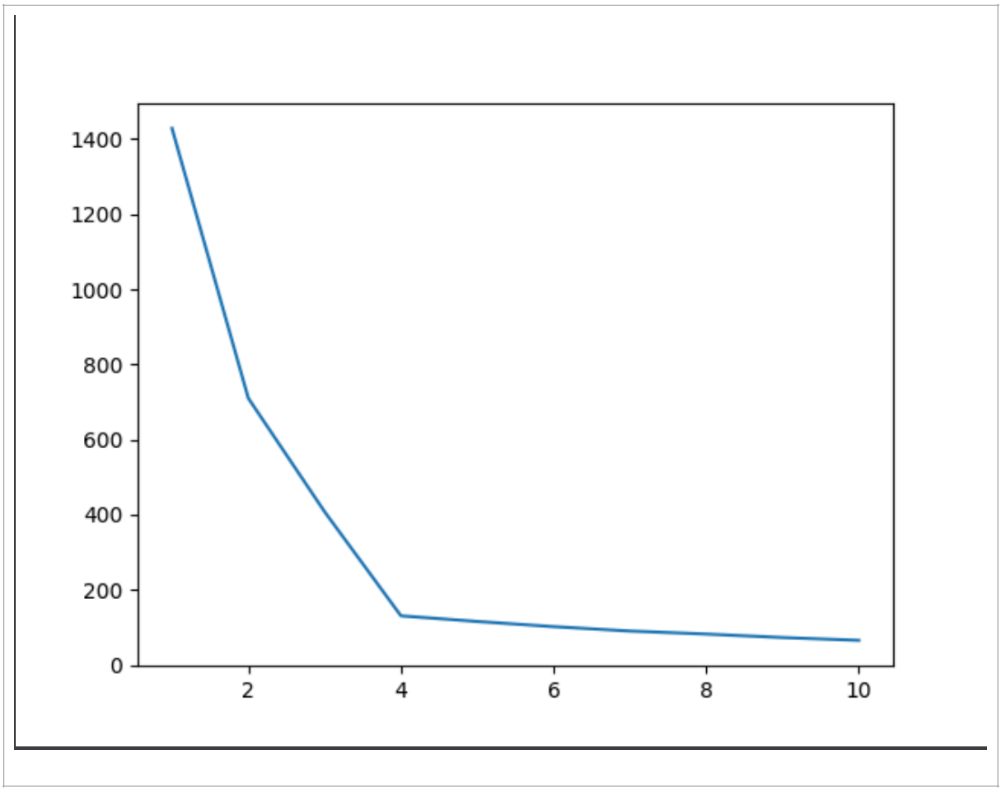
Dataset 3



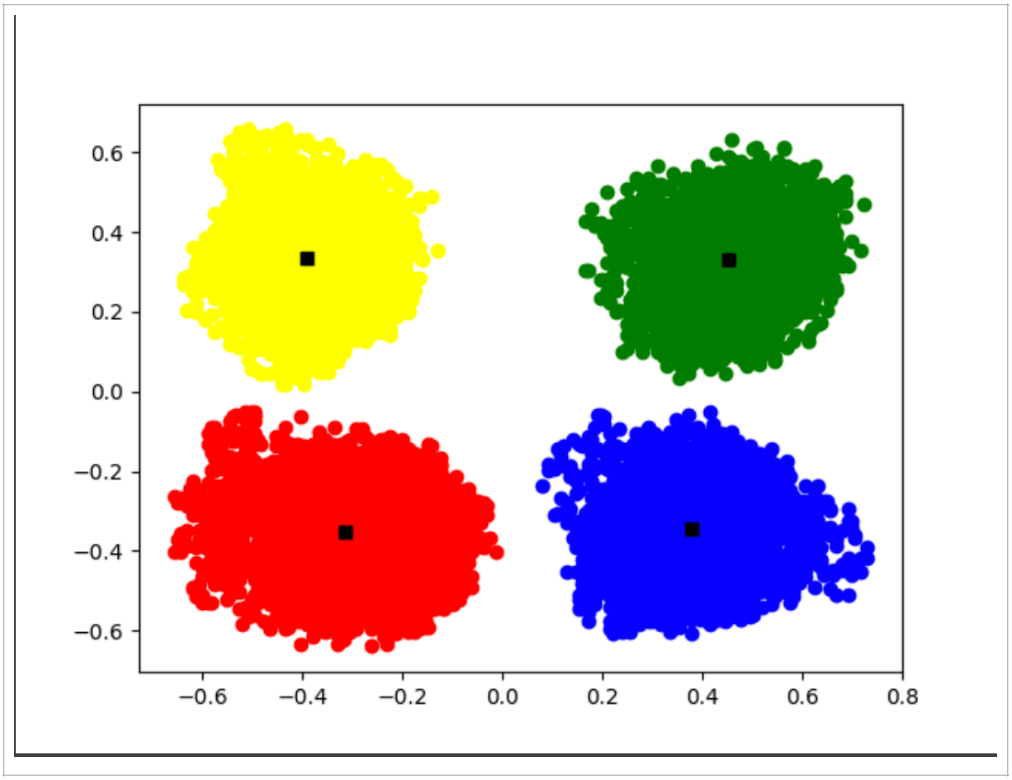
Best k is 4 in this dataset.So, if we apply K-means clustering with k = 4, we obtain:



