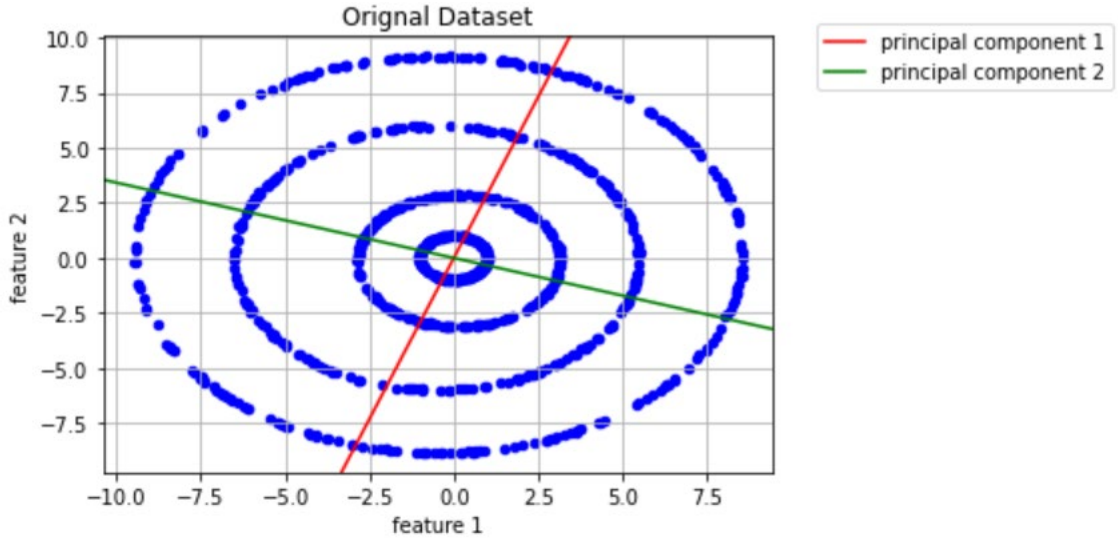
(1) You are given a data-set with 1000 data points each in $\mathbb{R}^2$.

i. Write a piece of code to run the PCA algorithm on this data-set. How much of the variance in the data-set is explained by each of the principal components?

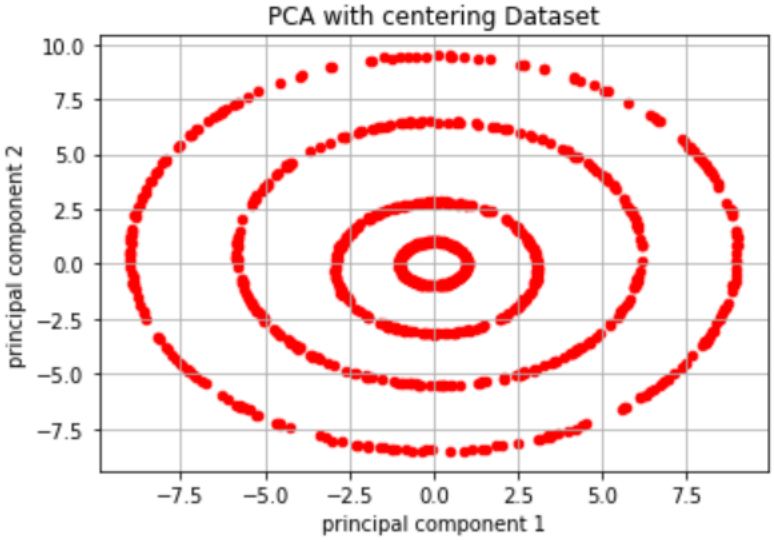
Answer

The variance explained by principal component 1 is `54.18%`

The variance explained by principal component 1 is `45.82%`



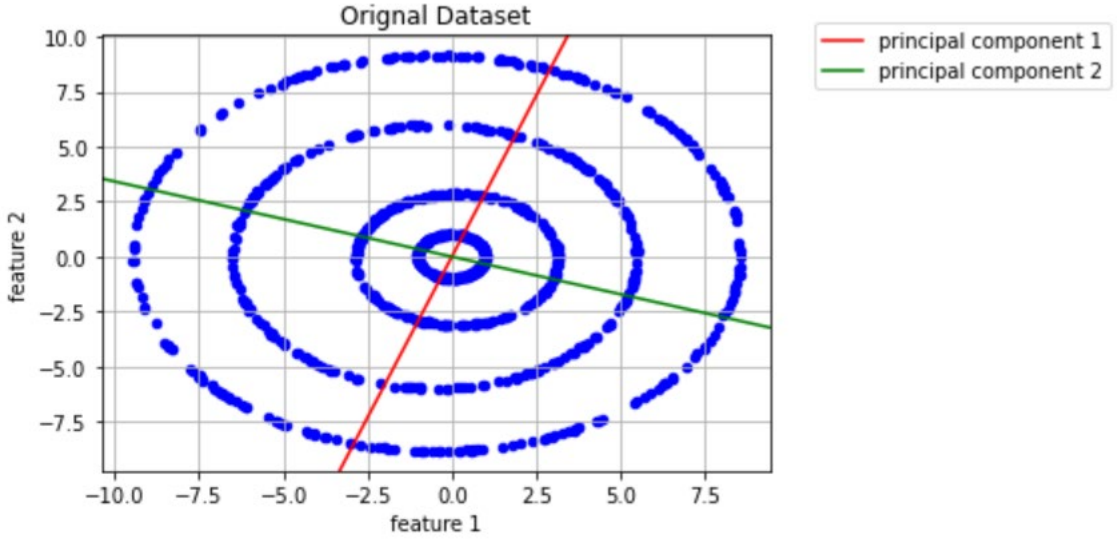

```
variance explained by 1 eigen vector is  54.17802452885221 %
variance explained by 2 eigen vector is  45.82197547114778 %
```



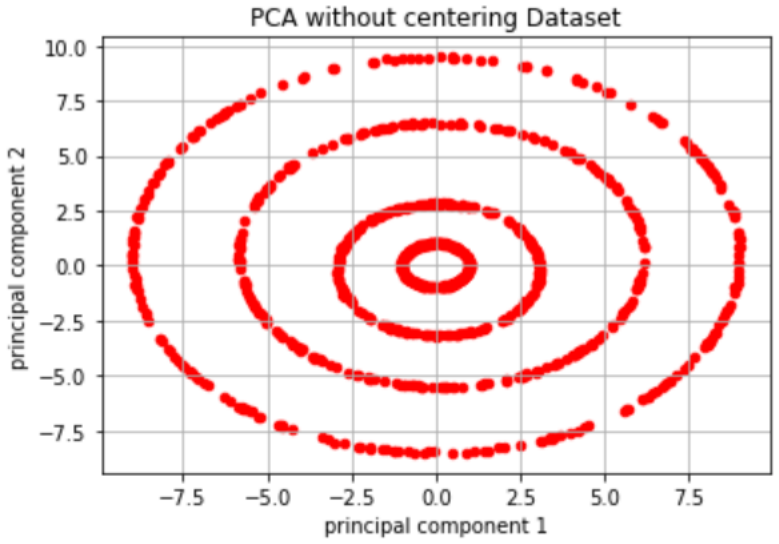
ii. Study the effect of running PCA without centering the data-set. What are your observations? Does Centering help?

Answer :

As the mean is very small [4.07e-07 2.22e-07] so subtracting the mean from the datasets do not give any significance difference from the earlier result



variance explained by 1 eigen vector is 54.17802452885223 %
variance explained by 2 eigen vector is 45.821975471147766 %



iii. Write a piece of code to implement the Kernel PCA algorithm on this dataset. Use the following kernels :

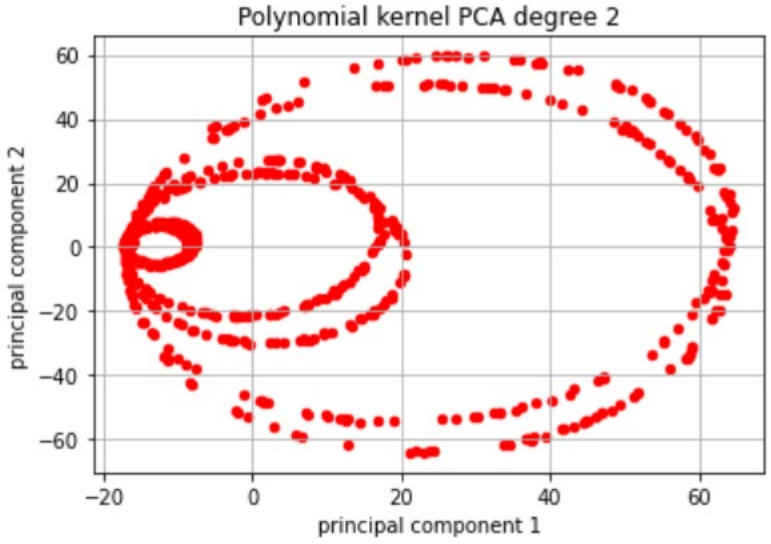
A. $\kappa(x, y) = (1 + x^T y)^d$ for $d = \{2, 3\}$

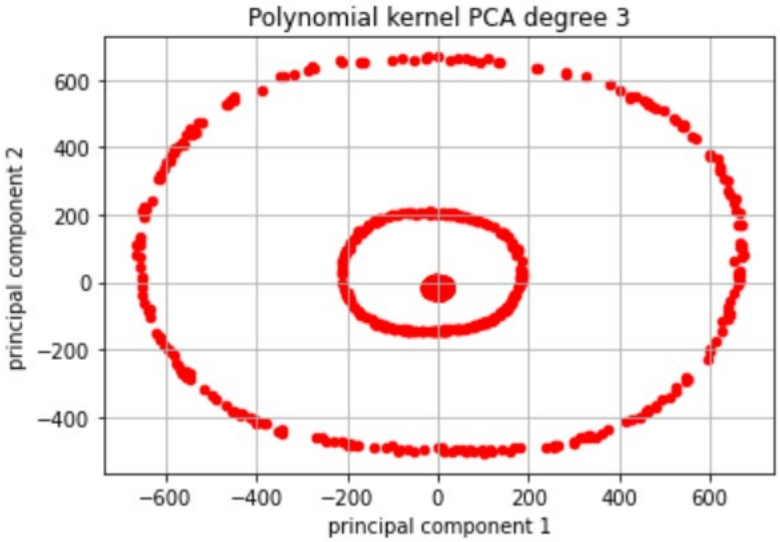
B. $\kappa(x, y) = \exp \frac{-(x-y)^T(x-y)}{2\sigma^2}$ for $\sigma = \{0.1, 0.2, \ldots, 1\}$

Plot the projection of each point in the dataset onto the top-2 components for each kernel. Use one plot for each kernel and in the case of (B), use a different plot for each value of σ.

