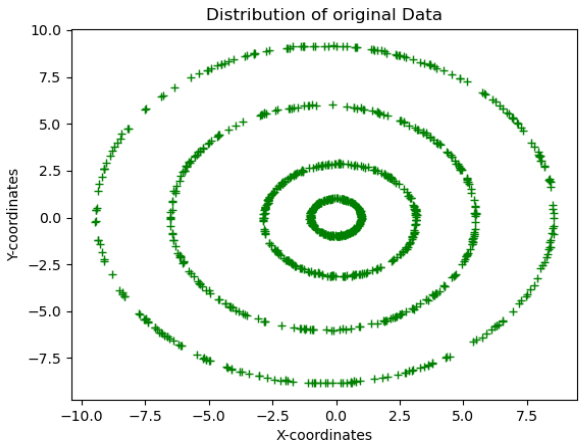
Name : Aashay Shah **Roll No : CS22M004**

PRML Assignment 1

Q.1) You are given a dataset with 1000 data points each in R2.

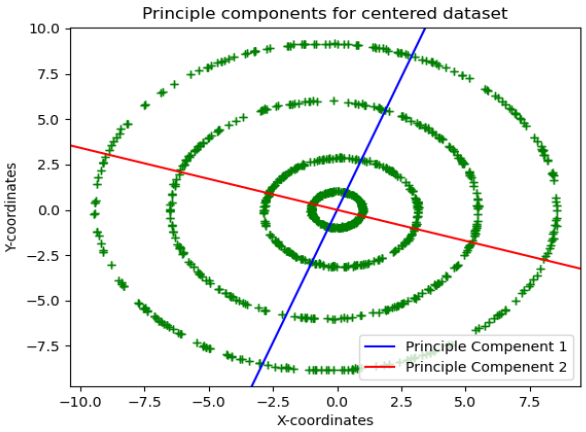
Data Distribution



1.) Write a piece of code to run the PCA algorithm on this dataset. How much of the variance in the dataset is explained by each of the principal components ?

Here, first of all, we center the dataset, and then apply the PCA algorithm.

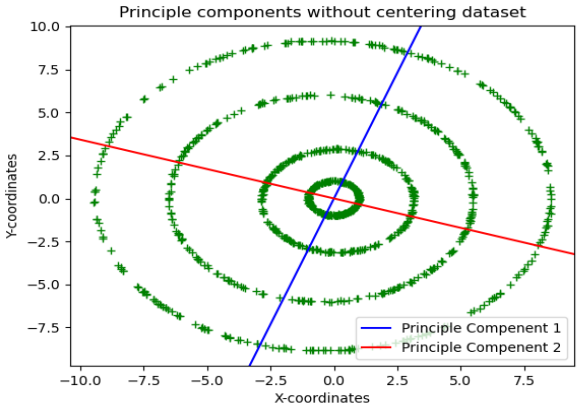
- For centering the dataset, we will find the mean corresponding to each of the dimensions and then subtract it from each data point. x' = x – mean(x), y' = y – mean(y)
- Variance explained by each of the principal components (eigenvector) is calculated by the formula : (Eigenvalue of the corresponding eigenvector) / (sum of all eigenvalues)
- Sum of all eigenvalues : 31.62151915177499
- Maximum eigenvalue : 17.1319144
- So, the variance explained by the first eigenvector is : 54.178024528852234 %
- The second maximum eigenvalue is : 14.489604749330631
- The variance explained by the second eigenvector is : 45.82197547114777 %



2.) Study the effect of running PCA without centering the data-set. What are your observations? Does Centering help?

In the second question Let's say we apply the PCA algorithm without centering our dataset.

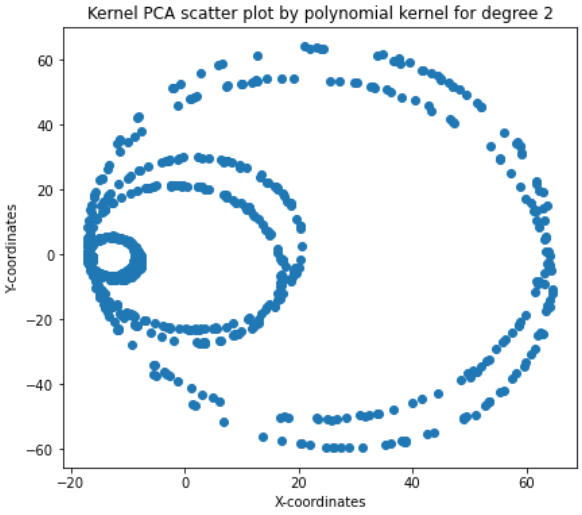
- The eigenvalues we got, in this case, are : 17.1319144 and 14.489604749330631
- The eigenvalues we got, in this case, are the same as we got when we applied the centering technique. This happened because the data we are provided is having the mean values corresponding to both the dimensions are : 4.075000e-07 and 2.227000e-07
- The initial dataset provided to implement PCA is in almost centered form. So, if we apply the centering technique or not, it will not affect the result much and the eigenvectors we got, are also the same as in the case we applied the centering technique. By that, the variance explained by both the eigenvectors are also of the same percentage as in case 1.

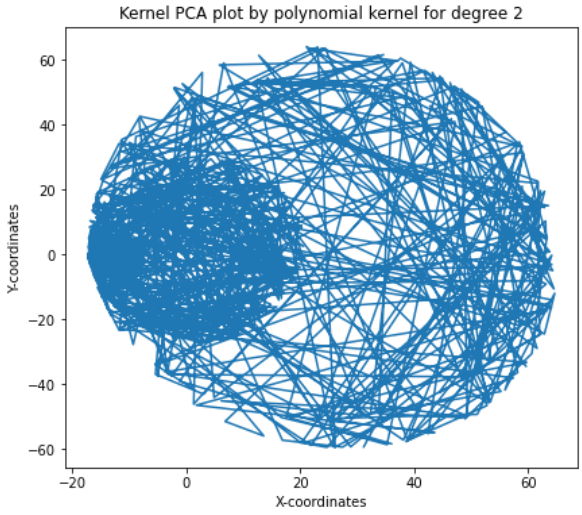


3.A) k(x,y) = (x.T*y + 1)d for d={2, 3}**

a) degree = 2

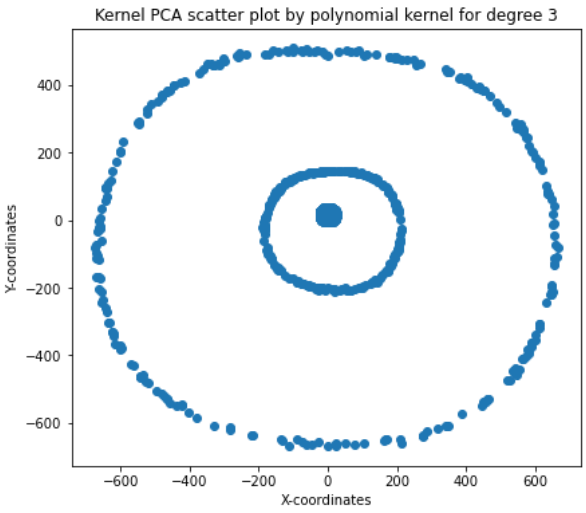
Graph of polynomial kernel function for degree 2 is as follows :

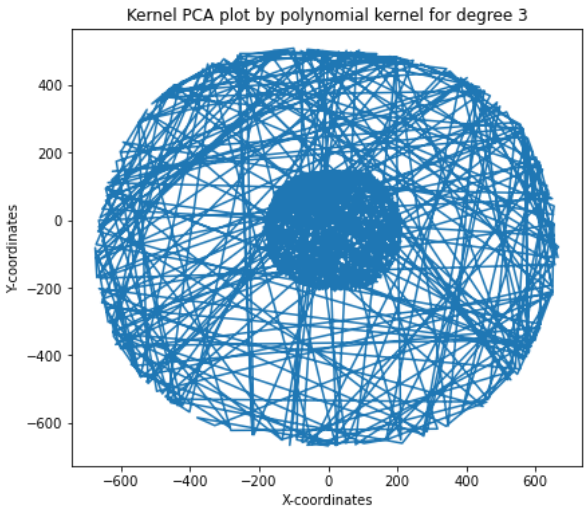




b) degree = 3

Graph of polynomial kernel function for degree 2 is as follows :





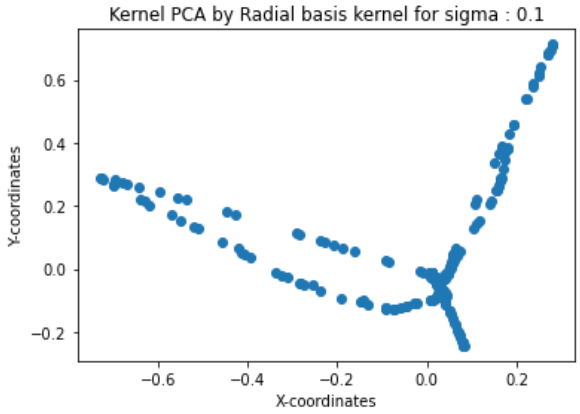
As the degree of polynomial function increases, its graph will contain more curvatures.