Dataset 4



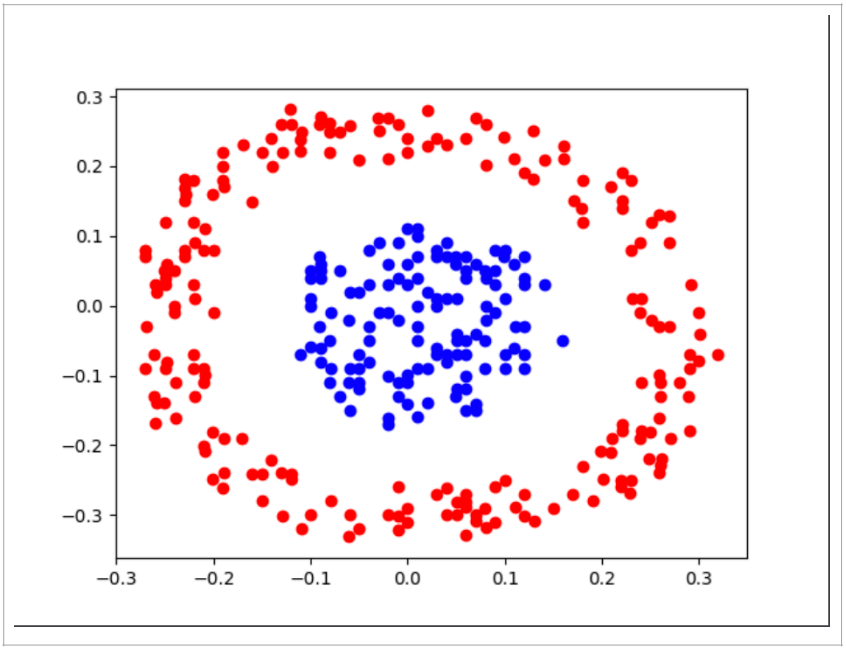
Best k is 4 in this dataset.So, if we apply K-means clustering with k = 4, we obtain:



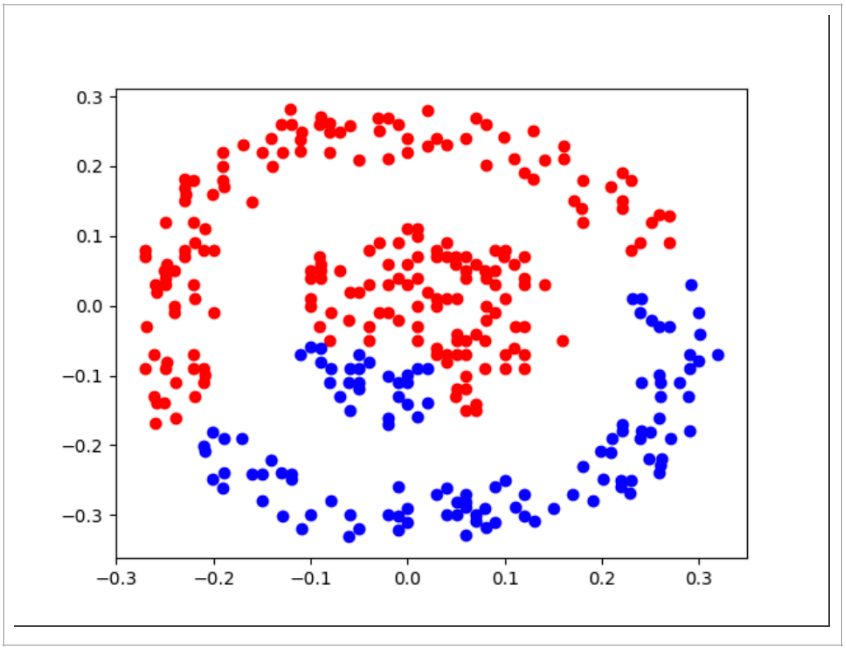
HIERARCHICAL AGGLOMERATIVE CLUSTERING

Dataset 1

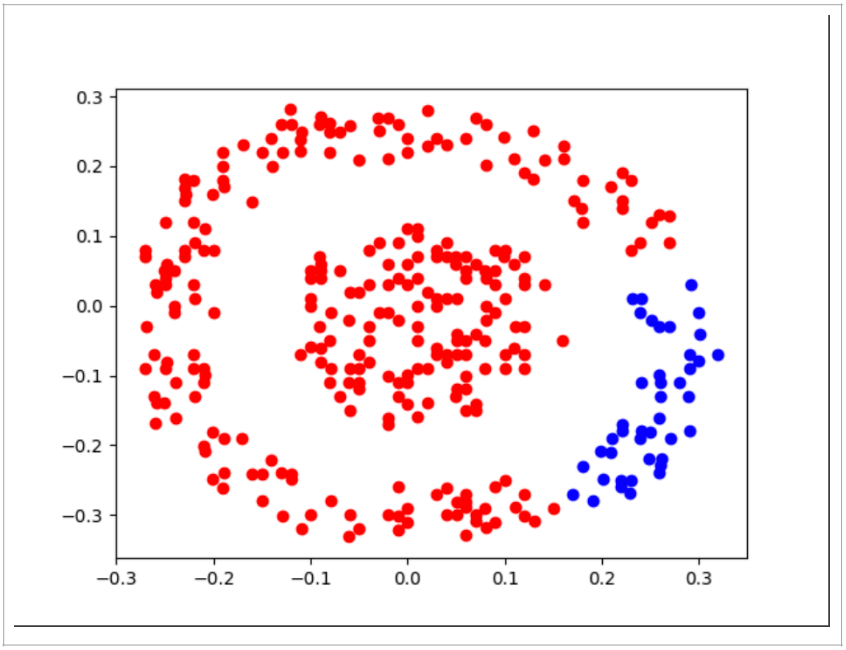
Single linkage is most successful method for this dataset which is globular data.Other methods are not proper for this dataset type.It does not have noisy datas much and data points at the center diverges from peripheral data points distinctly.So single linkage method gives more accurate cluster result.