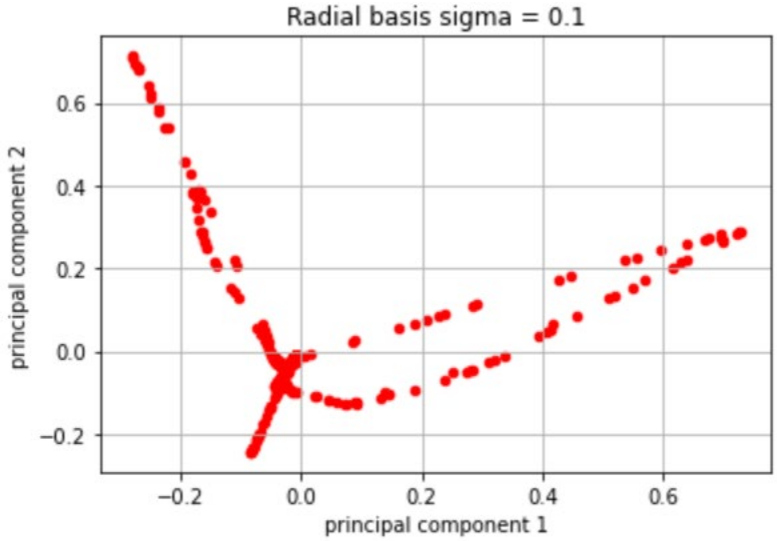
Answer

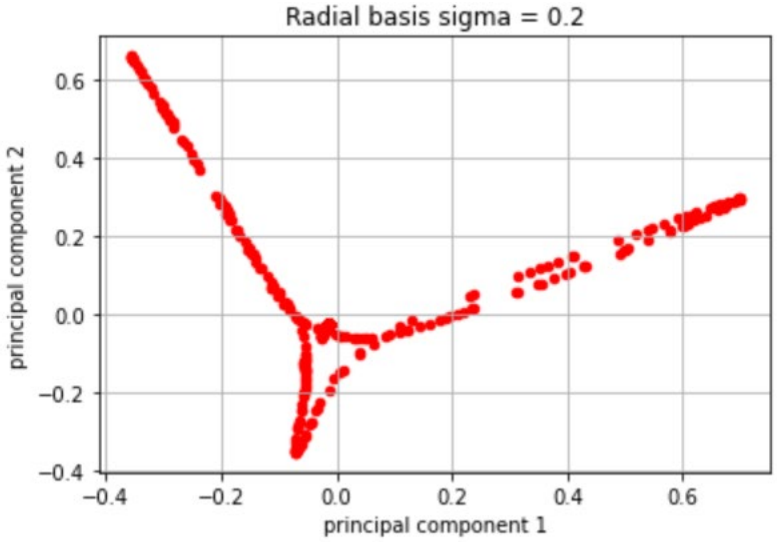
A) Polynomial kernel

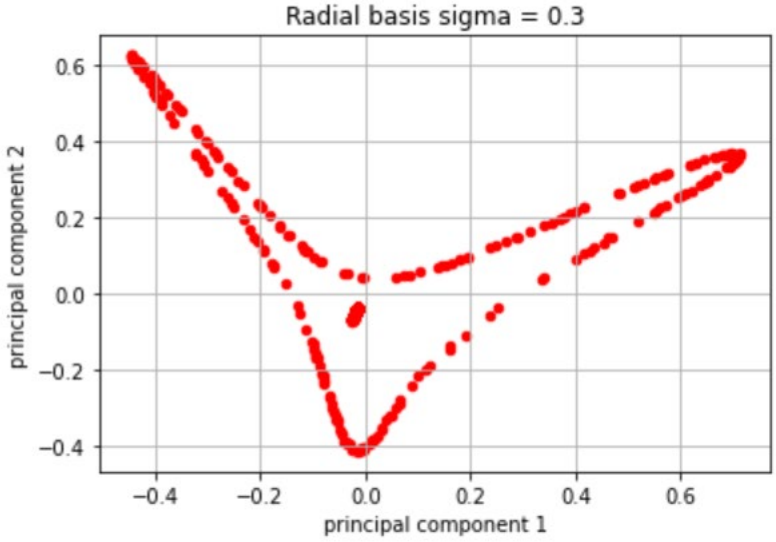


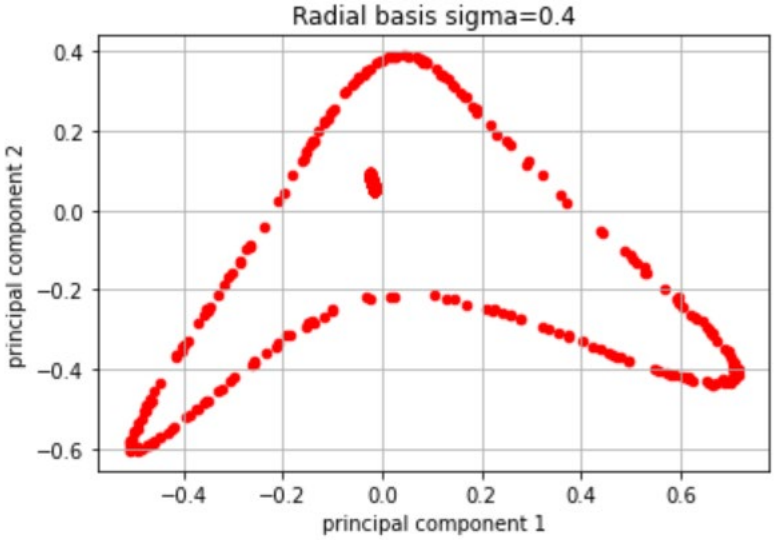


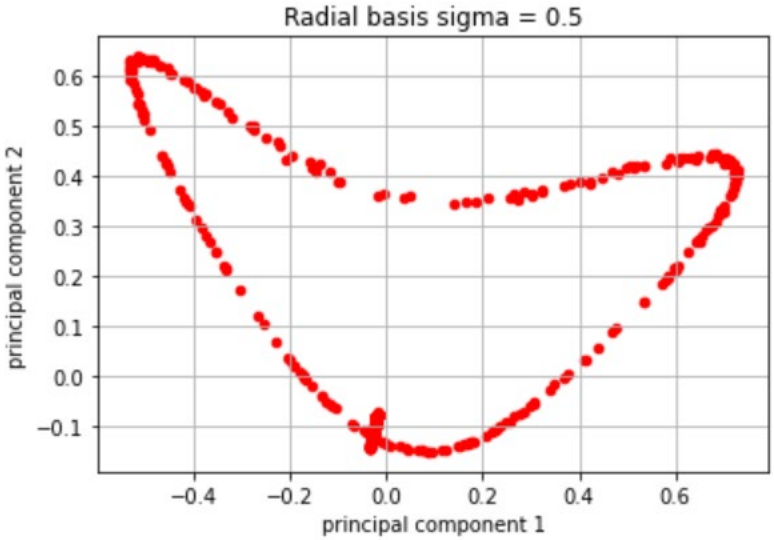
B) Radial Basis kernel

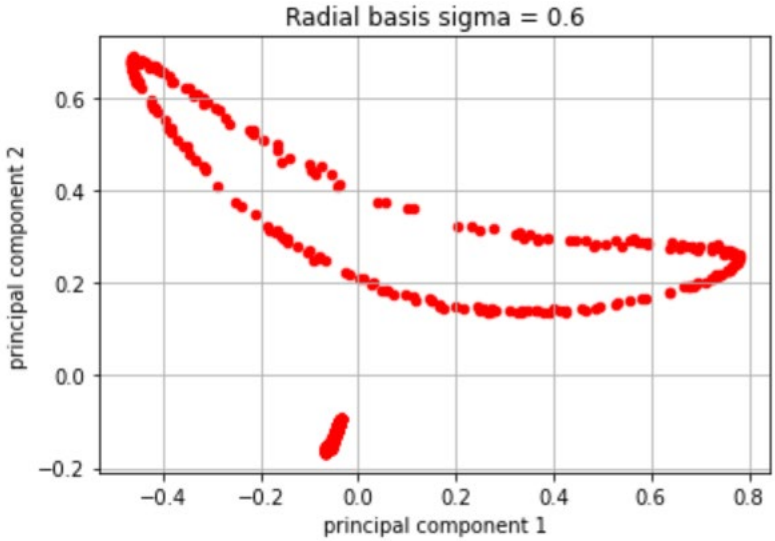


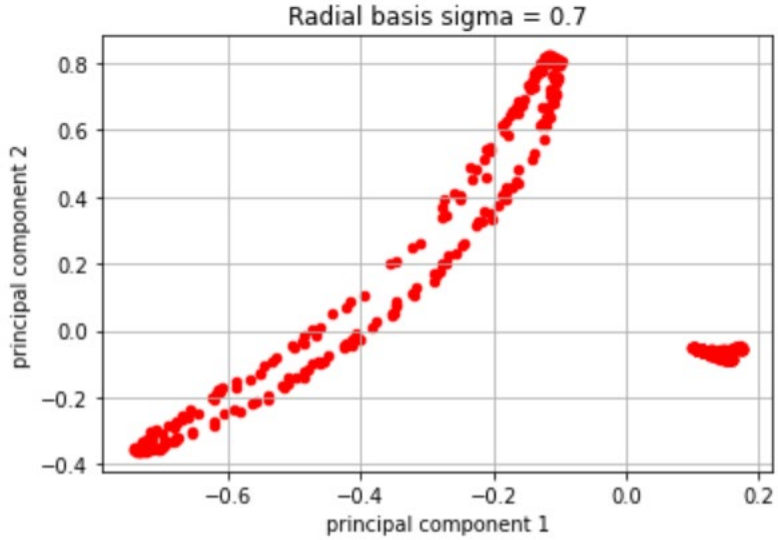


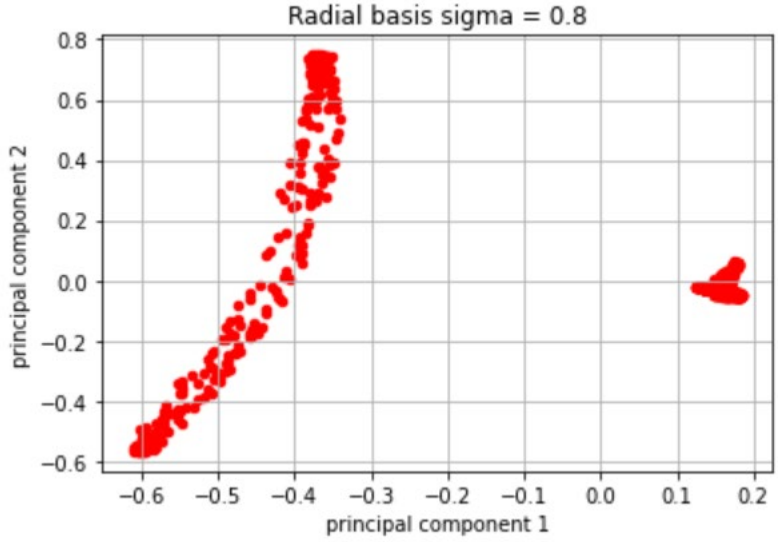


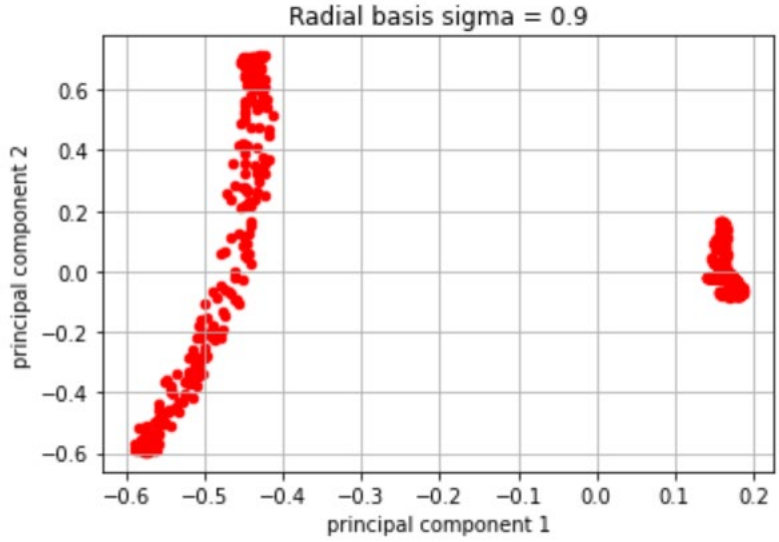

Radial basis sigma=0.4
principal component 2
principal component 1
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.4
−0.2
0.0
0.2
0.4
0.6