Conclusion : From the above two graphs, we can say that function with degree 3 is having more similar shape than the function with degree 2. So it is capturing more information from our original dataset and shows a similar cluster pattern.

3.B) exp((-(x-y).T * (x-y)) / 2*(sigma)2)**

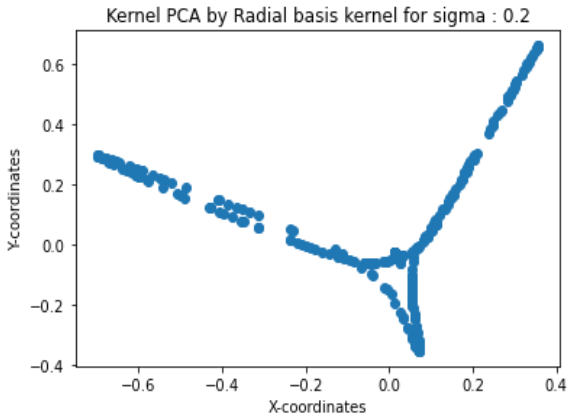
a) Sigma = 0.1

Graph of Radial Basis kernel function for sigma 0.1 is as follows :



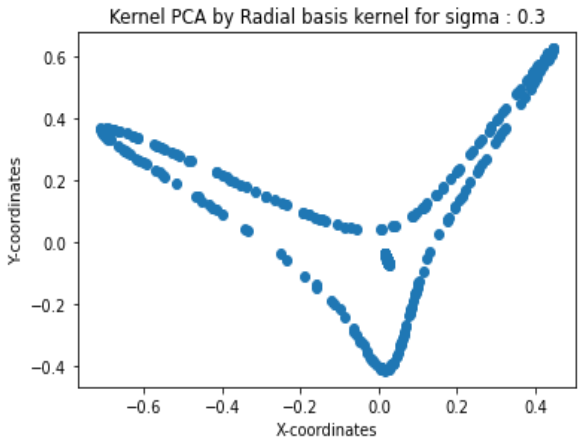
b) Sigma = 0.2

Graph of Radial Basis kernel function for sigma 0.2 is as follows :



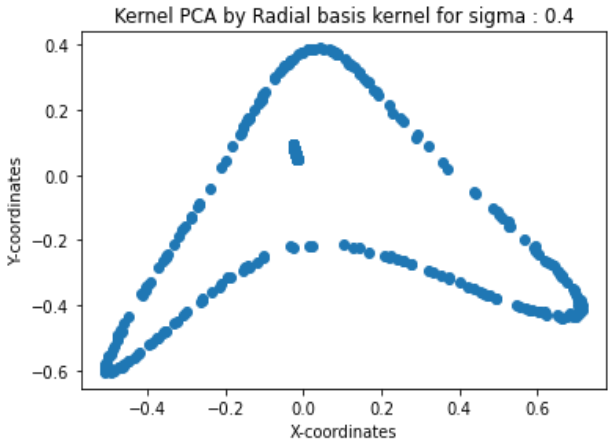
c) Sigma = 0.3

Graph of Radial Basis kernel function for sigma 0.3 is as follows :



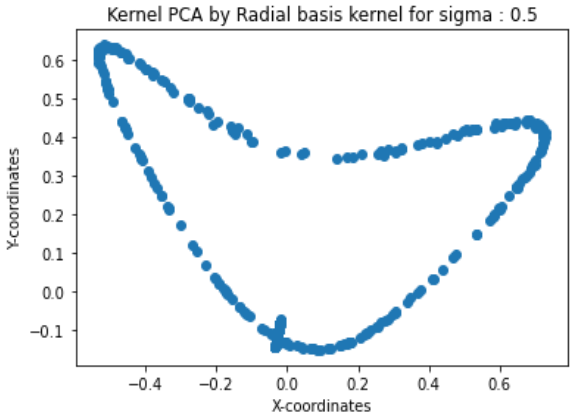
d) Sigma = 0.4

Graph of Radial Basis kernel function for sigma 0.4 is as follows :



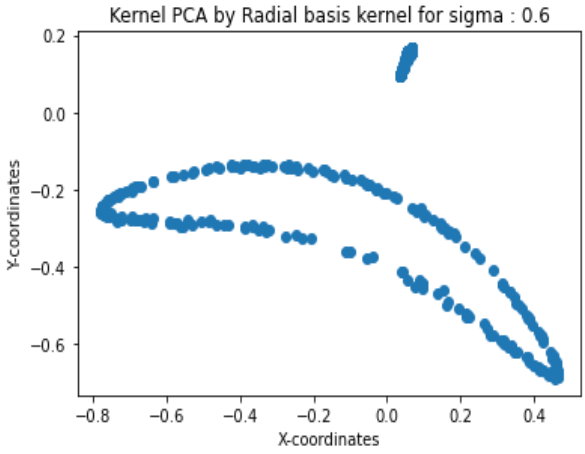
e) Sigma = 0.5

Graph of Radial Basis kernel function for sigma 0.5 is as follows :



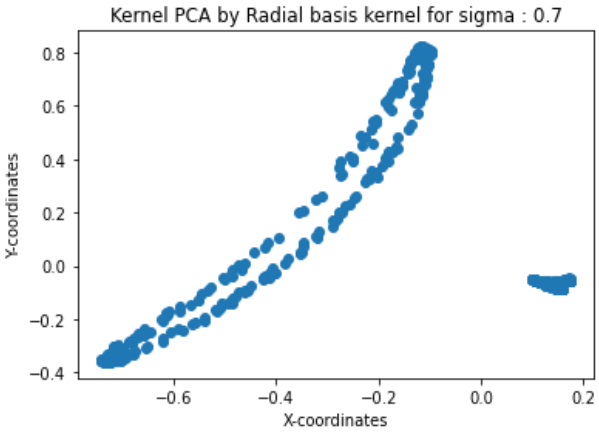
f) Sigma = 0.6

Graph of Radial Basis kernel function for sigma 0.6 is as follows :



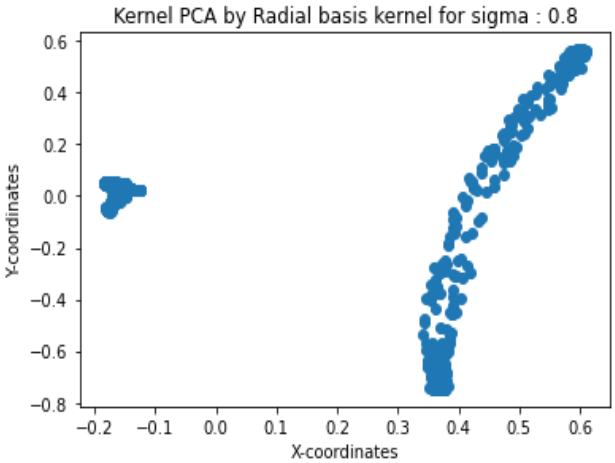
g) Sigma = 0.7

Graph of Radial Basis kernel function for sigma 0.7 is as follows :



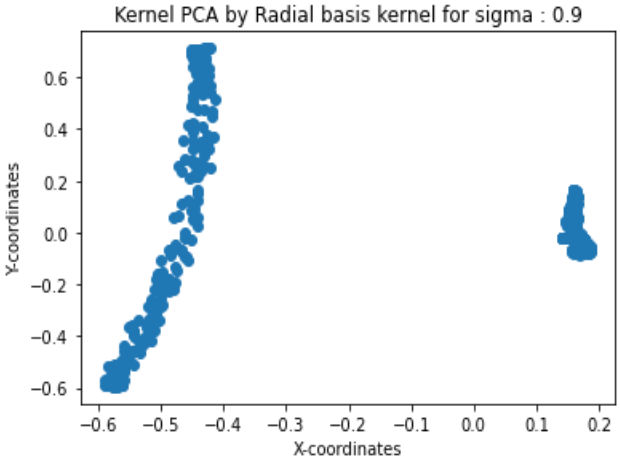
h) Sigma = 0.8

Graph of Radial Basis kernel function for sigma 0.8 is as follows :



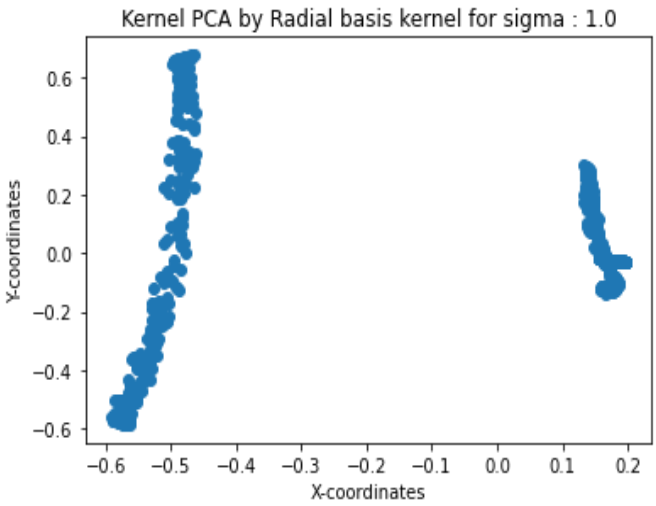
i) Sigma = 0.9

Graph of Radial Basis kernel function for sigma 0.9 is as follows :



j) Sigma = 1.0

Graph of Radial Basis kernel function for sigma 1.0 is as follows :



Conclusion : From the above graphs we can say that the original structure is being refined as we increased the value of sigma.