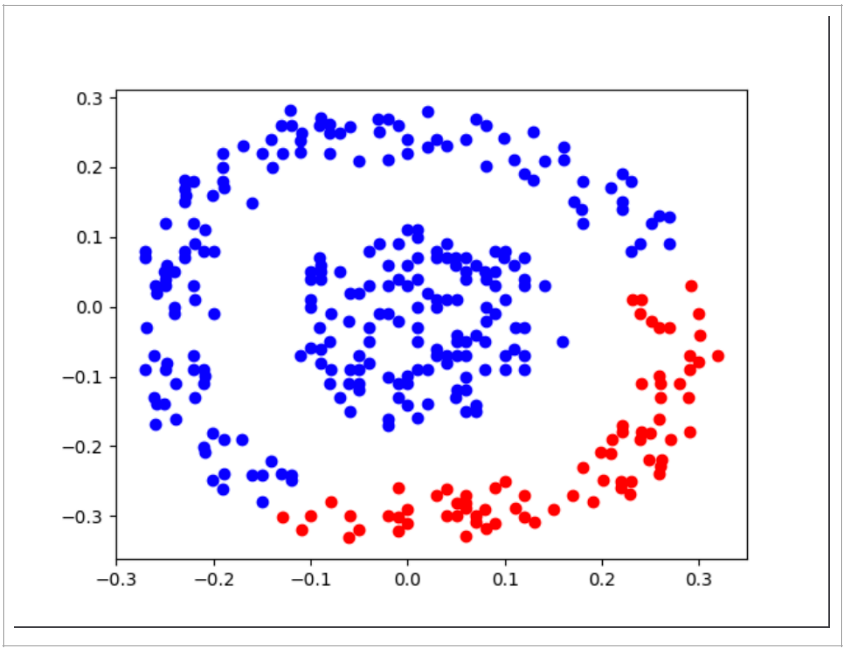
Single Linkage



Complete Linkage



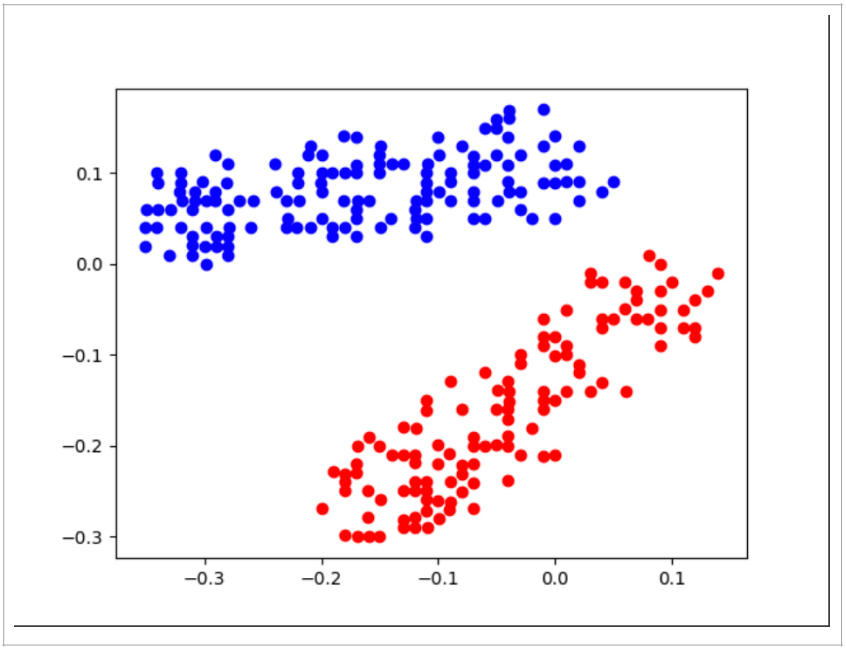
Average Linkage



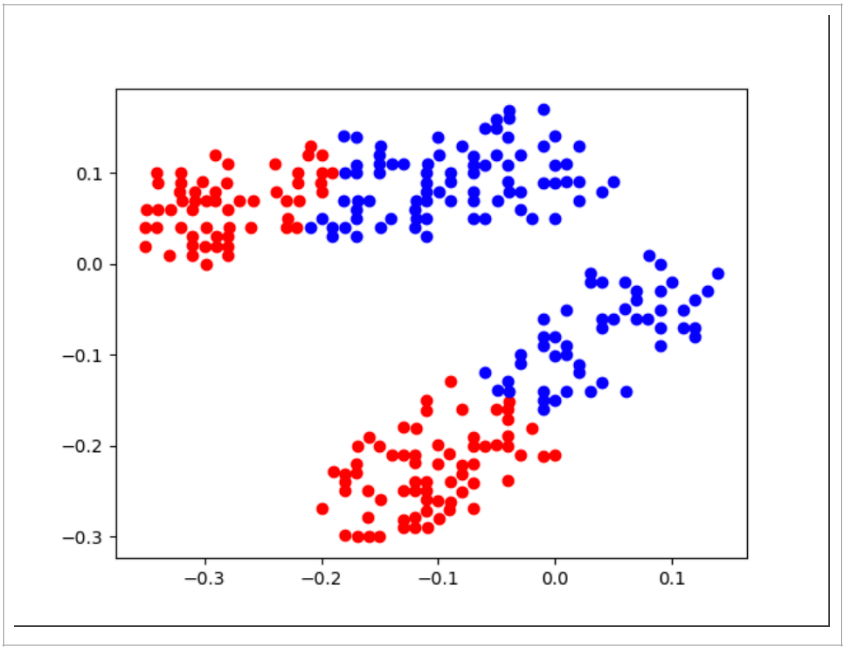
Centroid Linkage

Dataset 2

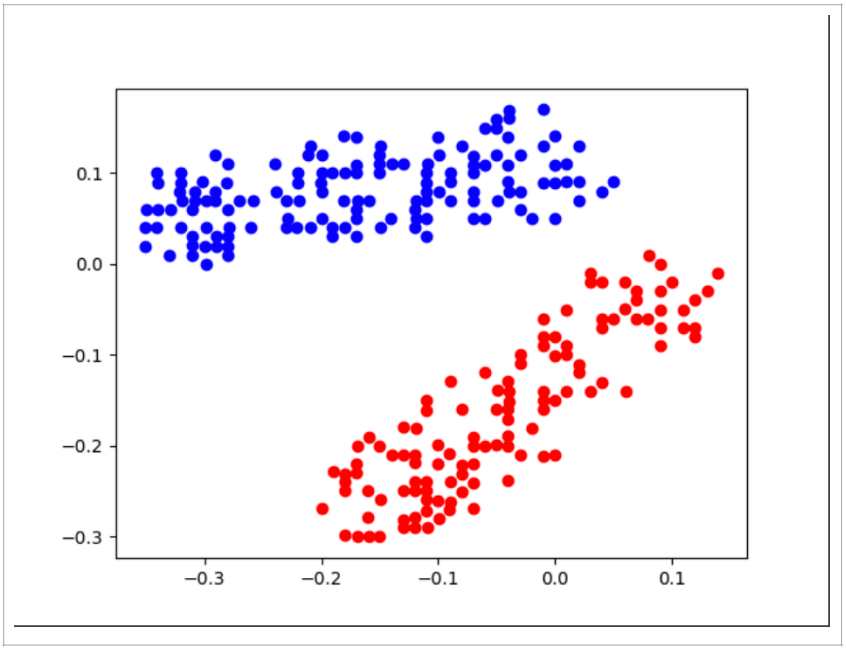
Single and average linkages determined clusters with same accuracy.Because data points are not distributed equally on the lower cluster, upper points of this cluster were chosen by upper cluster in centroid linkage method.In complete linkage, clusters were chosen symmetrically and upper part of upper cluster and lower part of lower cluster formed a cluster due to their distances are much than other parts.Therefore this two methods are not suitable for this dataset.