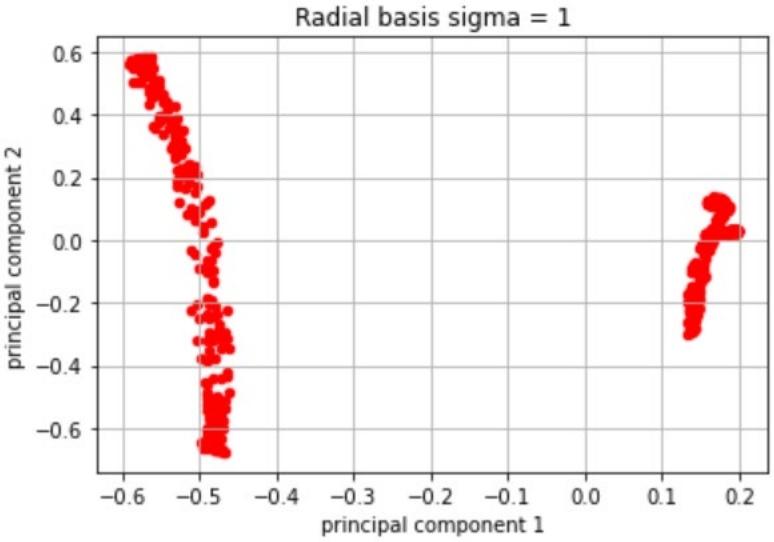

Radial basis sigma = 0.5
principal component 2
principal component 1
0.6
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
−0.4
−0.2
0.0
0.2
0.4
0.6


Radial basis sigma = 0.6
principal component 2
principal component 1
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.2
0.0
0.2
0.4
0.6
0.8


Radial basis sigma = 0.7
principal component 2
principal component 1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.4
−0.2
0.0
0.2


Radial basis sigma = 0.8
principal component 2
principal component 1
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.6
−0.5
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2


Radial basis sigma = 0.9
principal component 2
principal component 1
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.6
−0.5
−0.4
−0.3
−0.2
−0.1
0.0
0.1
0.2



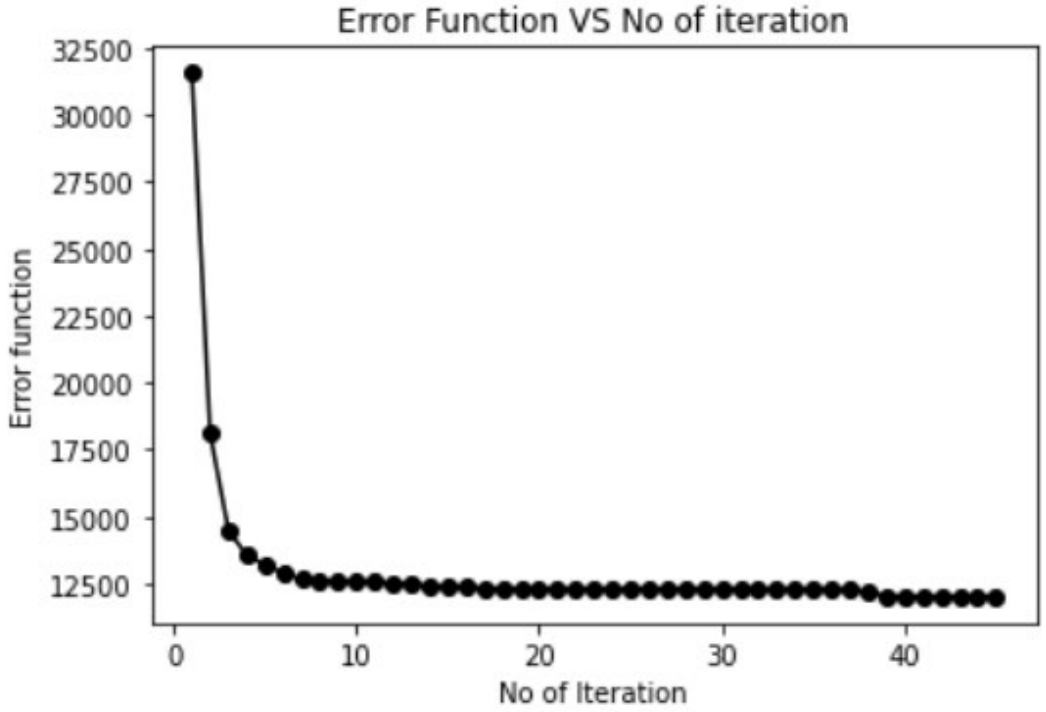
iv. Which Kernel do you think is best suited for this dataset and why?

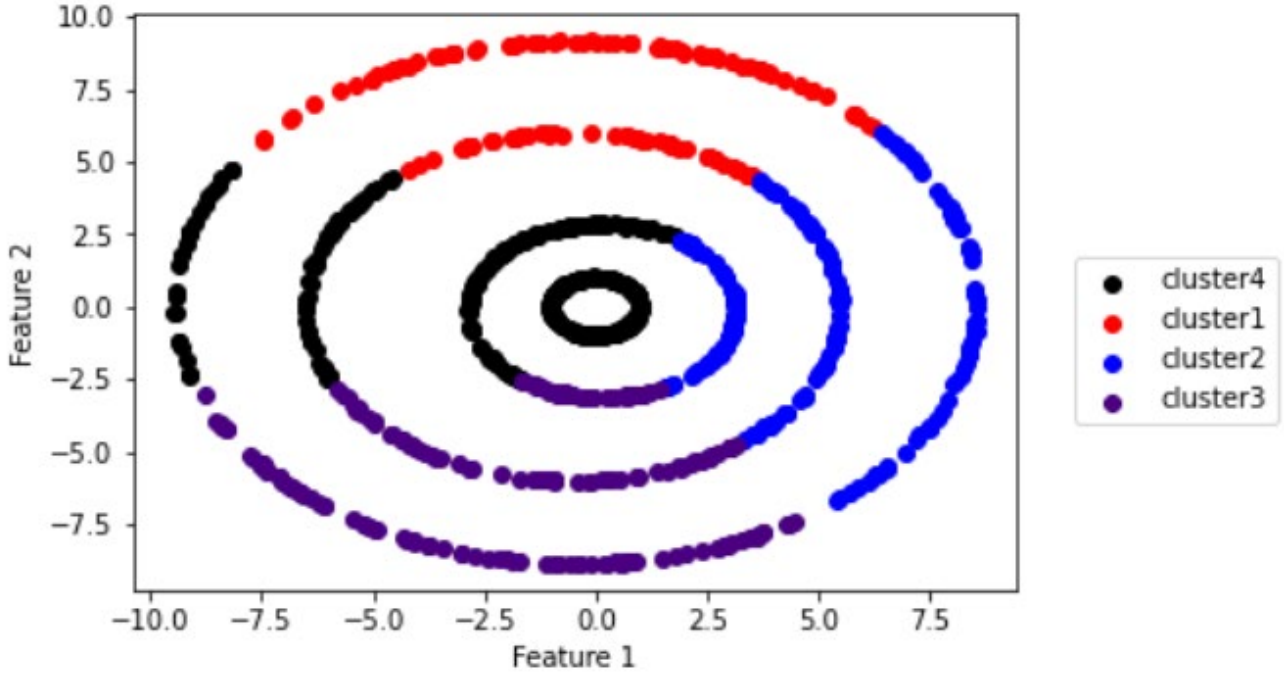
Answer :

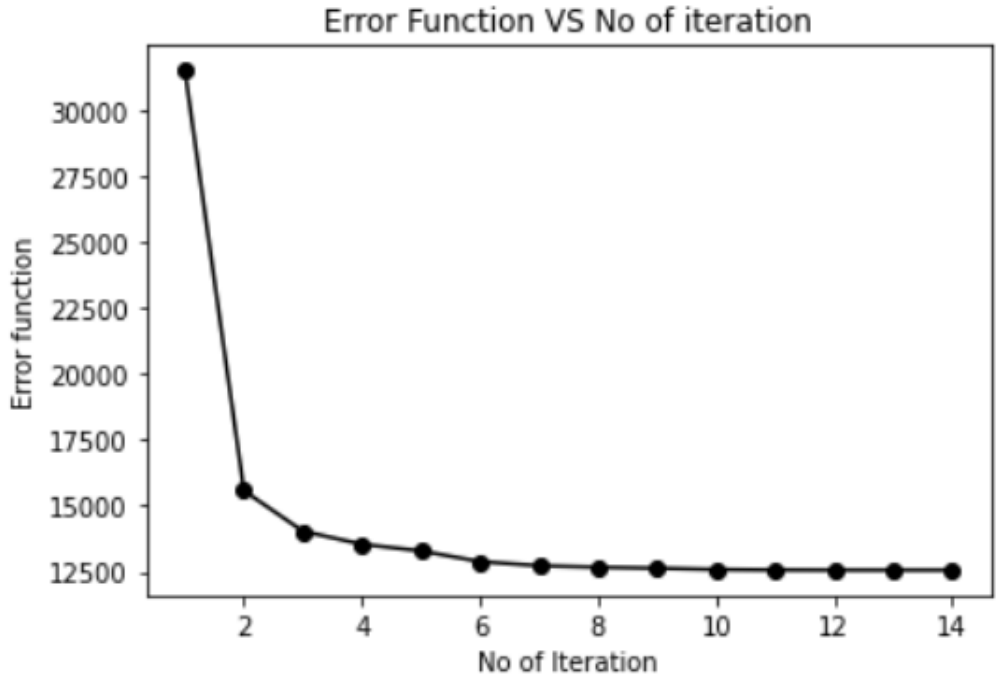
The Polynomial Kernel of degree three is best as the two components capture 73.41% of the data , degree two kernel first 2 principal component capture 68.53% data and in radial kernel basis the kernel with sigma=1 captures 15.64%