4.) Which Kernel do you think is best suited for this dataset and why ?

Variance observed by the polynomial kernel with degree 2 is : 68.53 %

Variance observed by the polynomial kernel with degree 3 is : 73.41 %

Variance observed by Radial basis kernel with sigma 0.1 is : 6.44 %

Variance observed by Radial basis kernel with sigma 0.2 is : 4.56 %

Variance observed by Radial basis kernel with sigma 0.3 is : 2.45 %

Variance observed by Radial basis kernel with sigma 0.4 is : 8.11 %

Variance observed by Radial basis kernel with sigma 0.5 is : 9.62 %

Variance observed by Radial basis kernel with sigma 0.6 is : 11.09 %

Variance observed by Radial basis kernel with sigma 0.7 is : 12.46 %

Variance observed by Radial basis kernel with sigma 0.8 is : 13.64 %

Variance observed by Radial basis kernel with sigma 0.9 is : 14.68 %

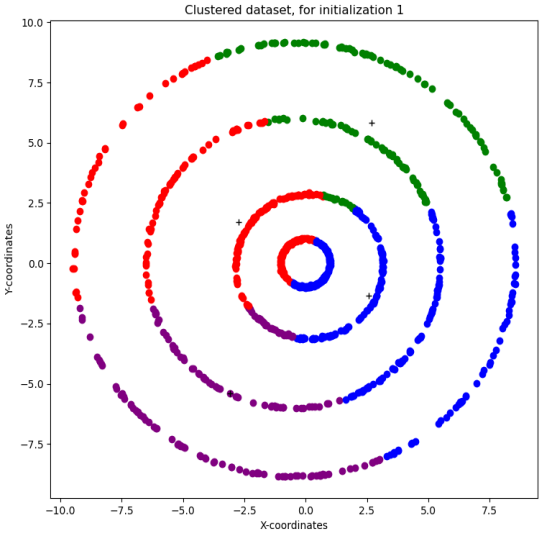
Variance observed by Radial basis kernel with sigma 1.0 is : 15.64 %

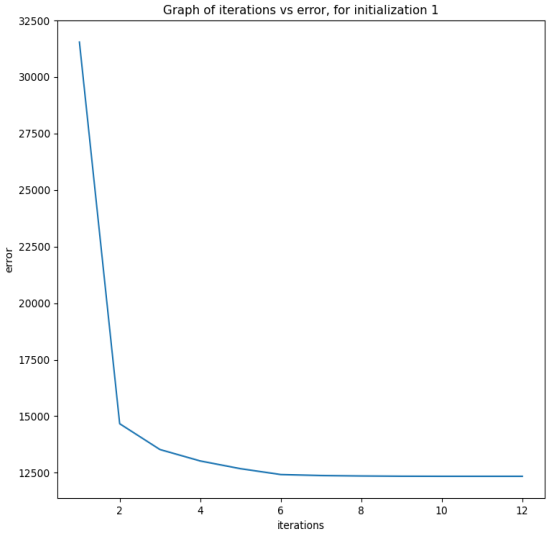
From the above observations we can say that the highest variance is obtained by the polynomial kernel with degree 3 = 73.41 %, So this is the best suitable kernel for the given dataset.

Q.2) You are given a data-set with 1000 data points each in R2

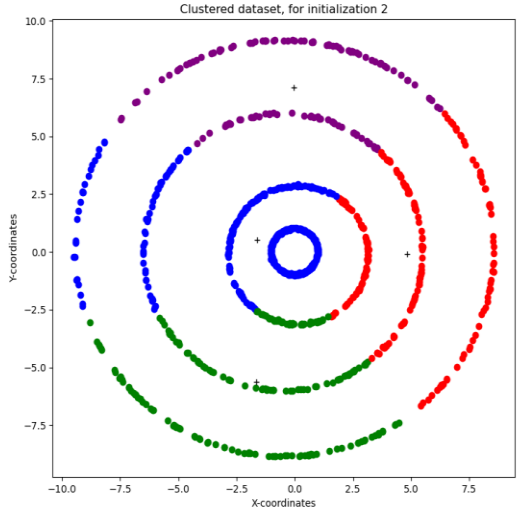
1.) K-means clustering algorithm for 5 different initialization for k = 4

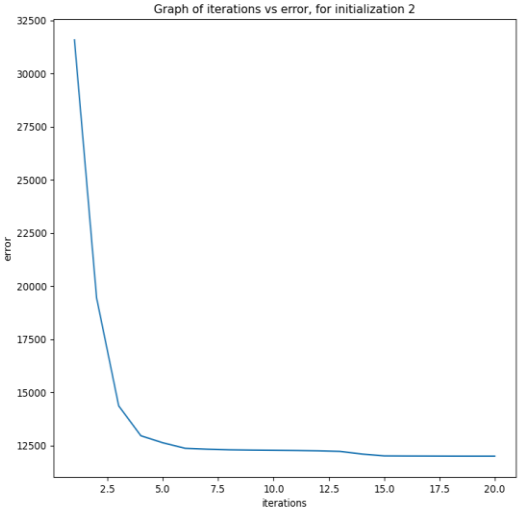
Graph for Clustered dataset and iterations vs error, for initialization 1