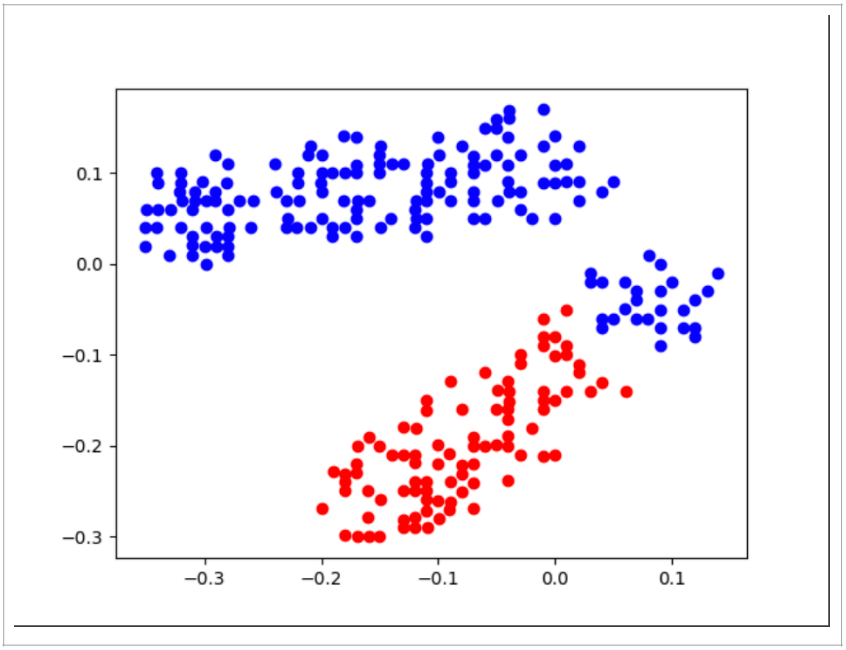
Single Linkage



Complete Linkage



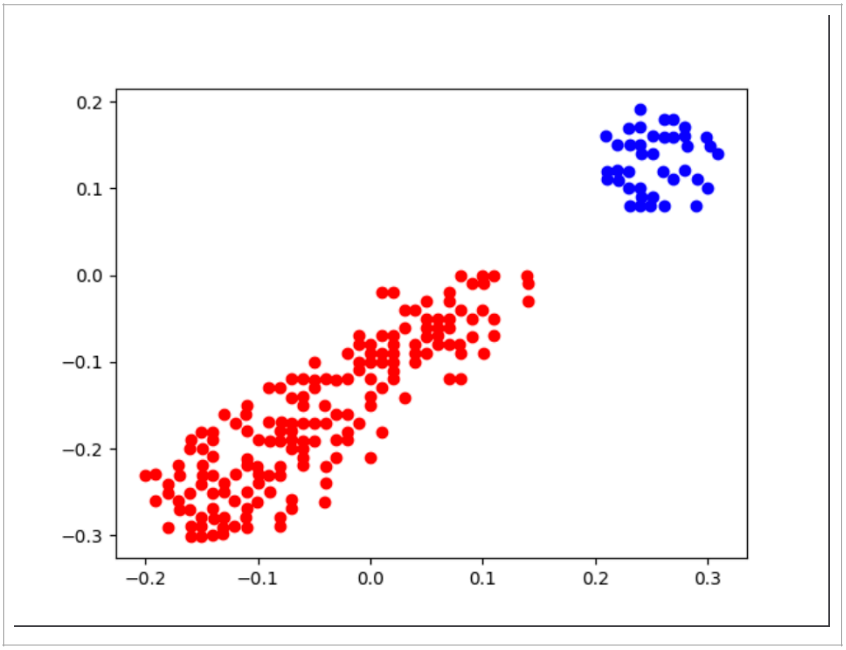
Average Linkage



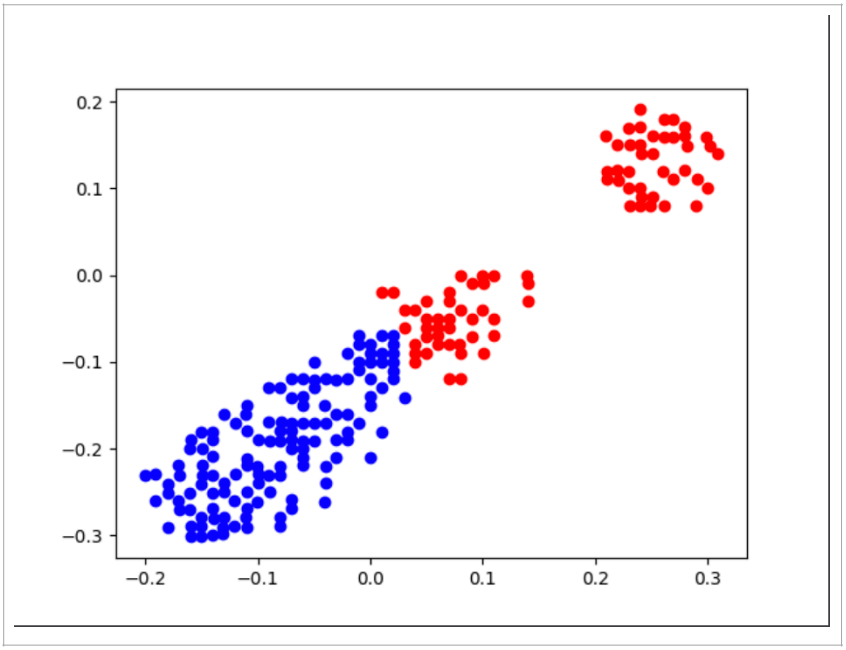
Centroid Linkage

Dataset 3

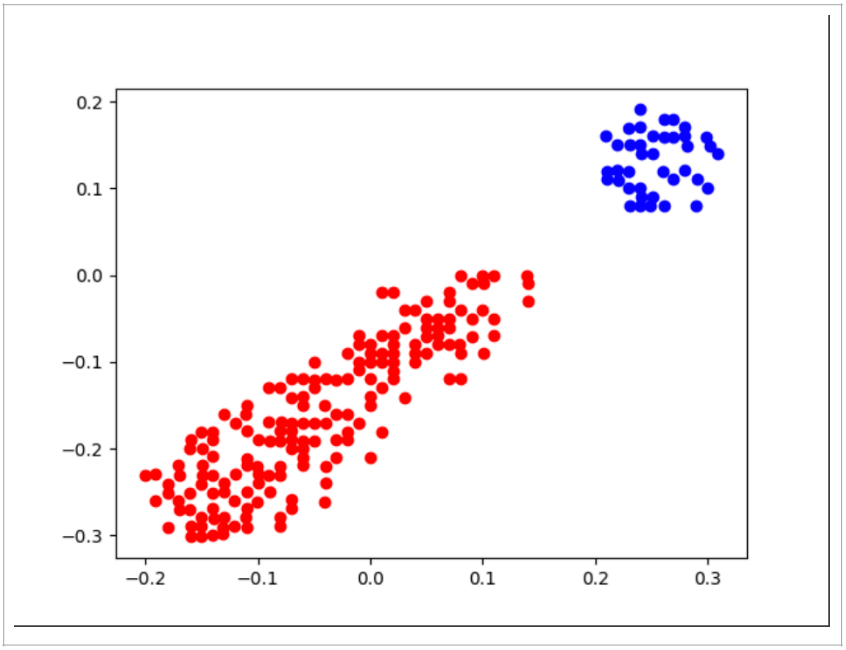
In this dataset, clusters are diverged from each other and distances are significant.Single, average and centroid linkage methods determined clusters correctly.Complete linkage method tends to assign clusters symmetrically and this makes distribute data points equally.Therefore, complete linkage is no suitable for this dataset.