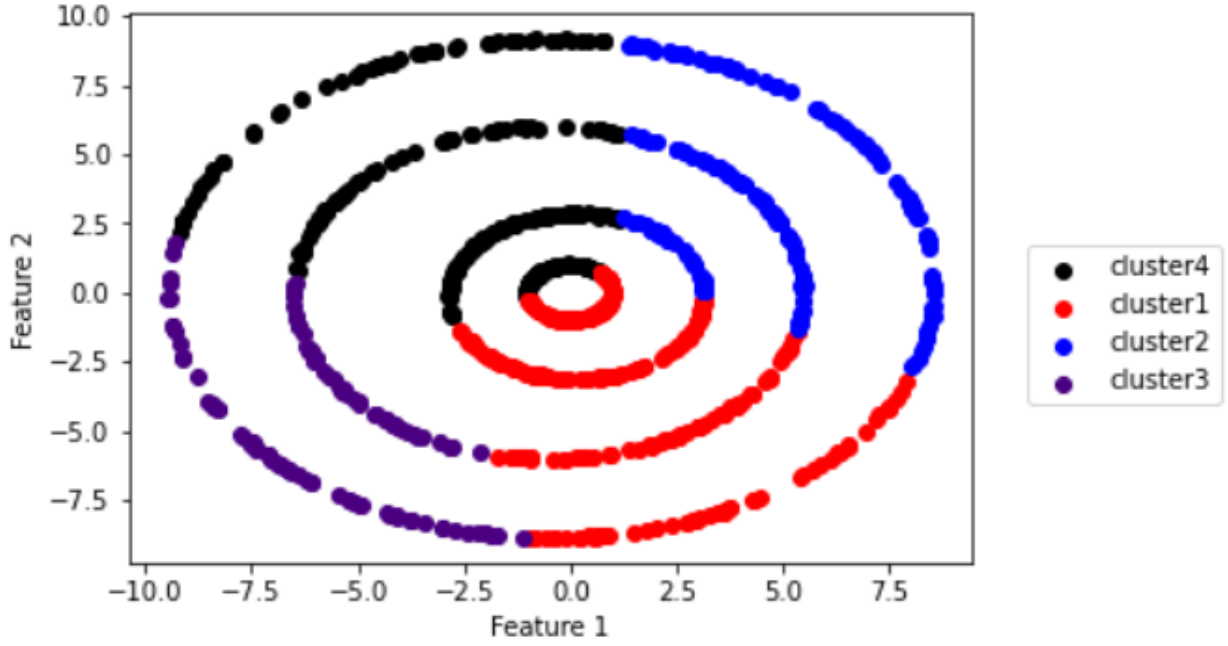
(2) You are given a data-set with 1000 data points each in $\mathbb{R}^2$.

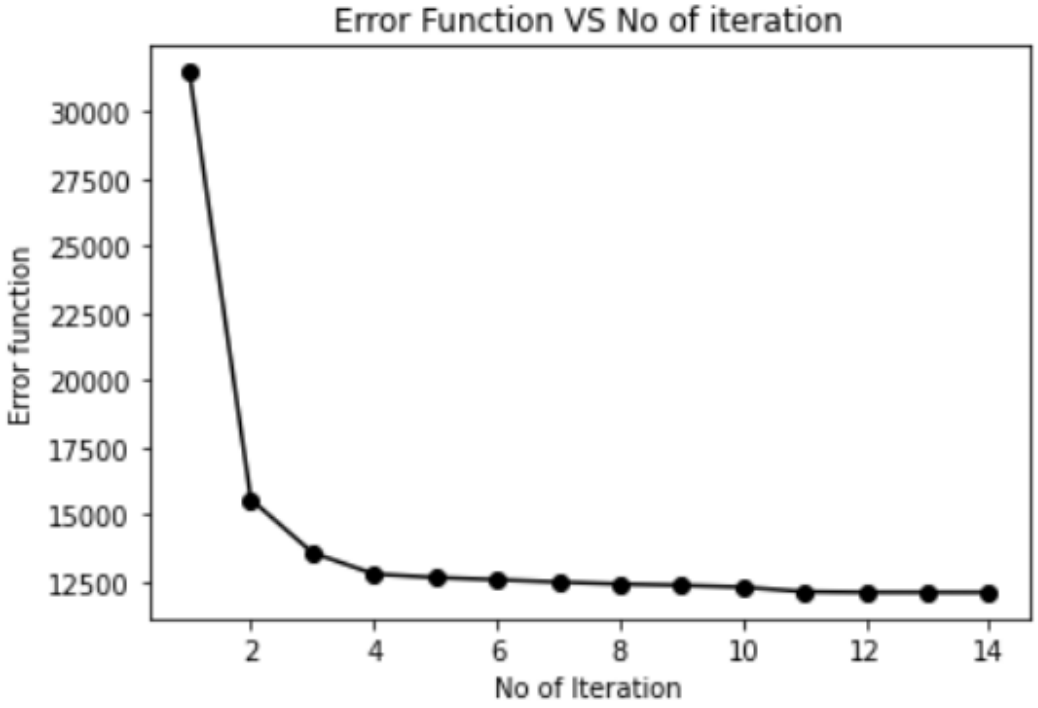
i. Write a piece of code to run the algorithm studied in class for the K-means problem with $k = 4$. Try 5 different random initialization and plot the error function w.r.t iterations in each case. In each case, plot the clusters obtained in different colors.

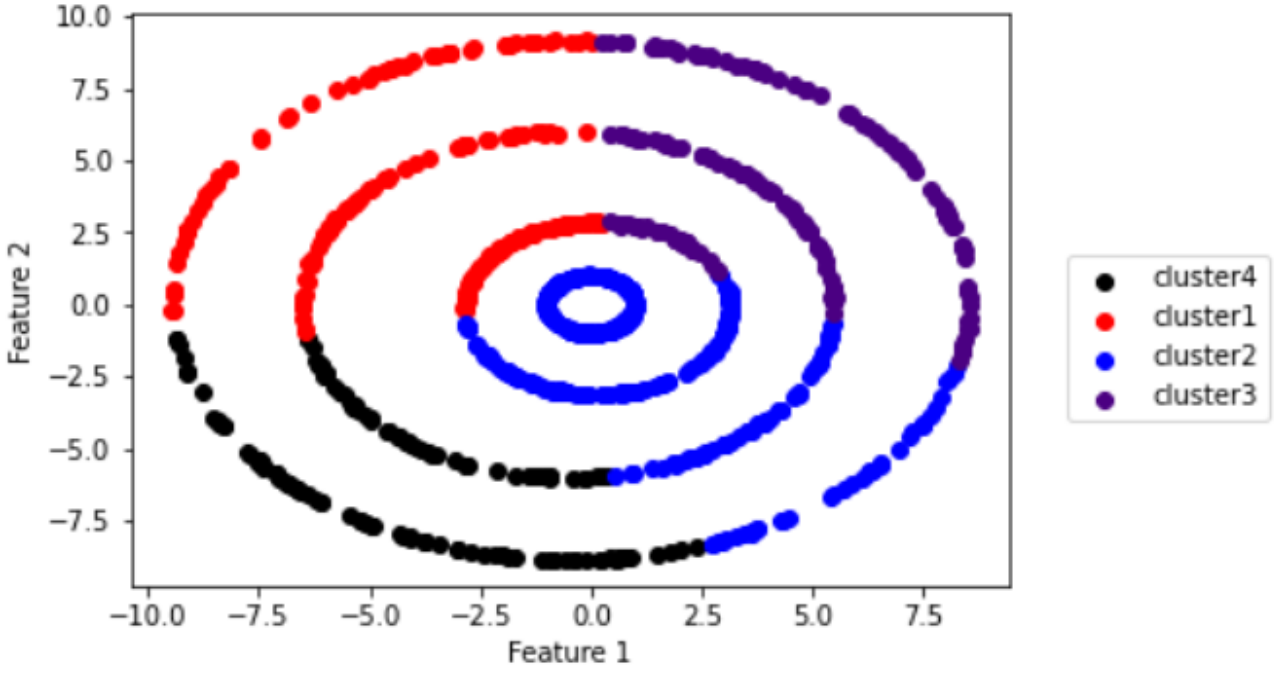


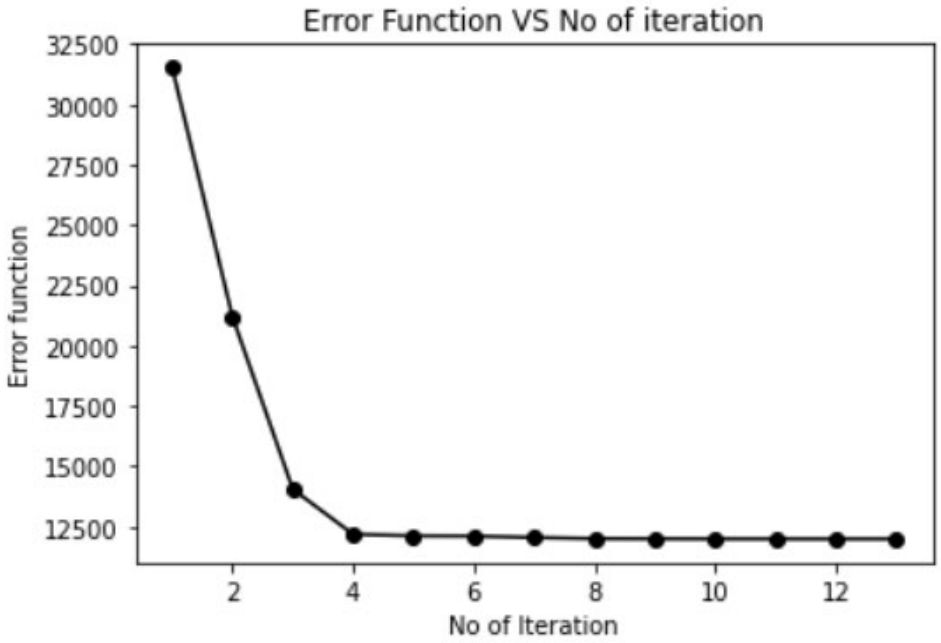


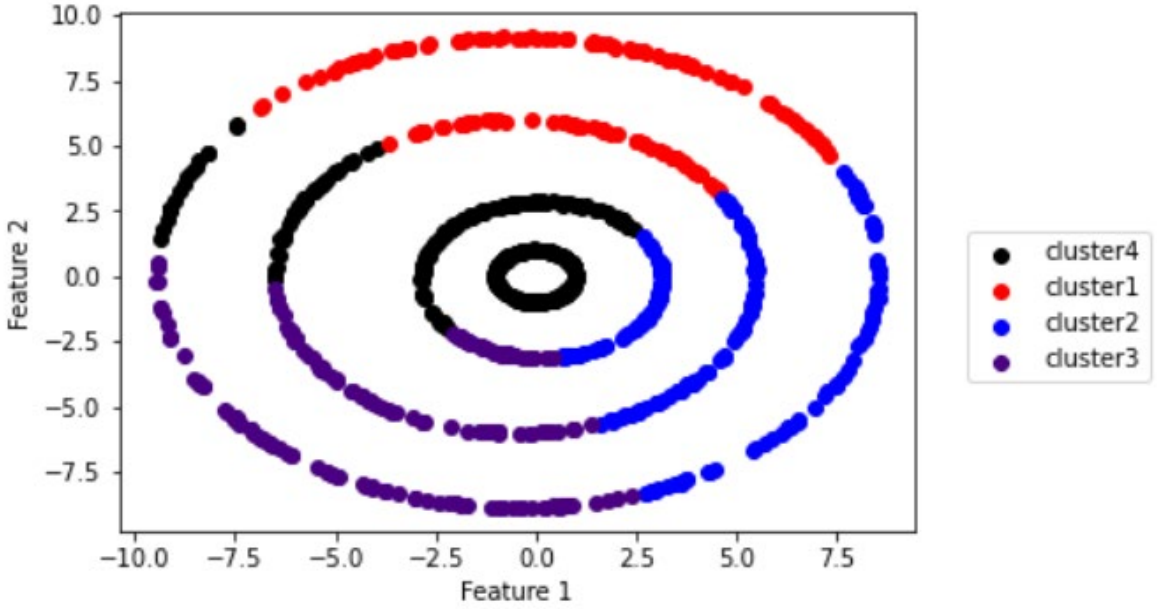

Error Function VS No of iteration
Error function
No of Iteration
30000
27500
25000
22500
20000
17500
15000
12500
2
4
6
8
10
12
14