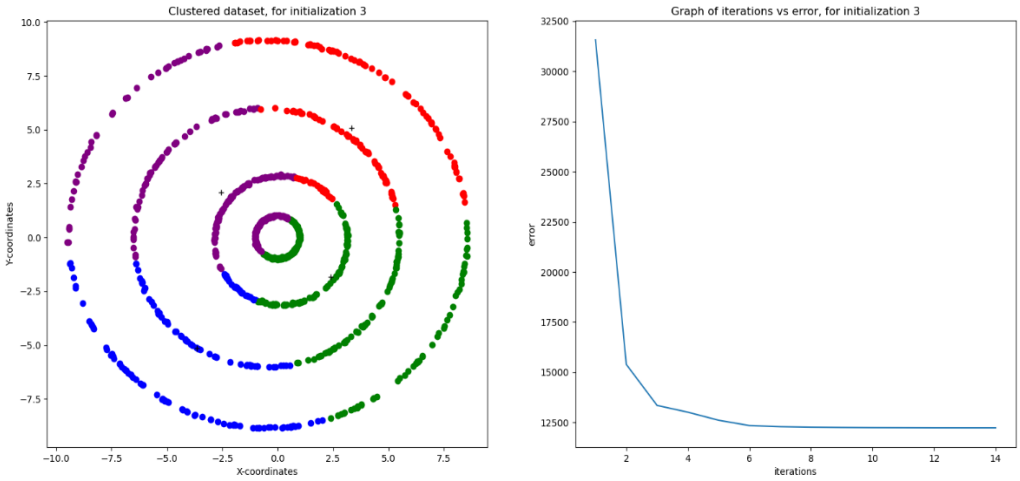


Graph for Clustered dataset and iterations vs error, for initialization 2

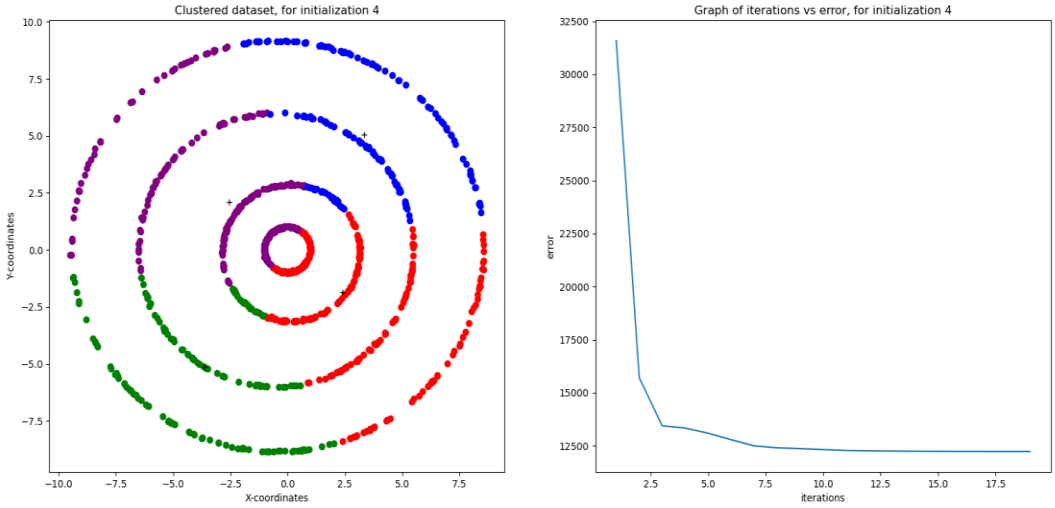




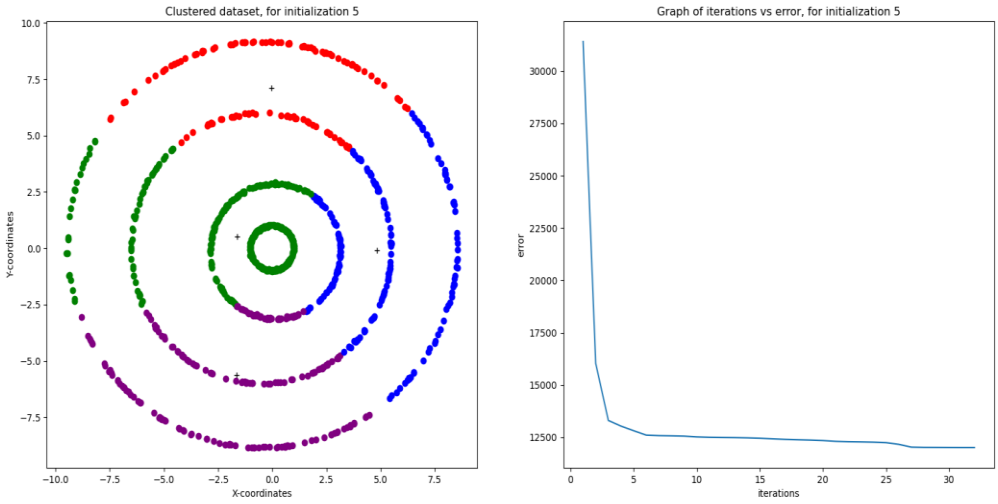
Graph for Clustered dataset and iterations vs error, for initialization 3



Graph for Clustered dataset and iterations vs error, for initialization 4



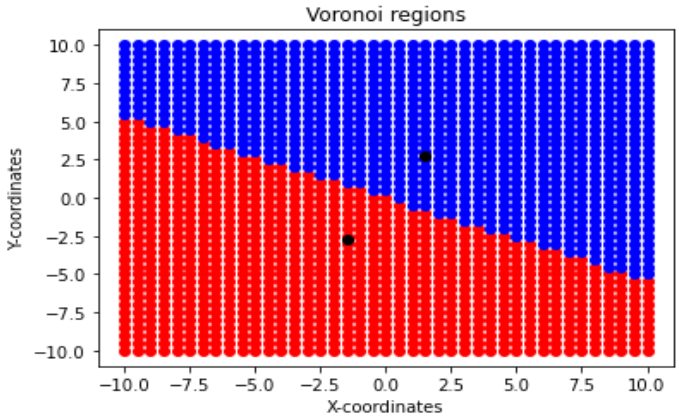
Graph for Clustered dataset and iterations vs error, for initialization

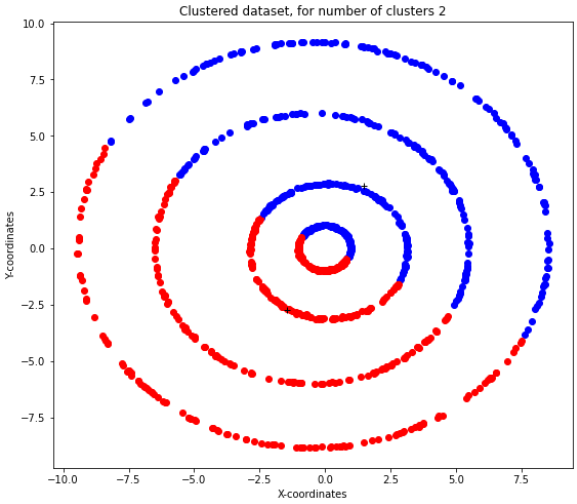


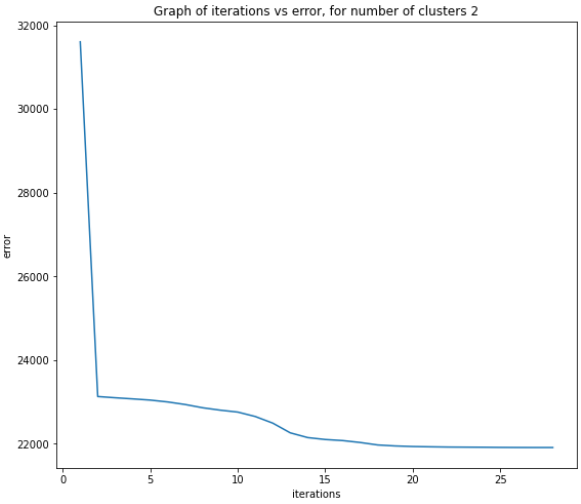
2.) K-means clustering algorithm for k = {2,3,4,5}

