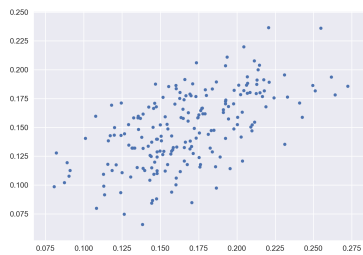
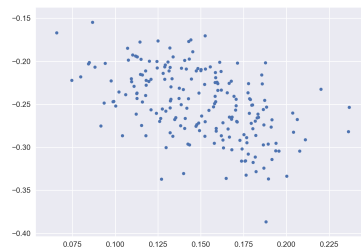


Problem 7

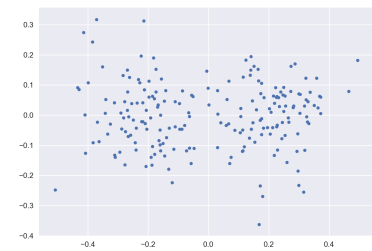
Part A



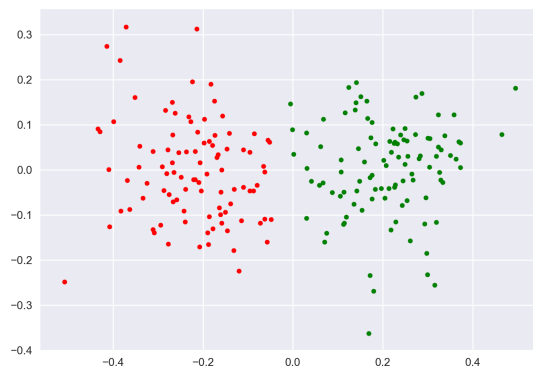
1



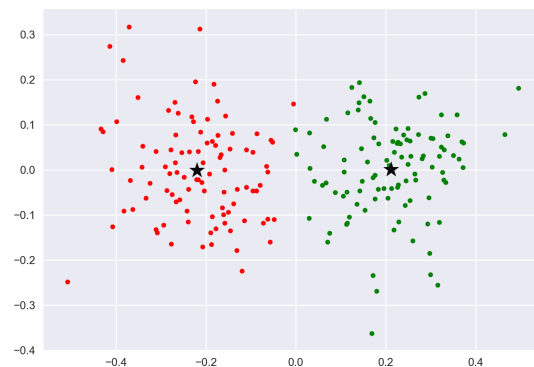
2



3



4



5

1) 2)

No specific structure can be seen from the plots. Because the plots only contain 2/40 of all the information in original data and all other information is lost.

3)

It can be seen on the plot that data are approximately in 2 clusters. Because although PCA generate 2-d points as well, it contains most of the original information by selecting the principal component.

4) 5)

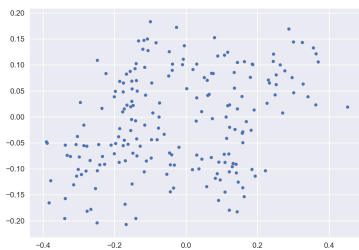
From the plots, we can see that both approaches with/without PCA are successful in clustering. Because PCA preserves the most important information. K-Means using 2/40 dimensions both gets the general structure of the original data.

6)

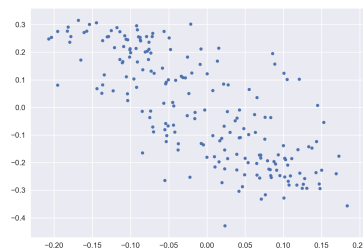
4) and 5) results only differ in one point close to the boundary.

For general problems, PCA can represent original data with lower-dimension and provide the principal components which implies the structure of data.

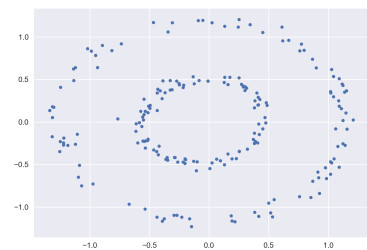
Part B



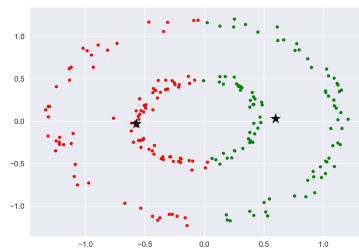
1



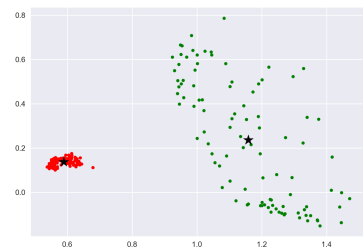
2



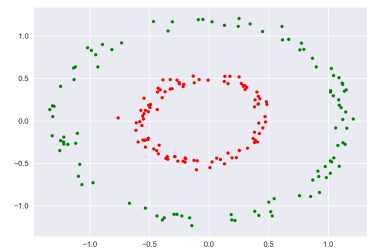
3



4



5



6

1) 2)

We are unable to tell the structure of data from plot 1,2. The reasons are the same as Part A.

3)

It can be seen that the data are grouped in 2 circular clusters.

4)

K-means fails after PCA because the two clusters are non-linear and not linear separable.

5) 6)

KPCA and Spectral Clustering succeed in recovering the true clustering of the data. Because KPCA and Spectral Clustering are able to separate the original non-linear data.