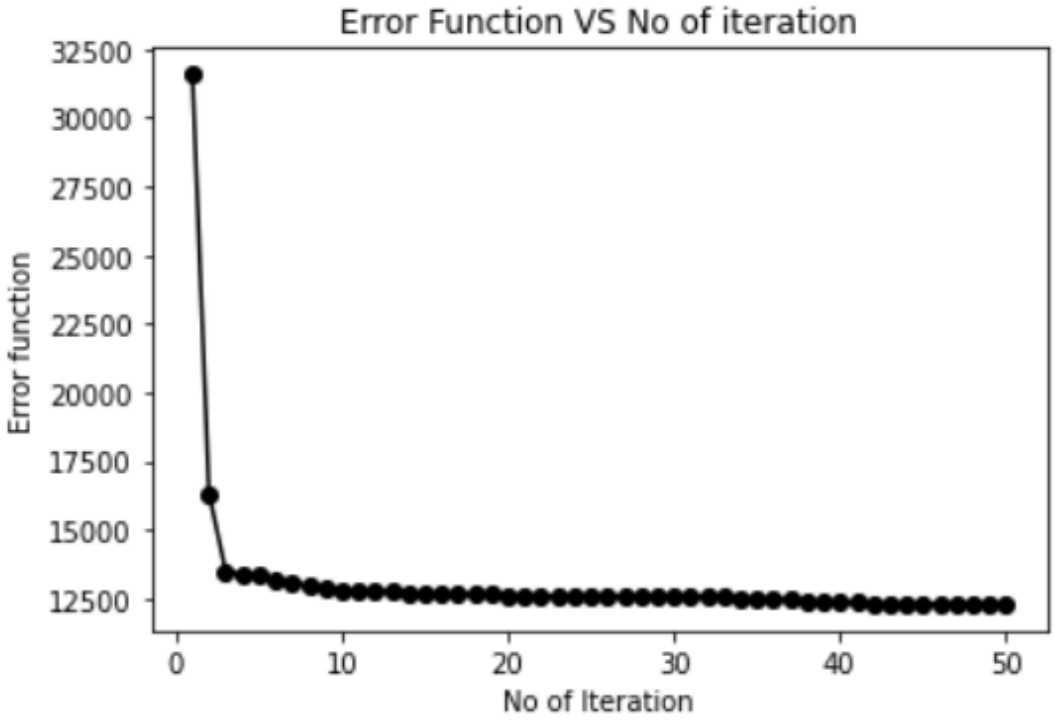

Feature 2
Feature 1
cluster4
cluster1
cluster2
cluster3
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5

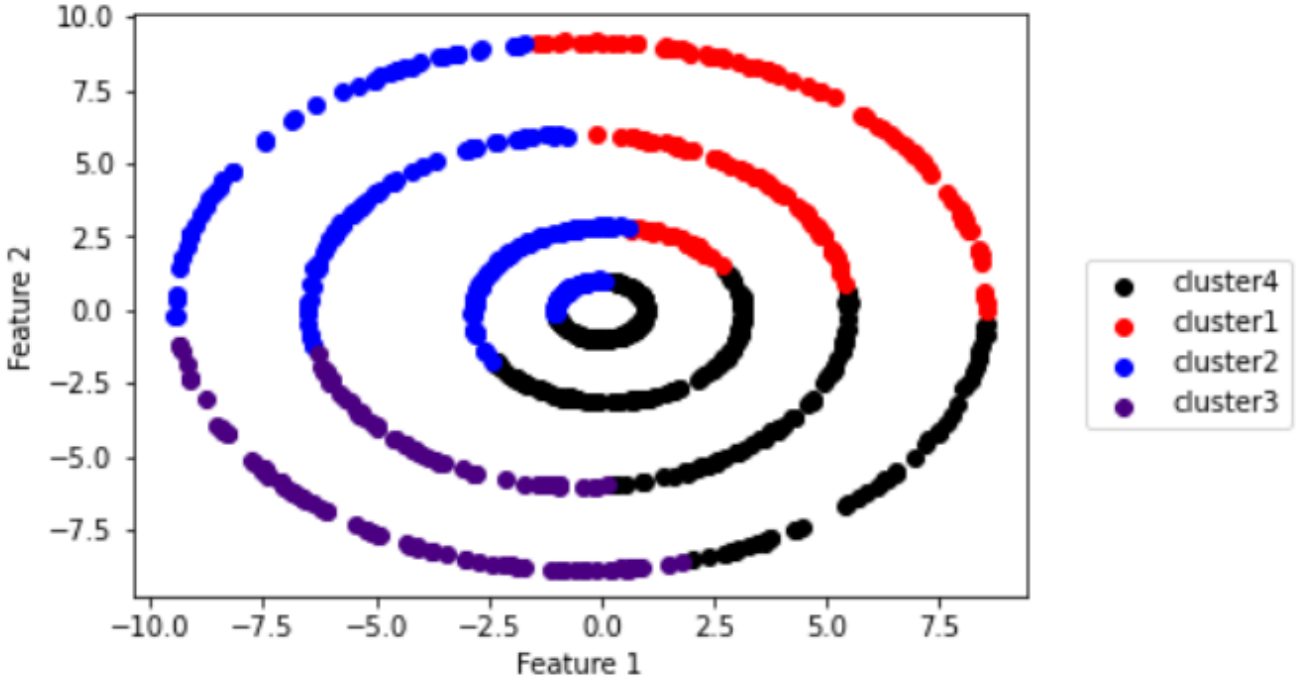

Error Function VS No of iteration
Error function
No of Iteration
30000
27500
25000
22500
20000
17500
15000
12500
2
4
6
8
10
12
14


Feature 2
Feature 1
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
cluster4
cluster1
cluster2
cluster3


Error Function VS No of iteration
Error function
No of Iteration
32500
30000
27500
25000
22500
20000
17500
15000
12500
2
4
6
8
10
12

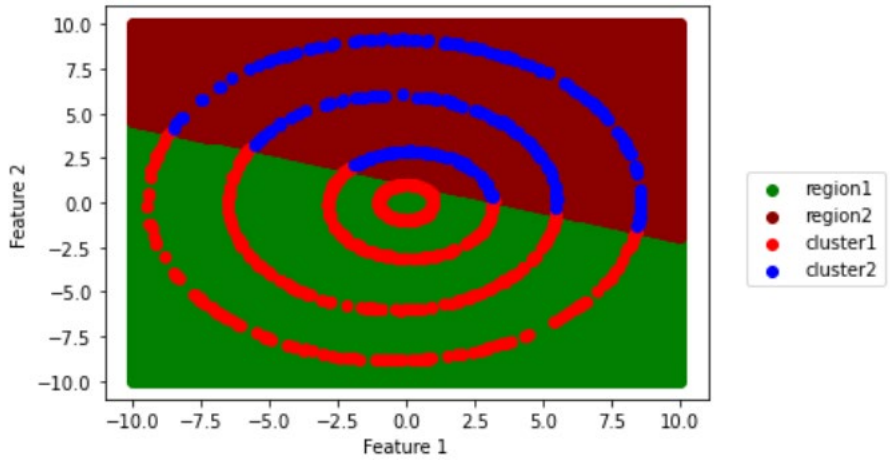

Feature 2
Feature 1
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
cluster4
cluster1
cluster2
cluster3


Error Function VS No of iteration
Error function
No of Iteration
32500
30000
27500
25000
22500
20000
17500
15000
12500
0
10
20
30
40
50

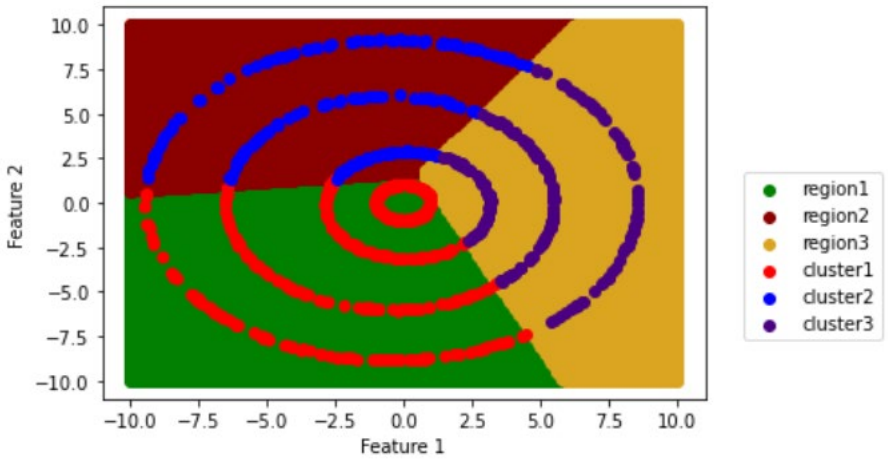

Feature 2
Feature 1
cluster4
cluster1
cluster2
cluster3
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5

ii. Fix a random initialization. For $K = \{2, 3, 4, 5\}$, obtain cluster centers according to K-means algorithm using the fixed initialization. For each value of K, plot the Voronoi regions associated to each cluster center. (You can assume the minimum and maximum value in the data-set to be the range for each component of $\mathbb{R}^2$).