Graph for Clustered dataset and iterations vs error, for 2 clusters

Voronoi regions

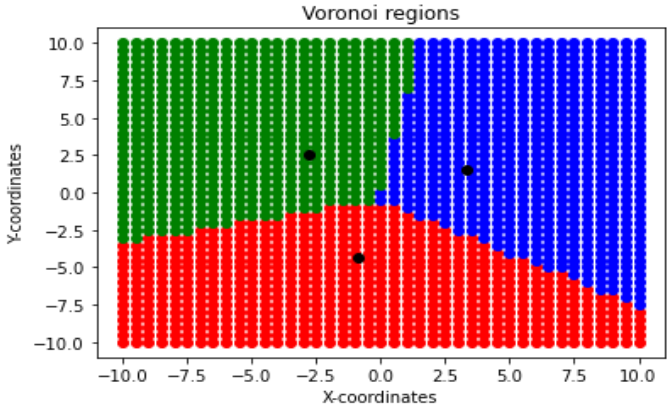


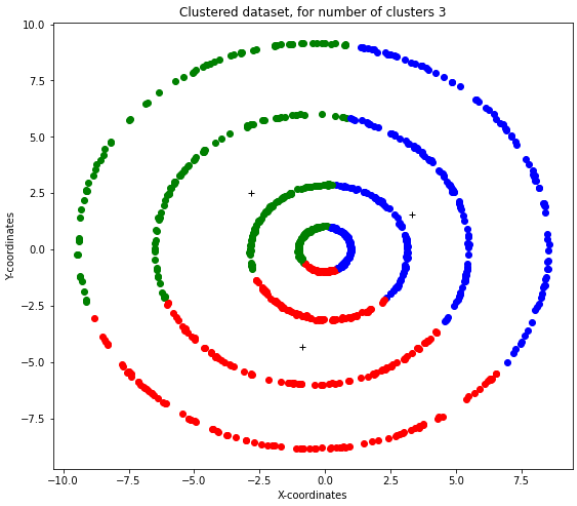


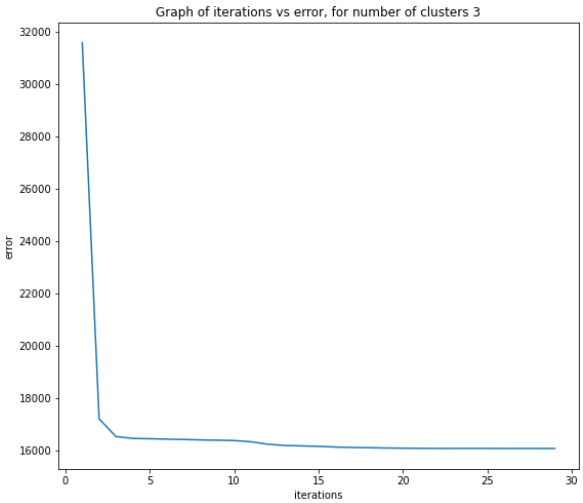


Graph for Clustered dataset and iterations vs error, for 3 clusters

Voronoi regions

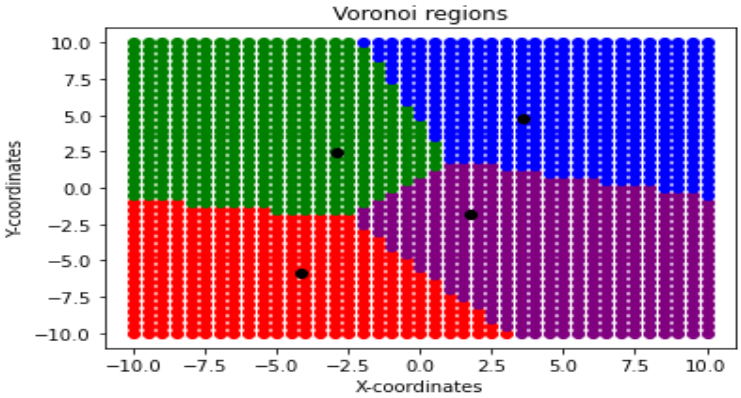


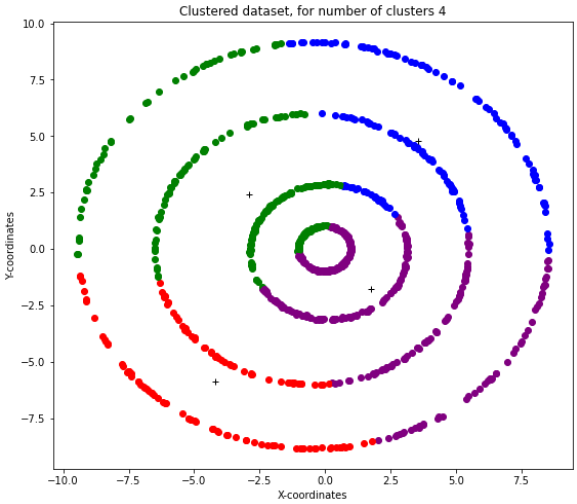


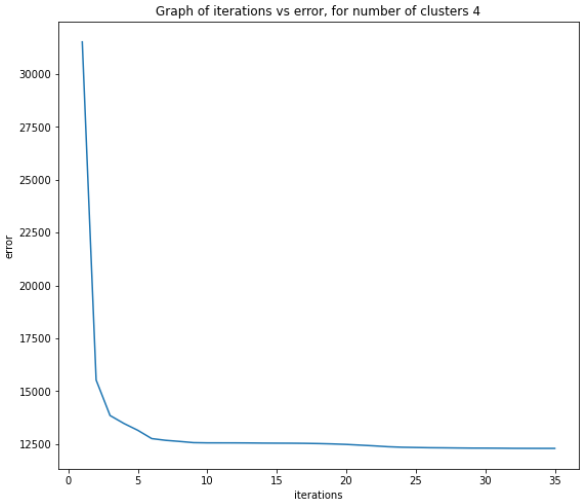


Graph for Clustered dataset and iterations vs error, for 4 clusters

Voronoi regions

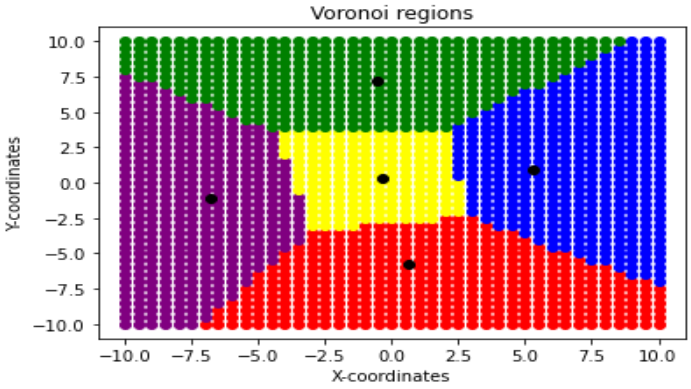


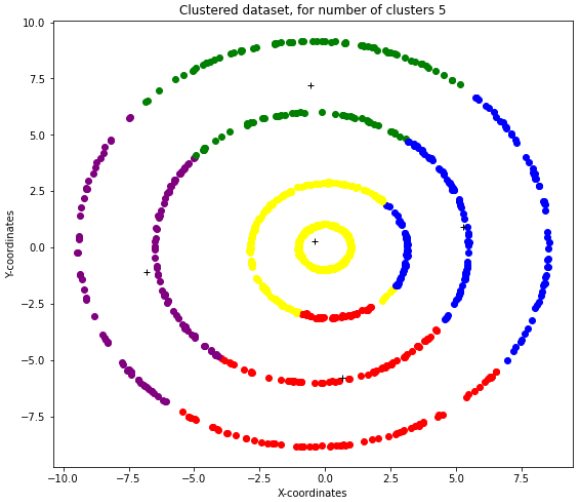


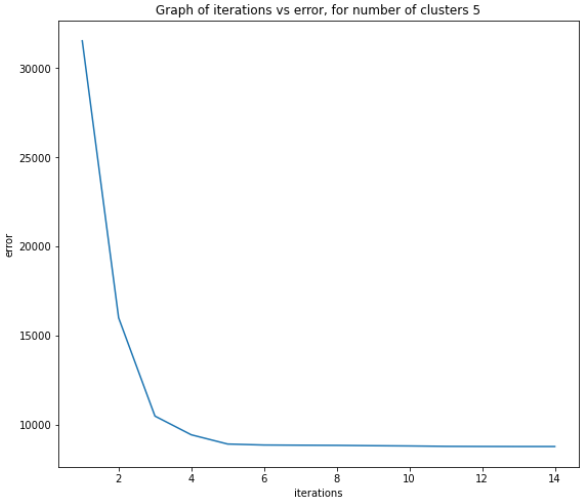


Graph for Clustered dataset and iterations vs error, for 5 clusters

Voronoi regions



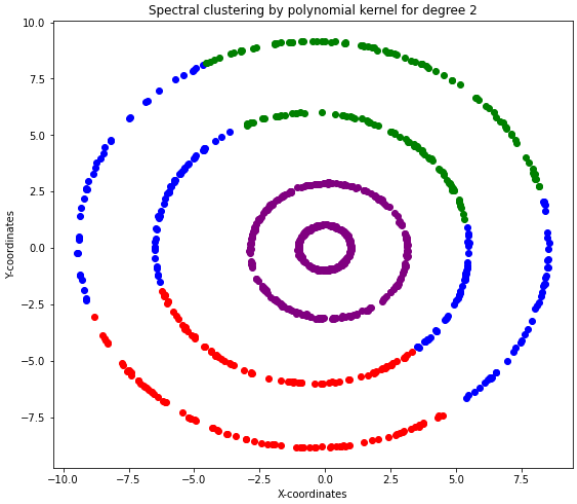


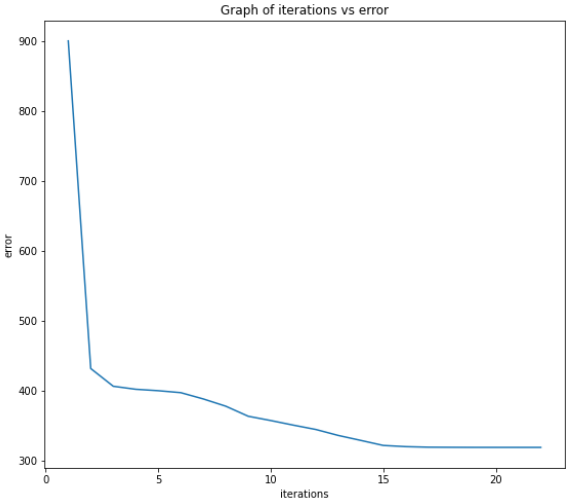


3.) Spectral Clustering algorithm for k = 4

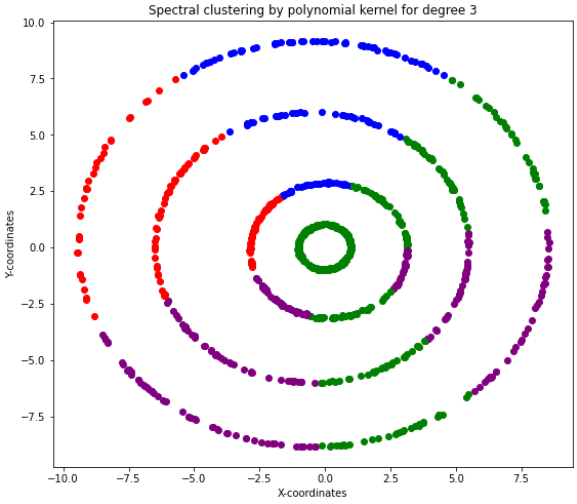
Clustering of a dataset by different kernel functions is shown below images :