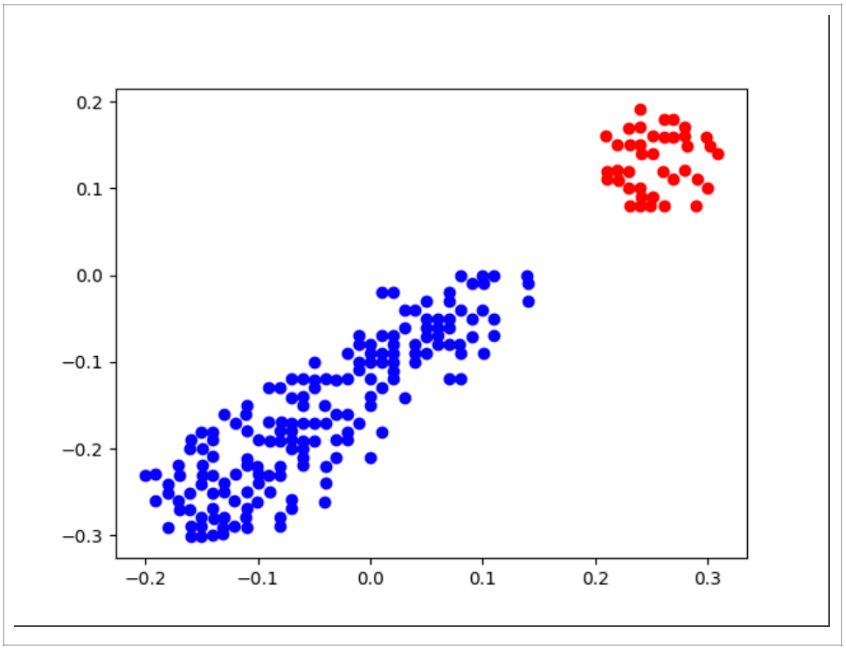
Single Linkage



Complete Linkage



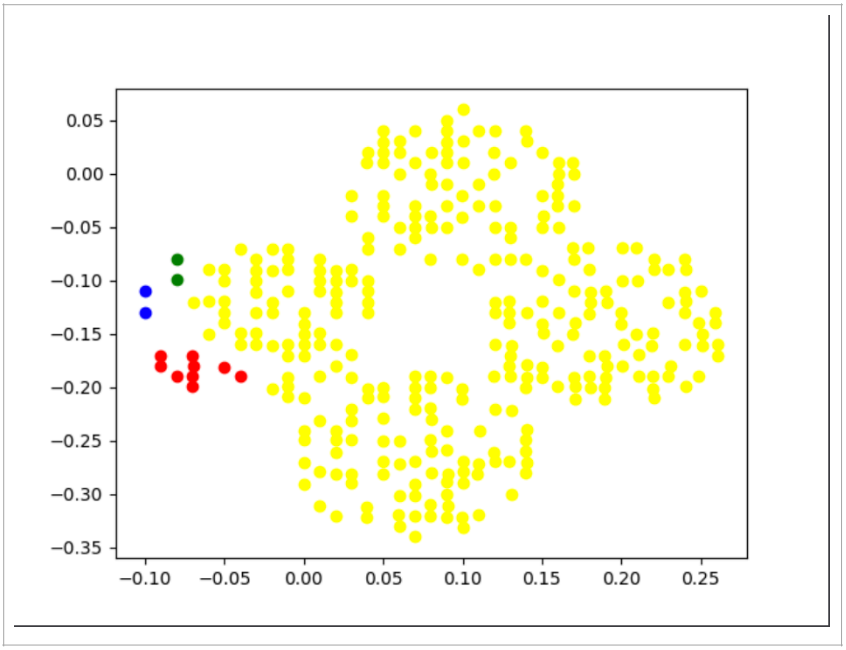
Average Linkage



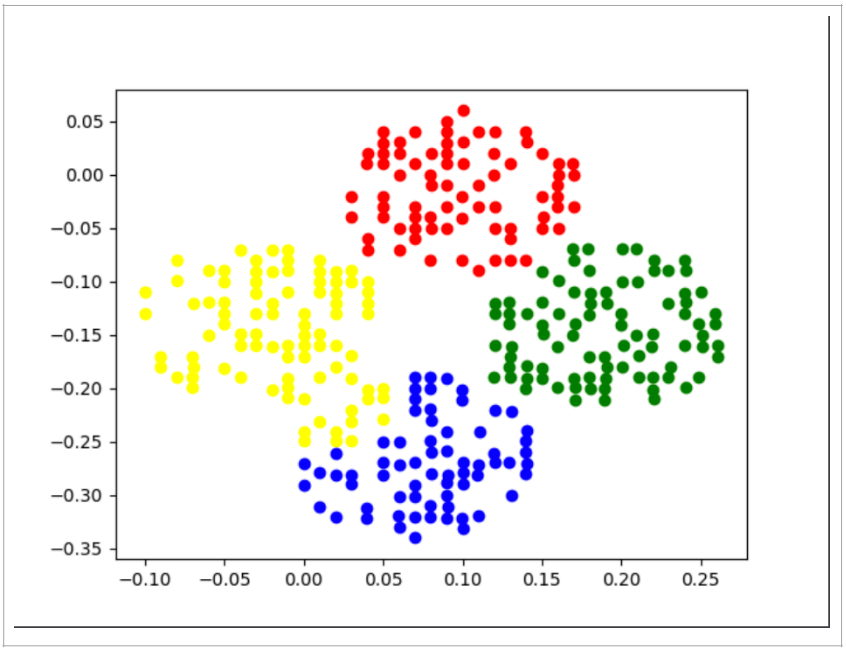
Centroid Linkage

Dataset 4

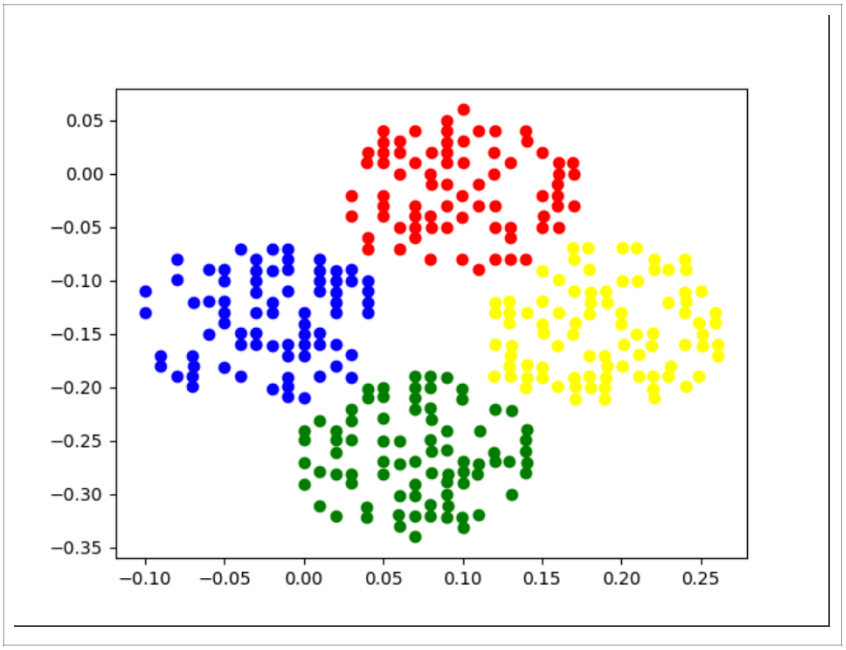