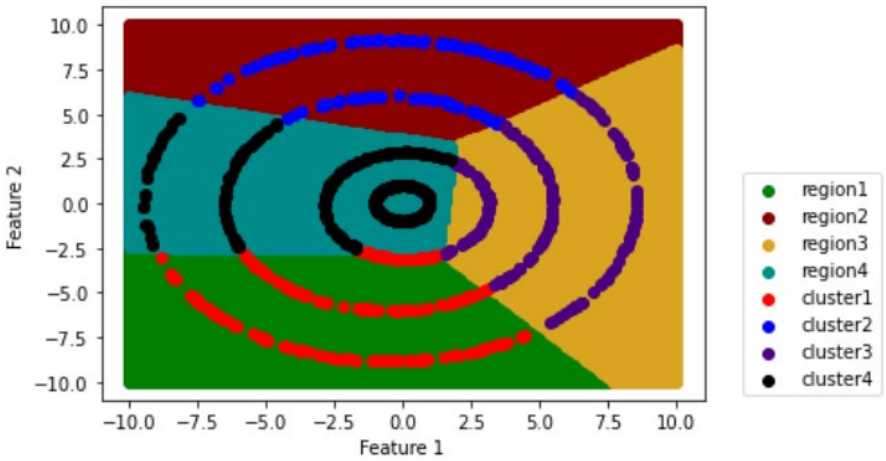
K=2



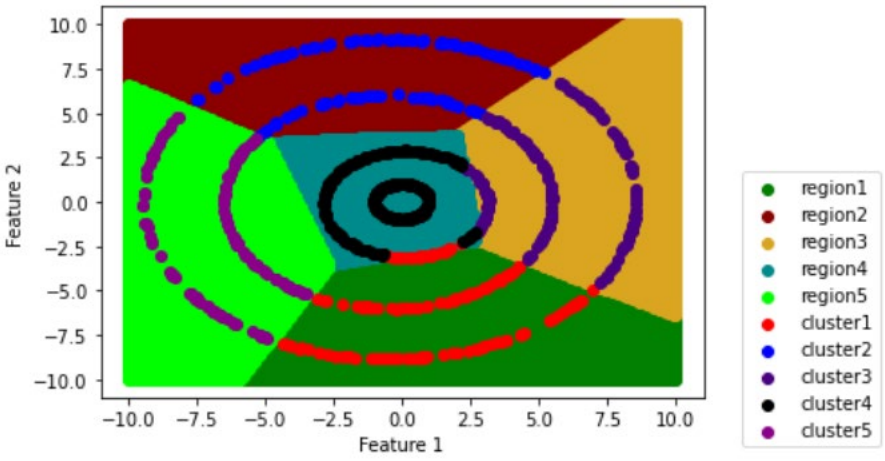
K=3



K=4



K=5

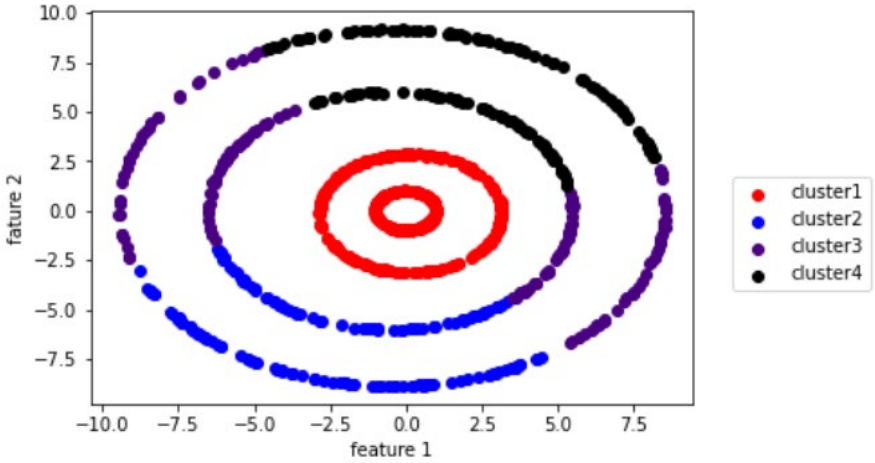


iii. Run the spectral clustering algorithm (spectral relaxation of K-means using Kernel-PCA) $k = 4$. Choose an appropriate kernel for this data-set and plot the clusters obtained in different colors. Explain your choice of kernel based on the output you obtain.

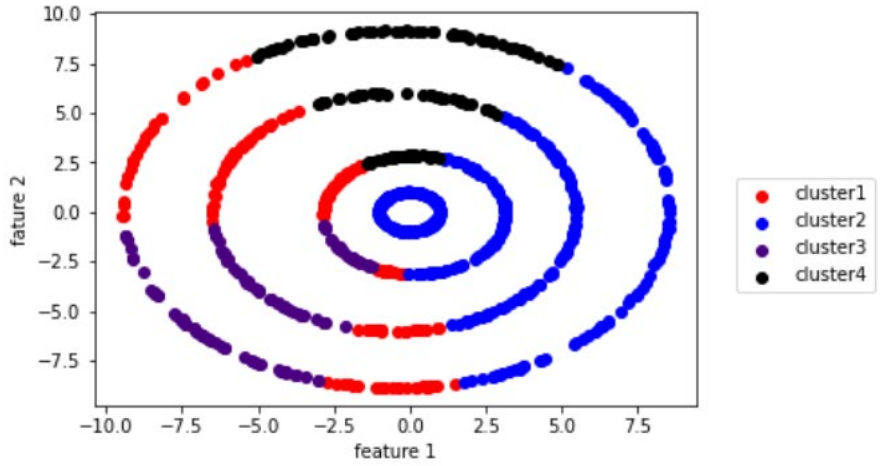
Answer :

Polynomial Kernel with degree 2 give better cluster compared to the other kernel model , as the ideal output is that all the 4 rings should belong to separate cluster , kernel with polynomial degree 2 is the closest