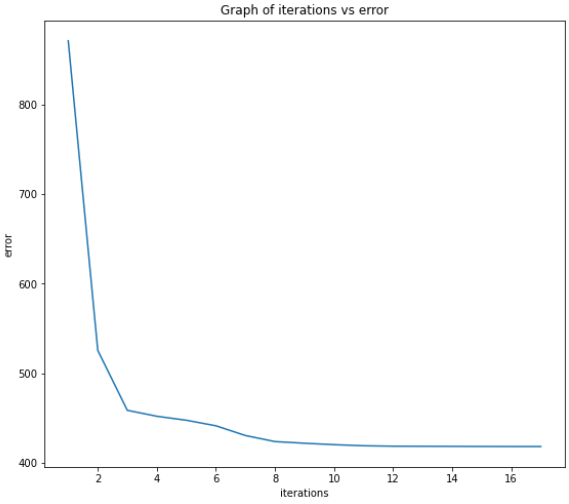
a) Polynomial kernel with degree 2





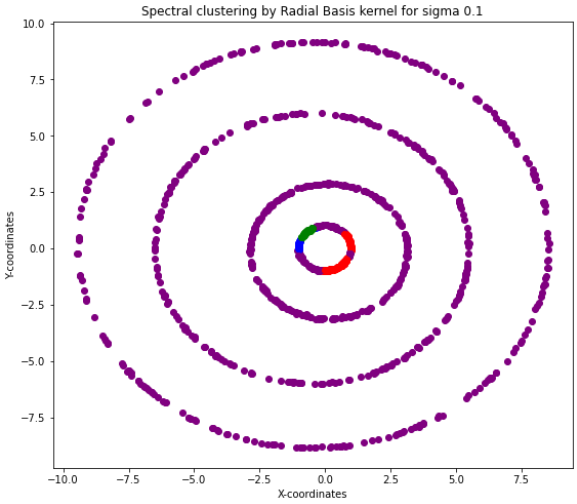
b) Polynomial kernel with degree 3

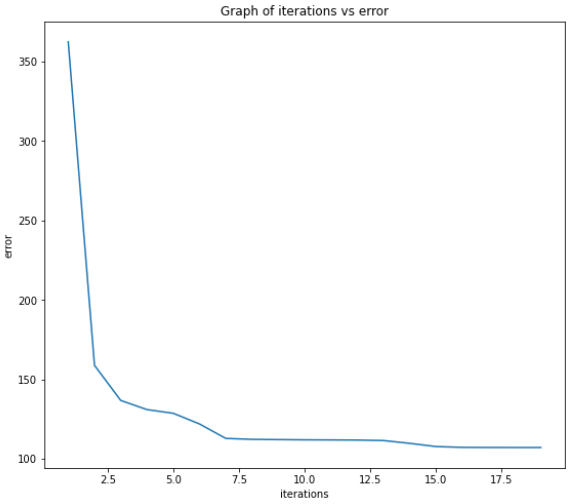




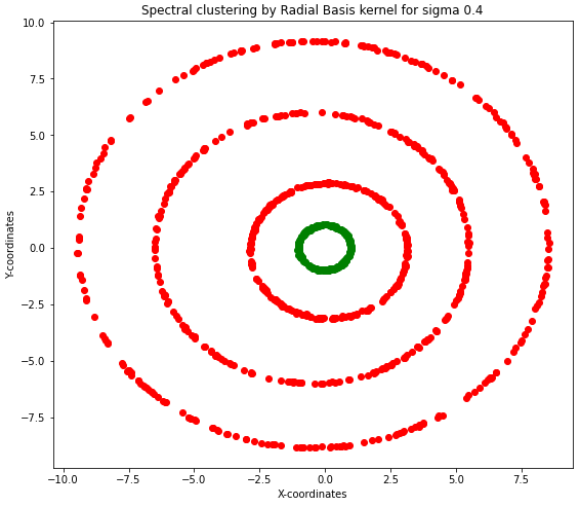
Radial Basis Kernel

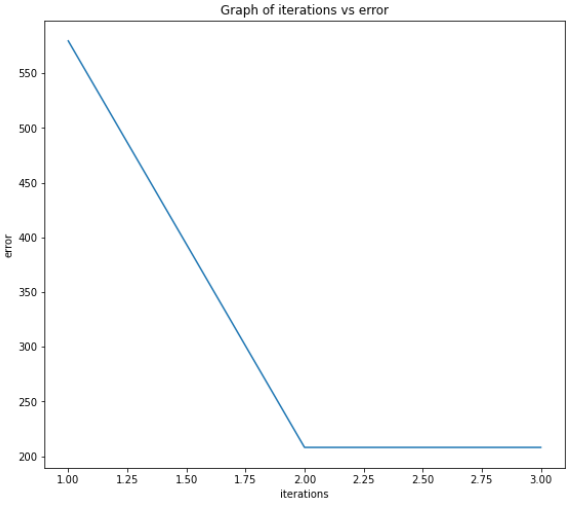
c) radial basis kernel with sigma 0.1



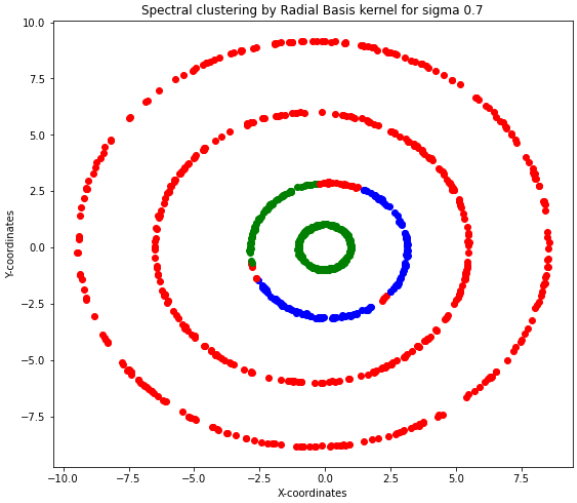


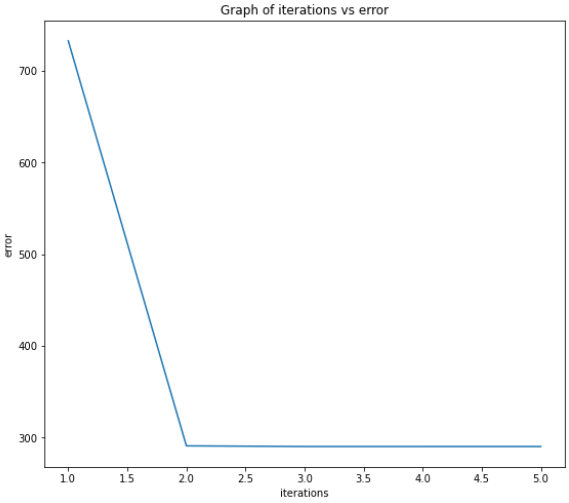
d) radial basis kernel with sigma 0.4



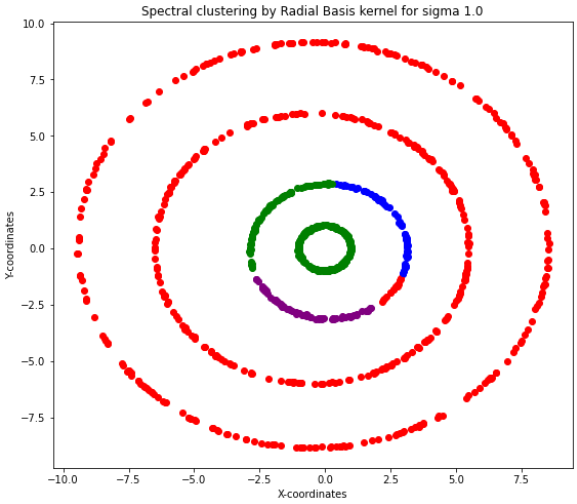


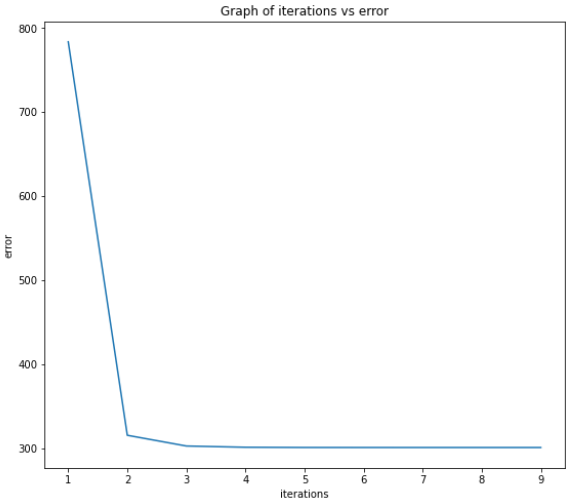
e) radial basis kernel with sigma 0.7





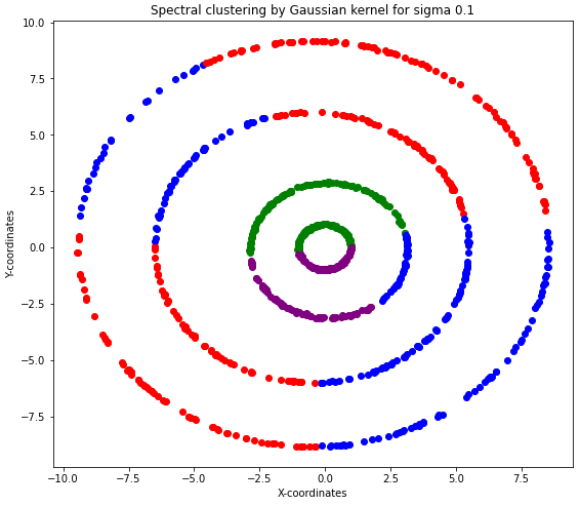
f) radial basis kernel with sigma 1.0