Due to noisy data points on the left side, single linkage method fails to distinguish cluster in this dataset.It seems that centroid and average linkage methods have almost same accuracy.



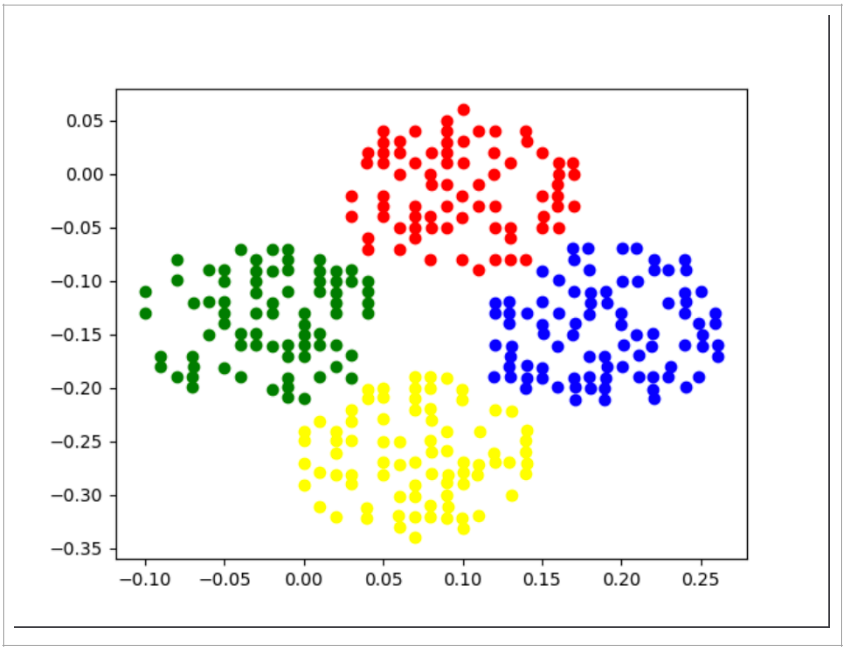
Single Linkage



Complete Linkage



Average Linkage



Centroid Linkage