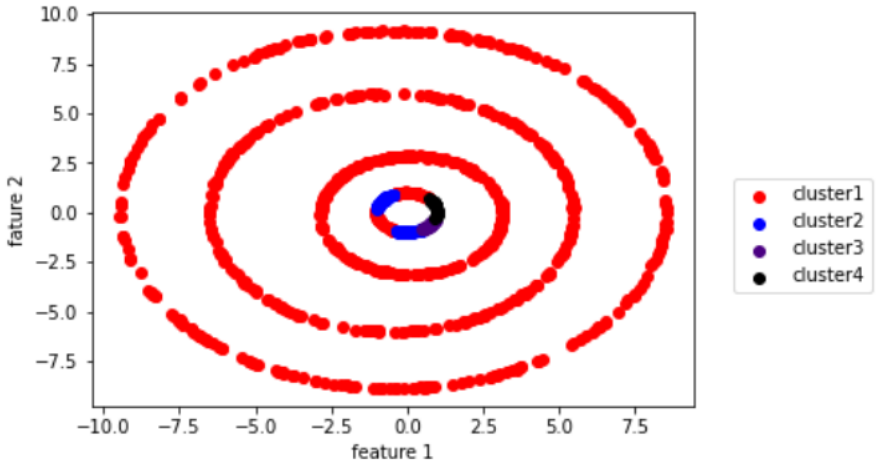
Polynomial degree 2



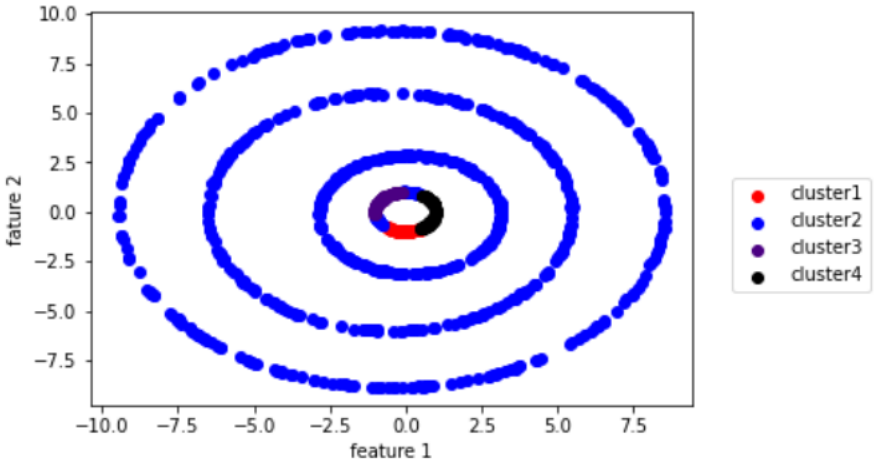
Polynomial degree 3



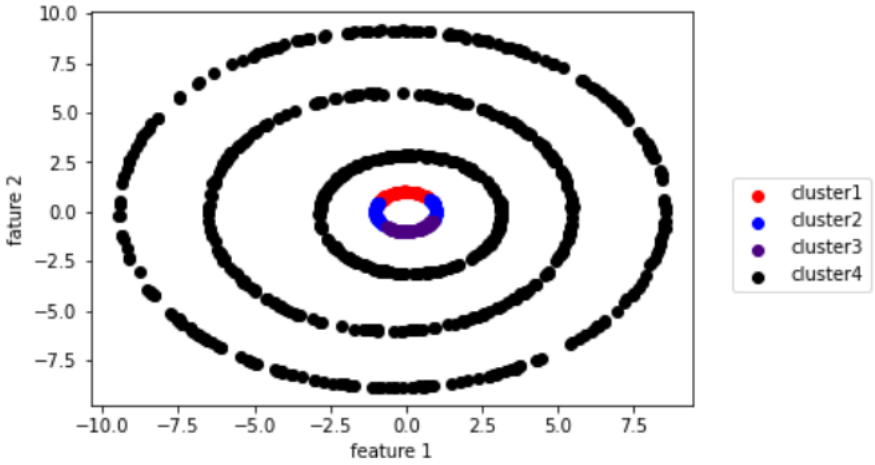
Radial kernel sigma 0.1



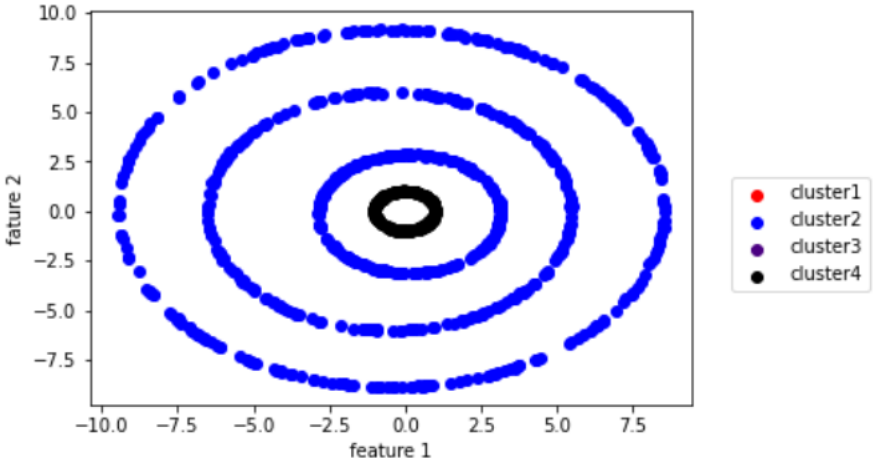
Radial kernel sigma 0.2



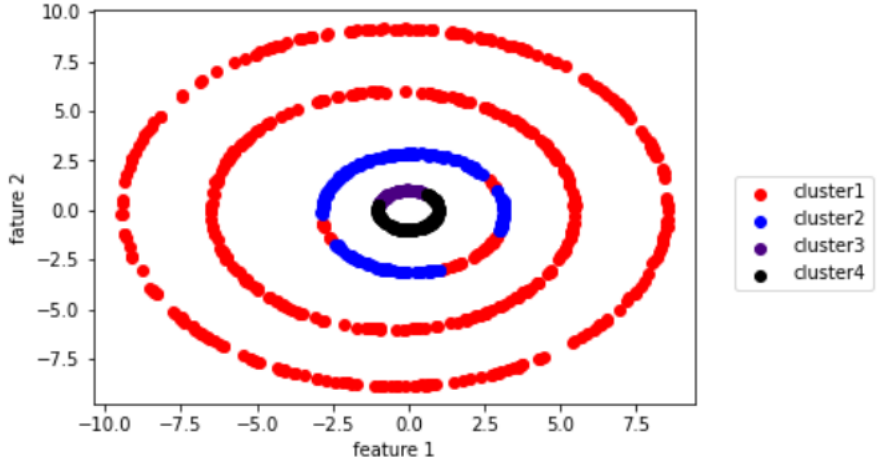
Radial kernel sigma 0.3



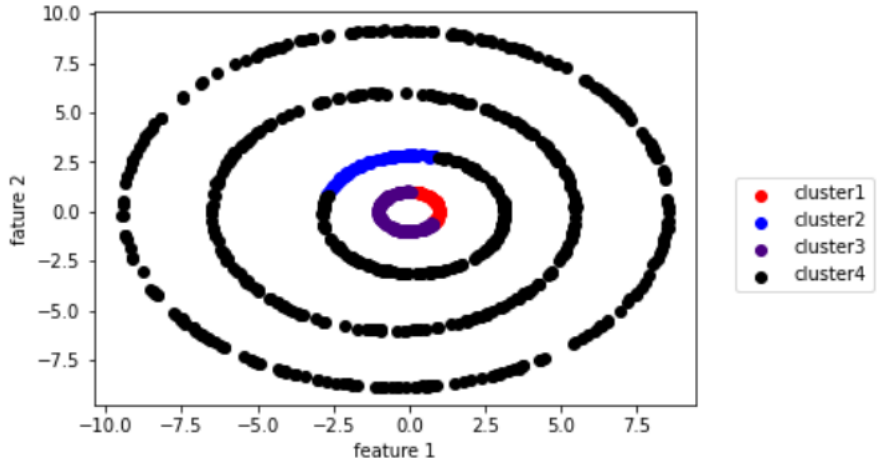
Radial kernel sigma 0.4



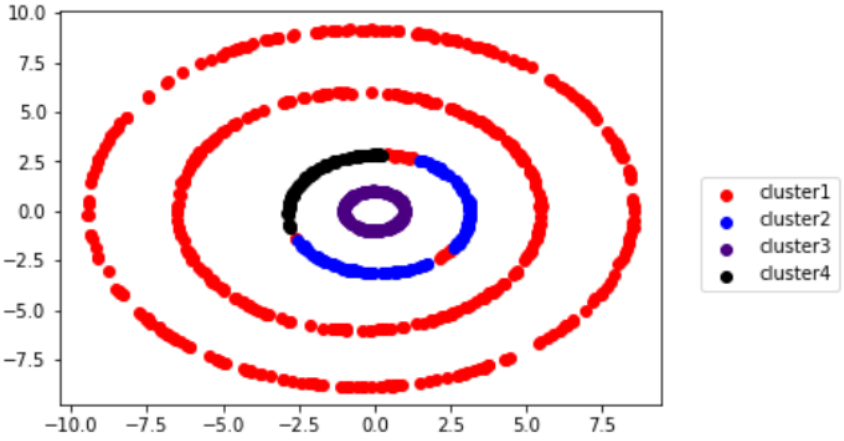
Radial kernel sigma 0.5



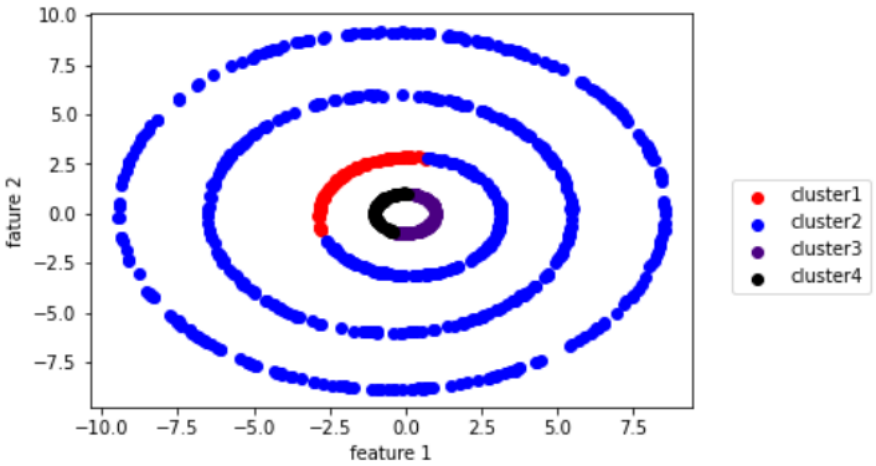
Radial kernel sigma 0.6



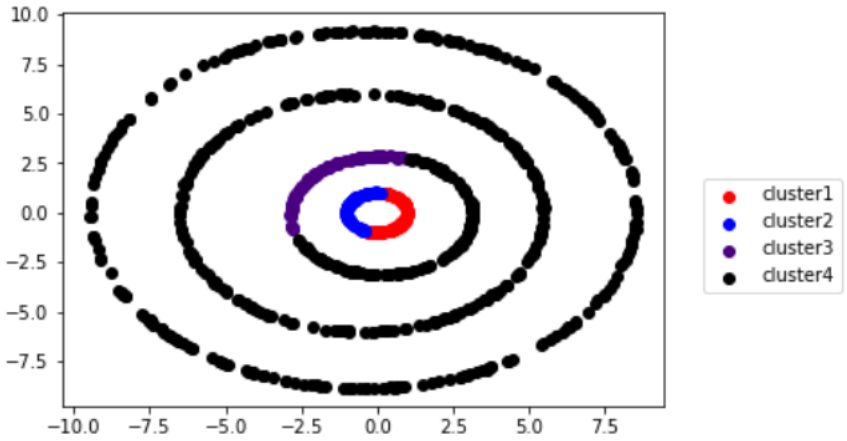
Radial kernel sigma 0.7



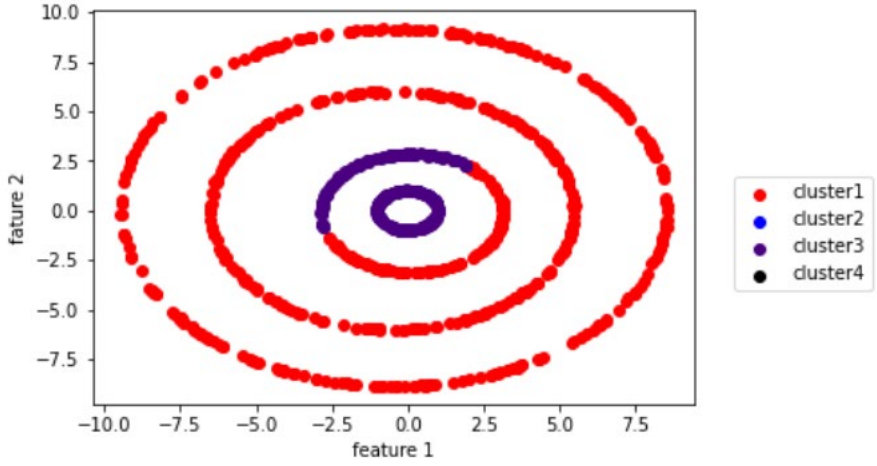
Radial kernel sigma 0.8



Radial kernel sigma 0.9



Radial kernel sigma 1



iv. Instead of using the method suggested by spectral clustering to map eigenvectors to cluster assignments, use the following method: Assign data point i to cluster ℓ whenever

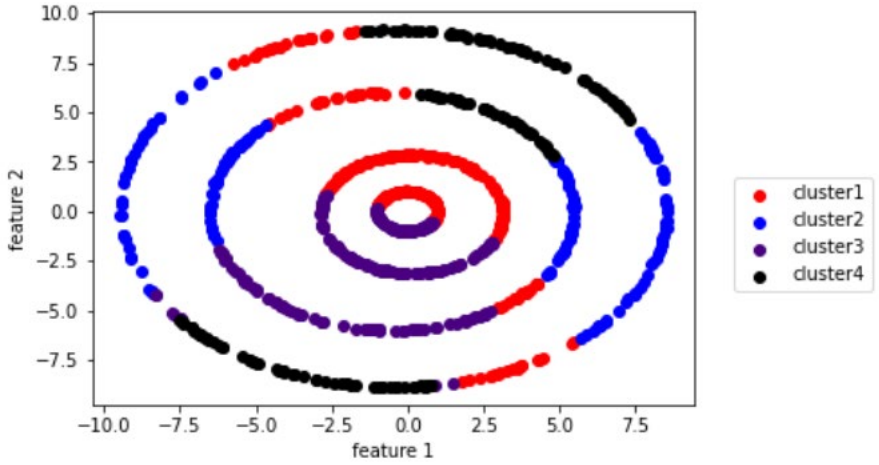
$$\ell = \arg \max_{j=1,\ldots,k} v_i^j$$

where $v^j \in \mathbb{R}^n$ is the eigenvector of the Kernel matrix associated with the j-th largest eigenvalue. How does this mapping perform for this dataset?. Explain your insights.

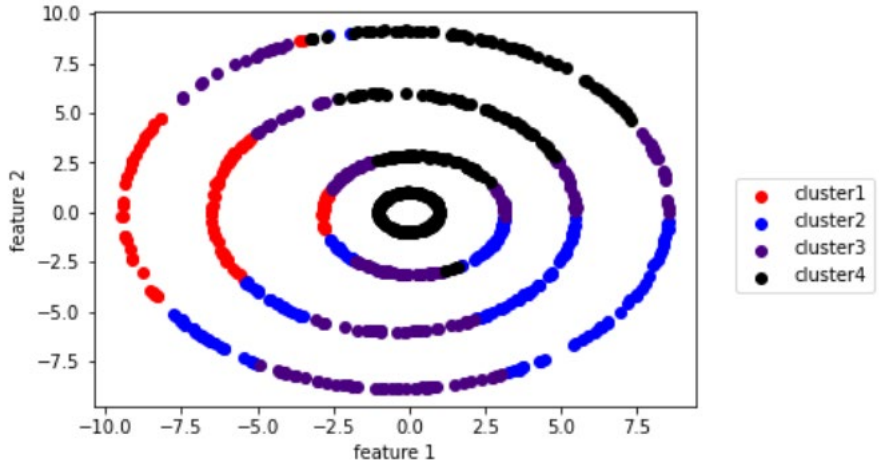
Answer

The mapping perform better than k means for the dataset as the dataset have nonconvex figure , when we do the compute the kernel matrix it contain the dot product of the data points in higher dimension as the top 4 eigen vector of this matrix will maximize the trace of the Kernel matrix , and the datapoints which are similar will belong the same cluster