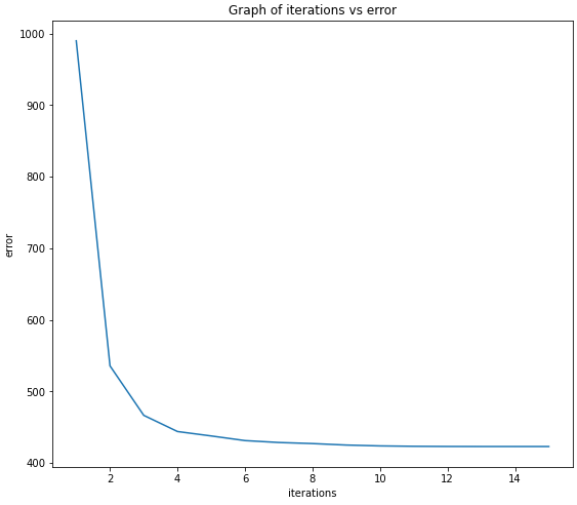




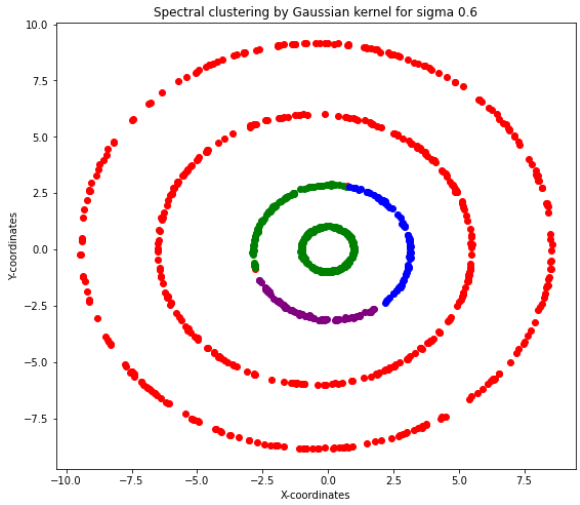
Gaussian Kernel

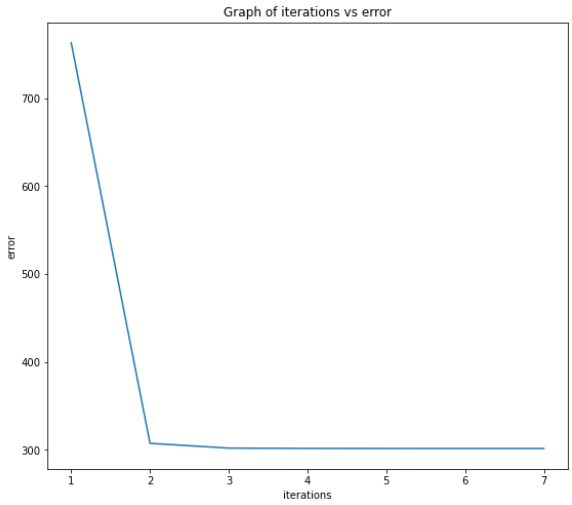
g) Gaussian kernel with sigma 0.1



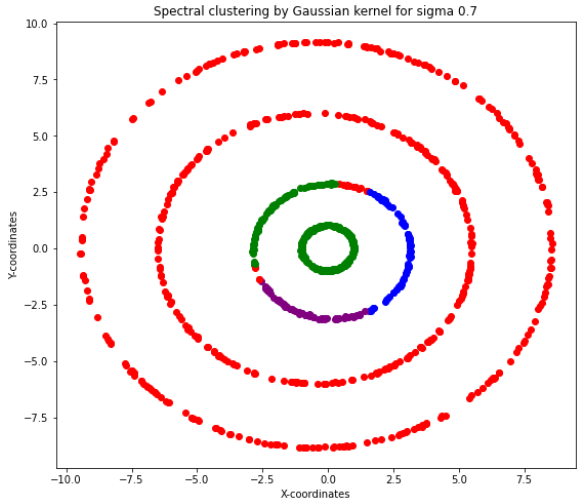


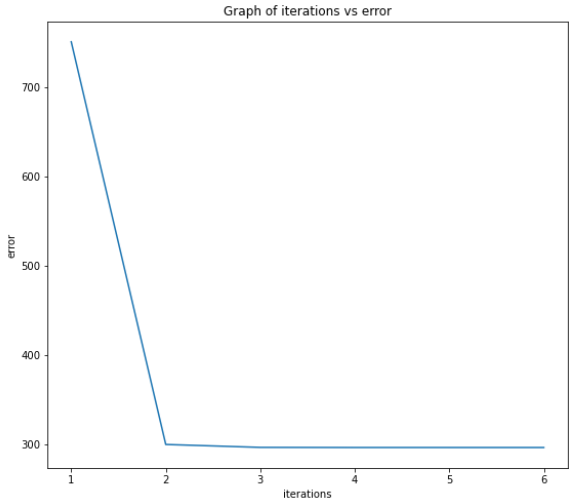
h) Gaussian kernel with sigma 0.6





i) Gaussian kernel with sigma 0.7





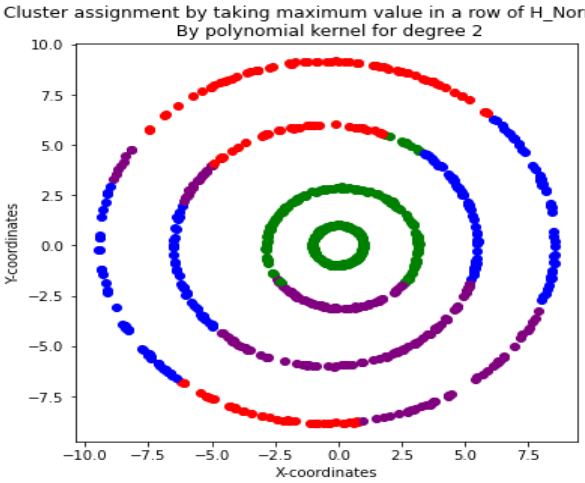
Conclusion : From the above three different kind of kernel functions, we can say that the polynomial kernel with degree 2 gives a better result, since the equation of a circle is (x-a)**2 + (y-b)**2 = r**2. The degree of this equation is 2, so we can say that polynomial function with degree 2 is better than the other functions.