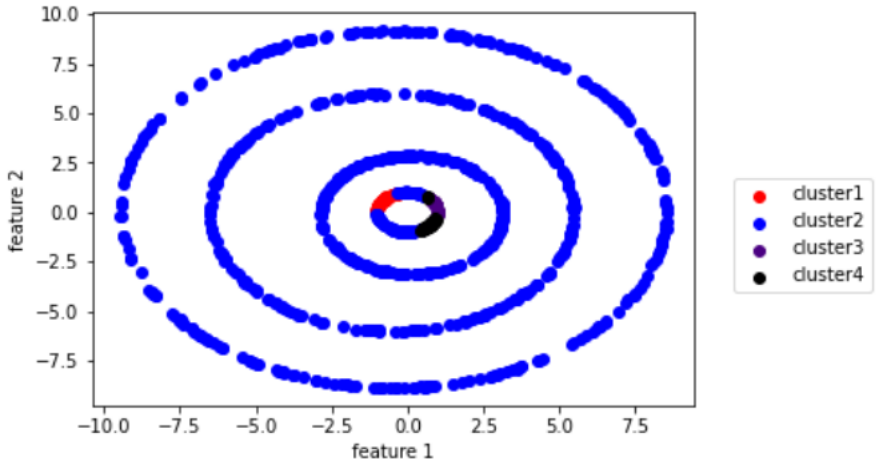
Polynomial kernel degree =2



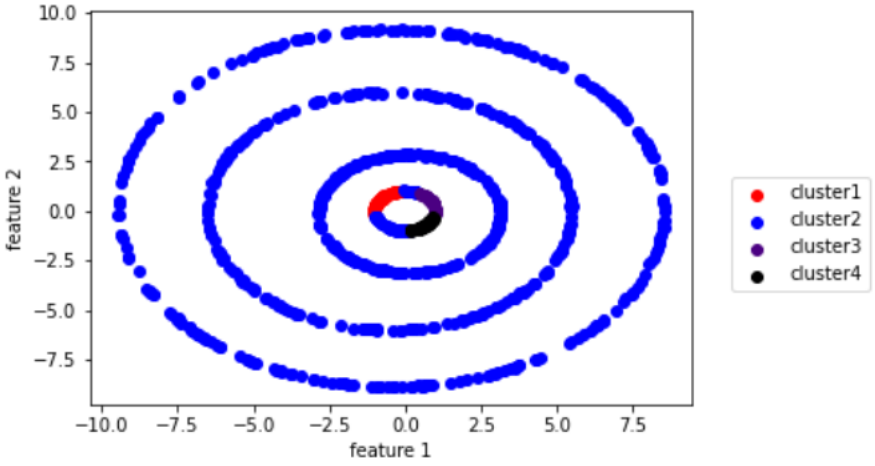
Polynomial kernel degree =3



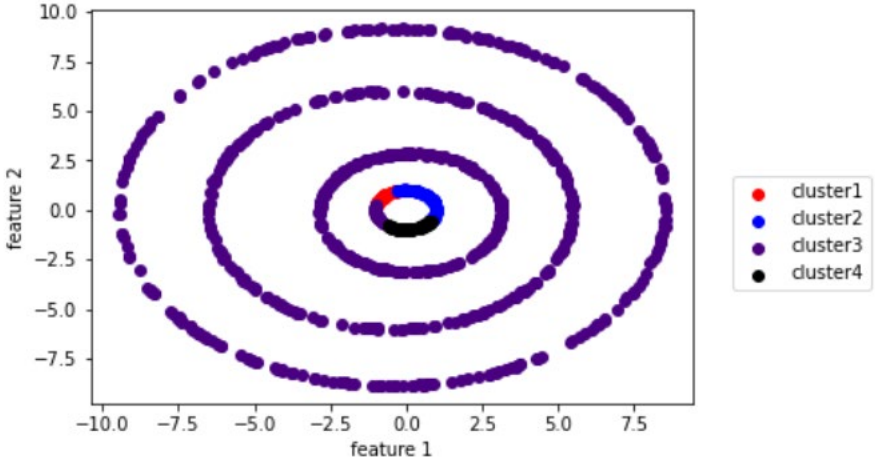
Radial basis sigma =0.1



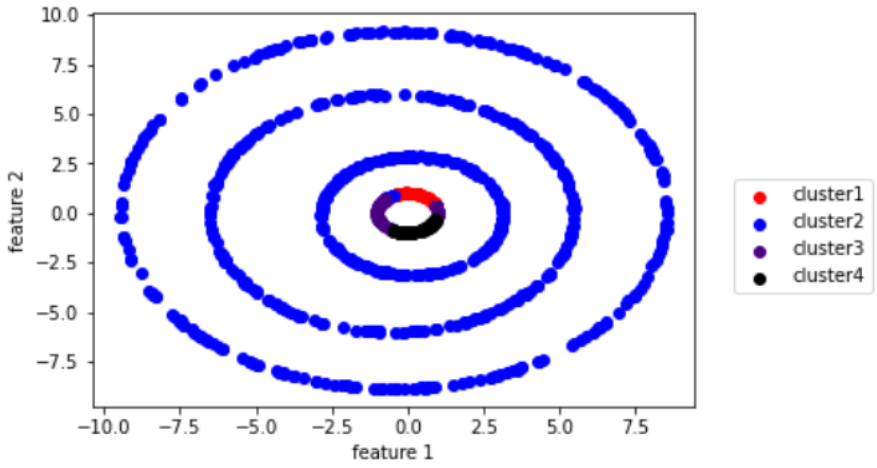
Radial basis sigma =0.2



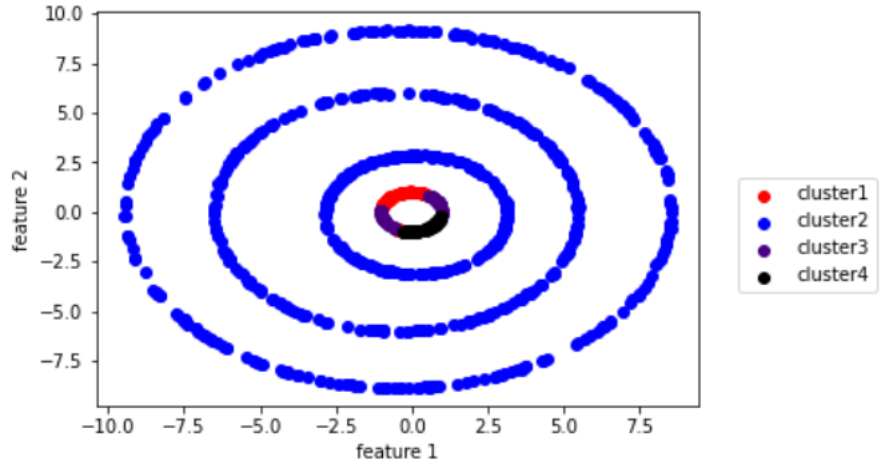
Radial basis 0.3



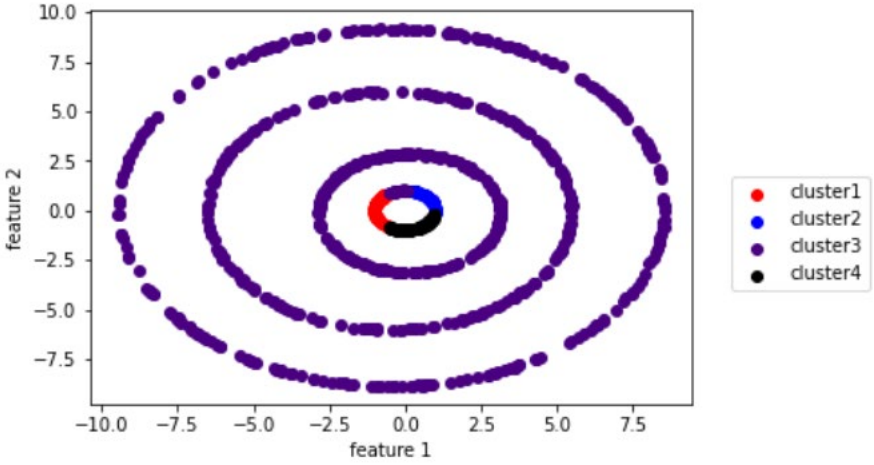
Radial basis sigma = 0.4



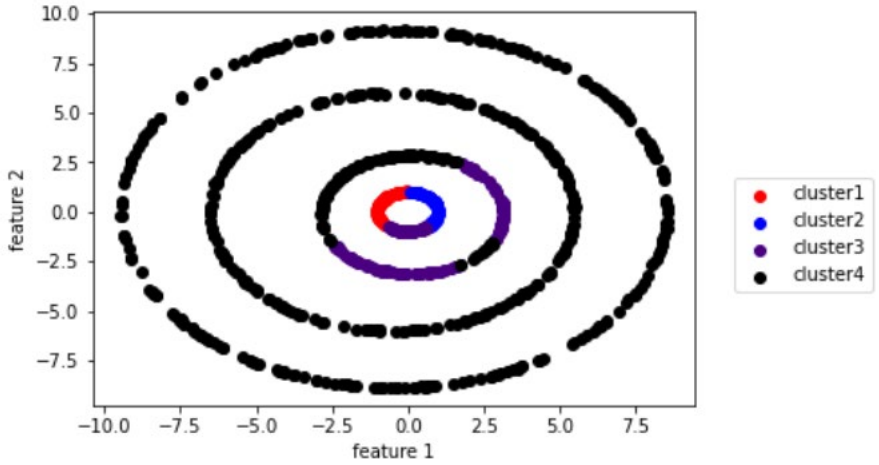
Radial basis sigma = 0.5



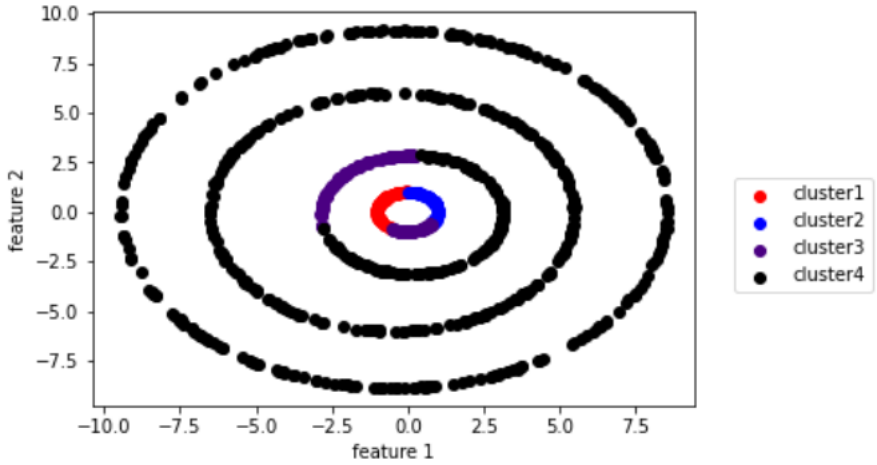
Radial basis sigma =0.6



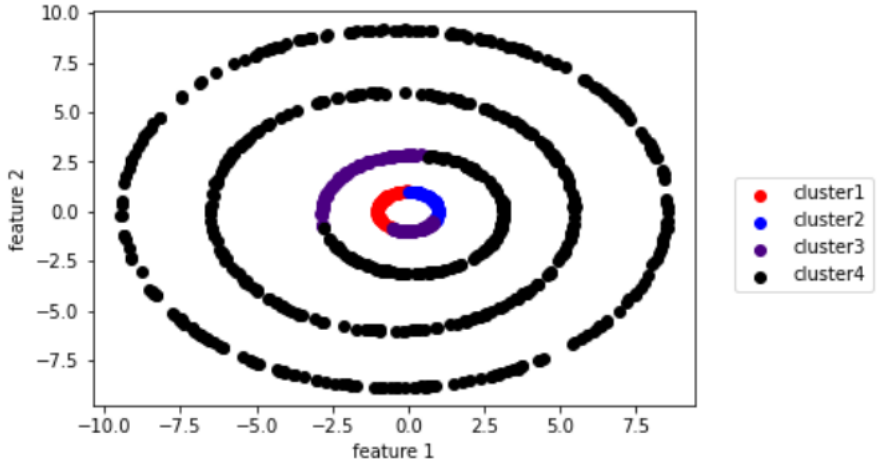
Radial basis sigma = 0.7



Radial basis sigma = 0.8



Radial basis sigma = 0.9



Radial basis sigma =1