4.) Spectral Clustering by taking the maximum value and assigning to the cluster

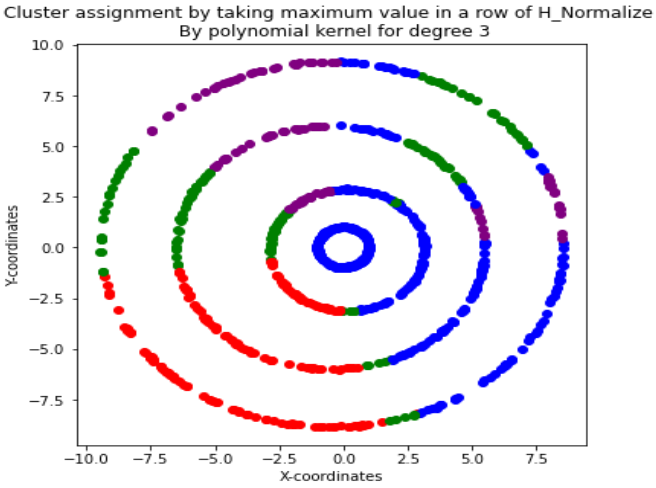
Clustering of a dataset by different kernel functions is shown below images :

Polynomial Kernel

a) Polynomial kernel with degree 2

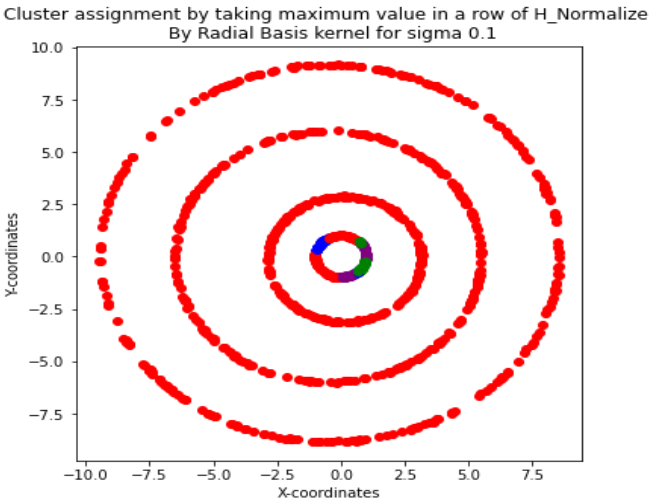


b) Polynomial kernel with degree 3

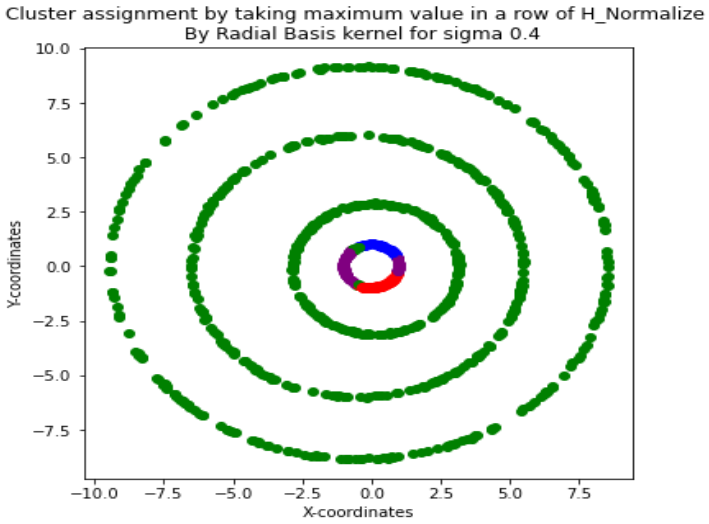


Radial Basis kernel

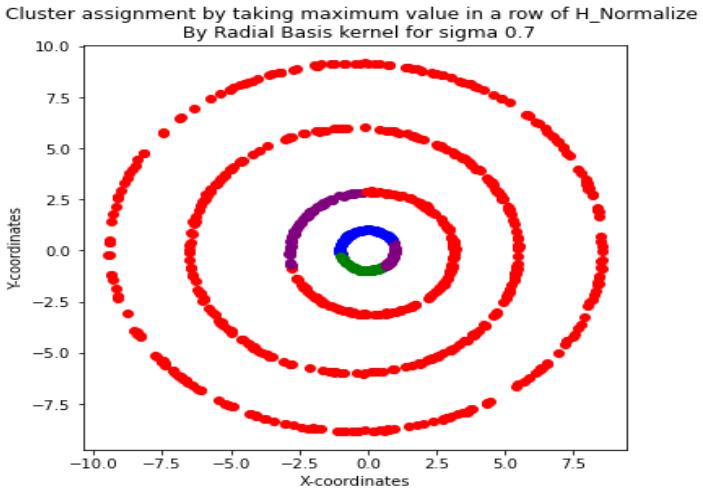
c) Radial basis kernel with sigma 0.1



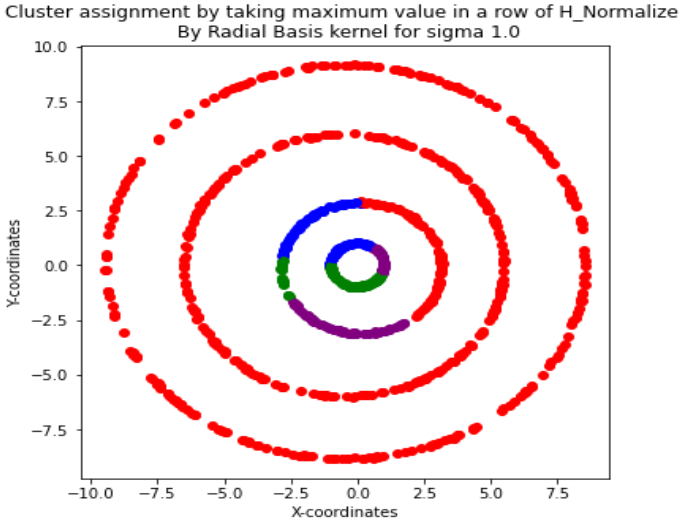
d) Radial basis kernel with sigma 0.4



e) Radial basis kernel with sigma 0.7

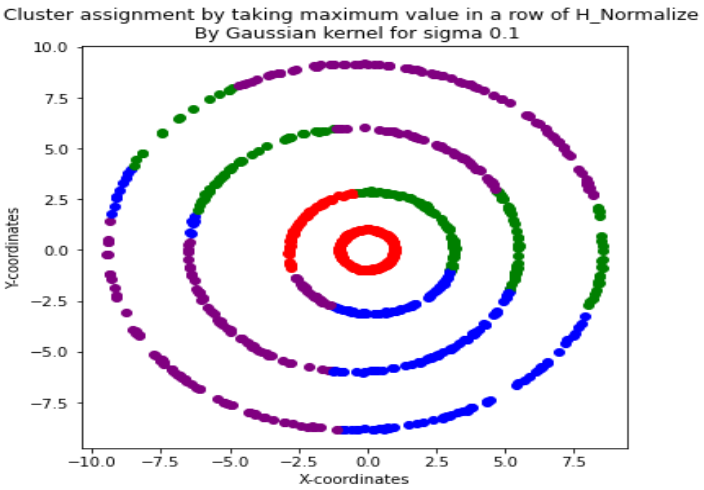


f) Radial basis kernel with sigma 1.0

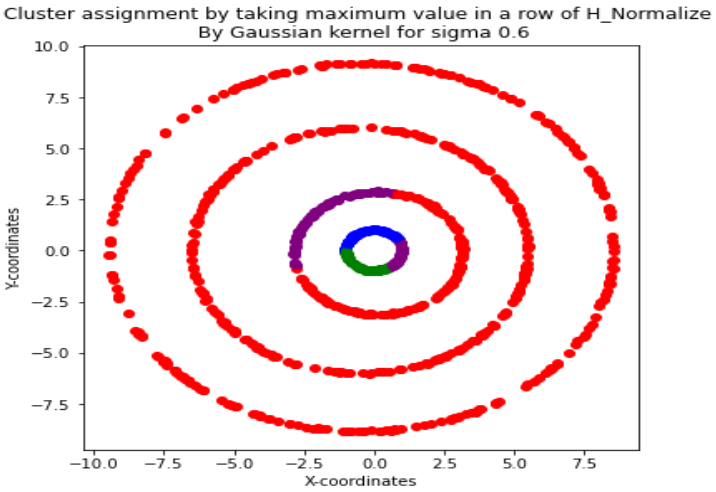


Gaussian kernel

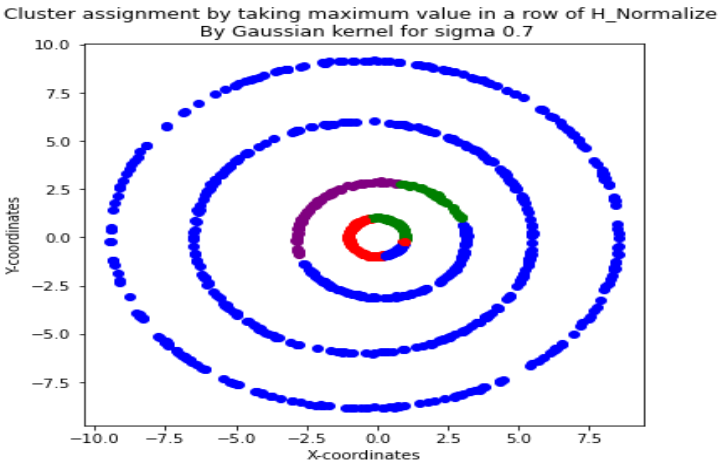
g) Gaussian kernel with sigma 0.1



h) Gaussian kernel with sigma 0.6



i) Gaussian kernel with sigma 0.7



The mapping given in the question is not perform well for the given dataset, because this is a kind of deterministic method of implementing clustering. We are not even checking whether the datapoint has a similar kind of property assigned to its cluster. The convergence of error is not even considered for assigning the clusters.