Q1

Q1 (i):

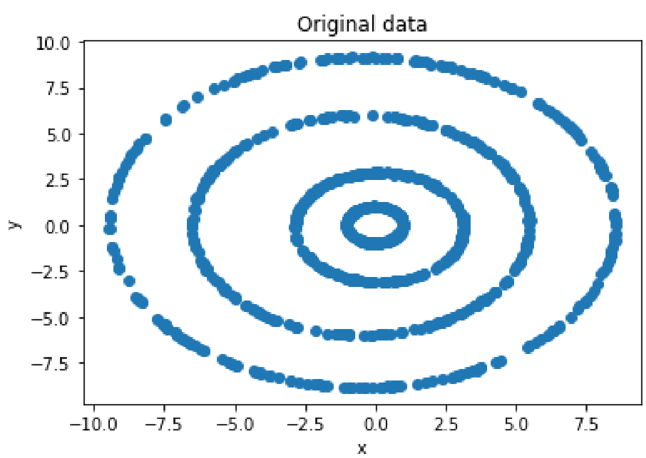
The variance explained by a particular principal component is essentially the associated eigenvalue Therefore, the following results were obtained:

- Variance explained by principal component 1: 17.1319 (54.18%)
- Variance explained by principal component 2: 14.4896 (45.82%)

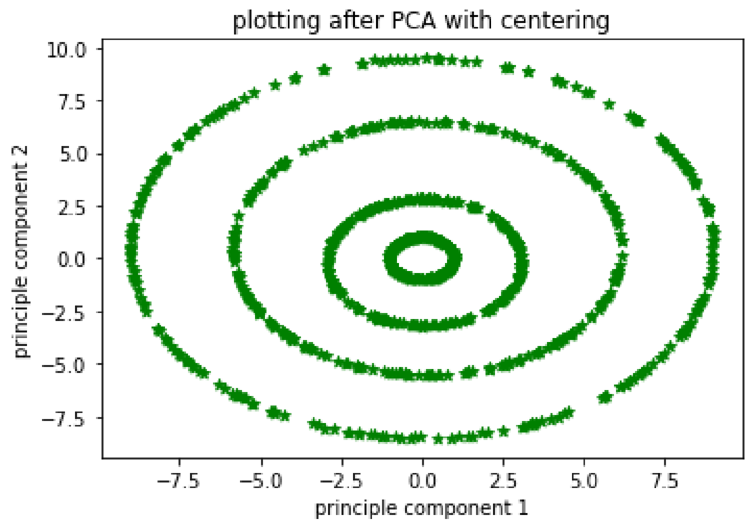
Q1 (ii):

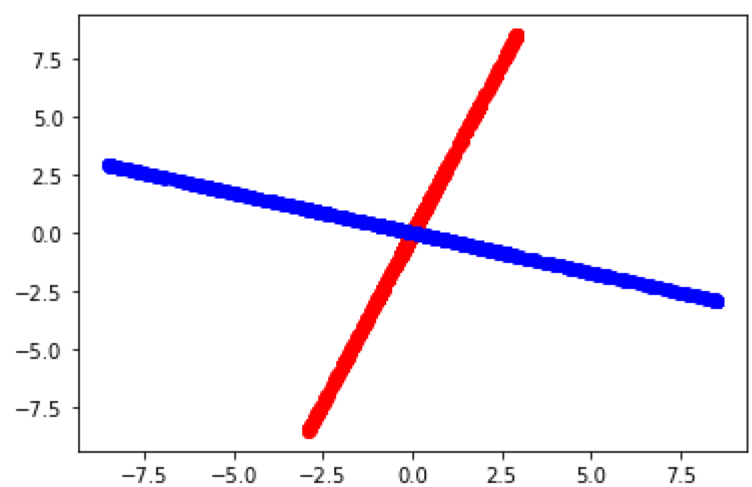
- Centering does not affect the data.
- Centering the data doesn't have any effect on the variance as seen by each of the principal components
- This means that the data is already a centered data

- Original Data:

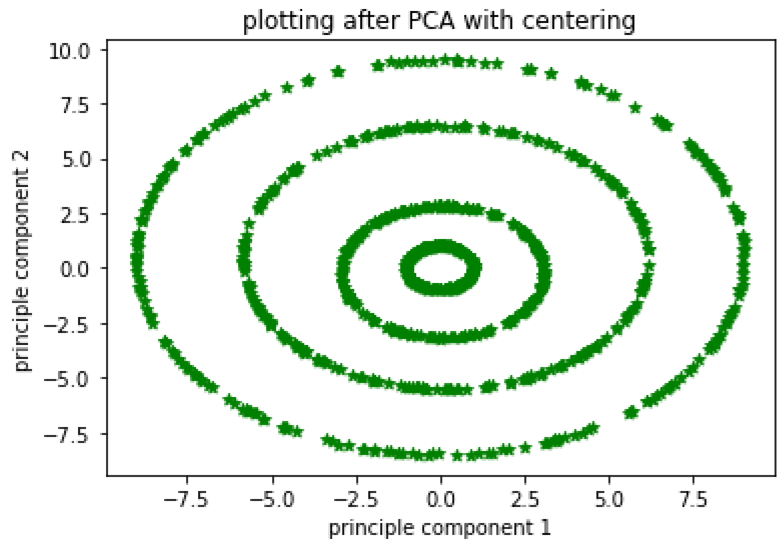


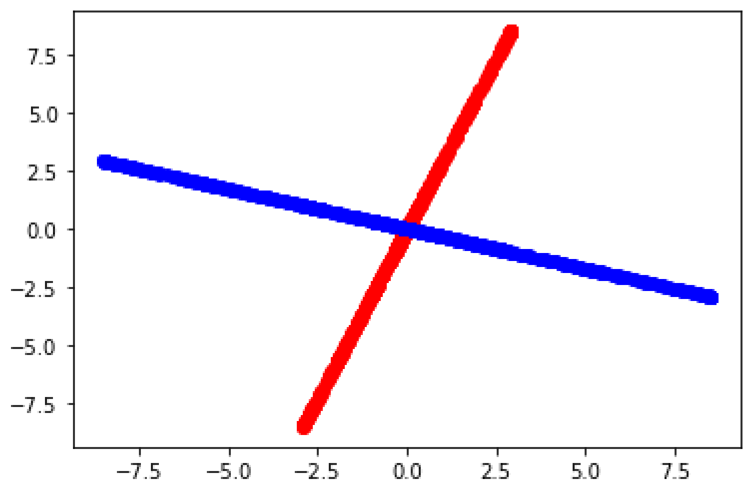
- Centered:





- Without Centering:

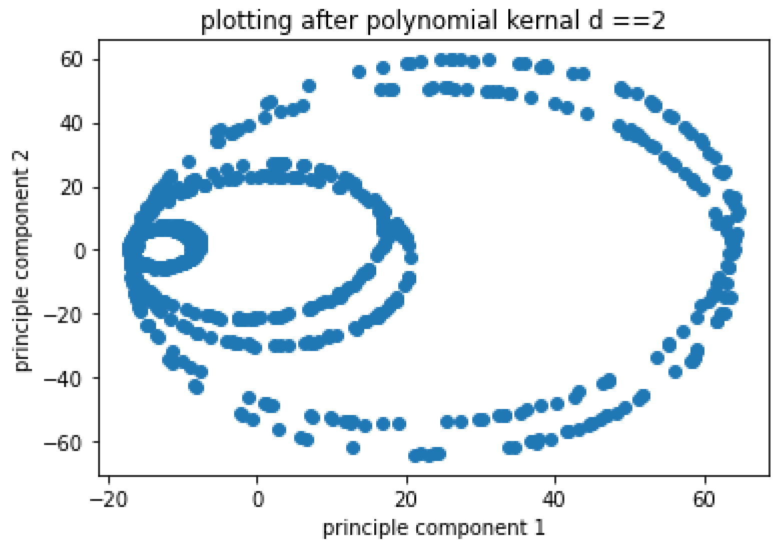




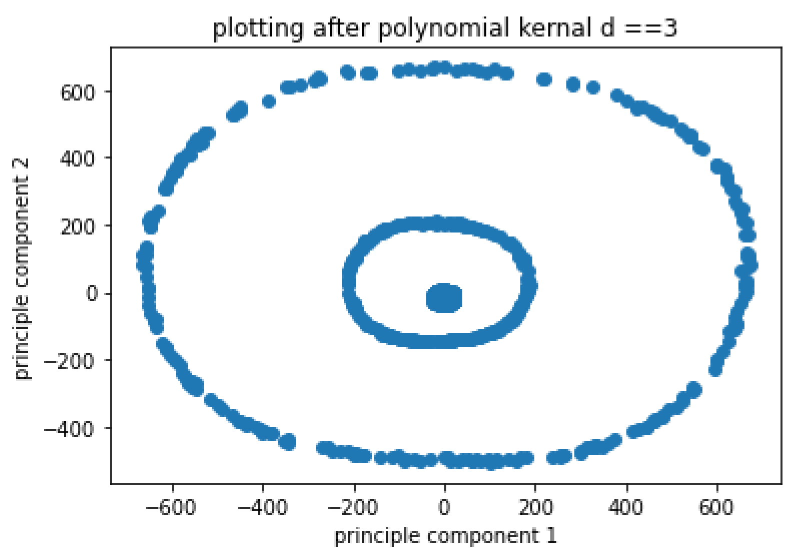
Q1 (iii) A:

Kernel PCA was implemented using the given kernels

- Polynomial Kernel : k (x, y) = (1 + x T y) ^d for d=2

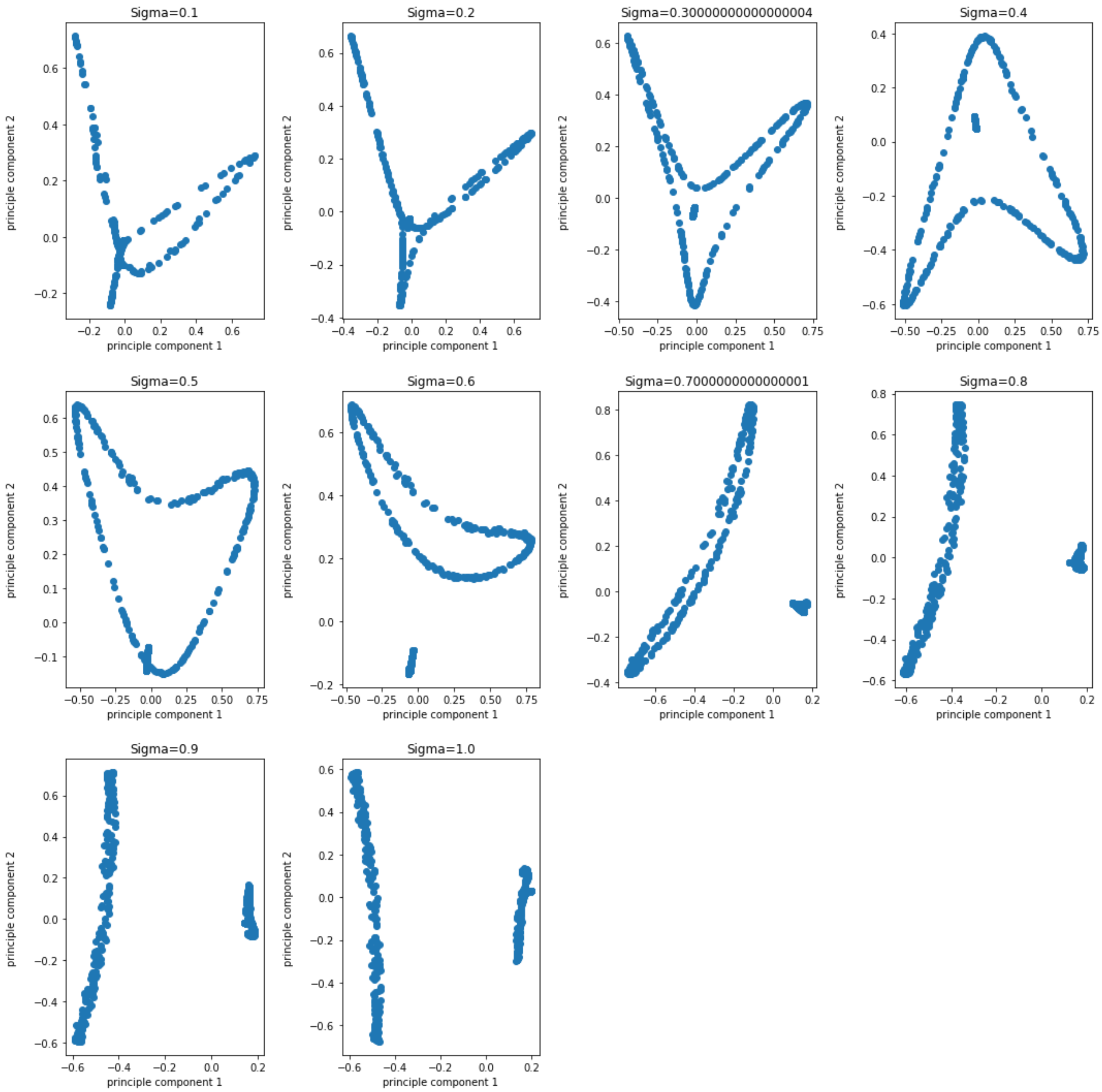


- Polynomial Kernel : k (x, y) = (1 + x T y) ^d   for d=3



Q1 (iii) B:

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



Q1 (iv):

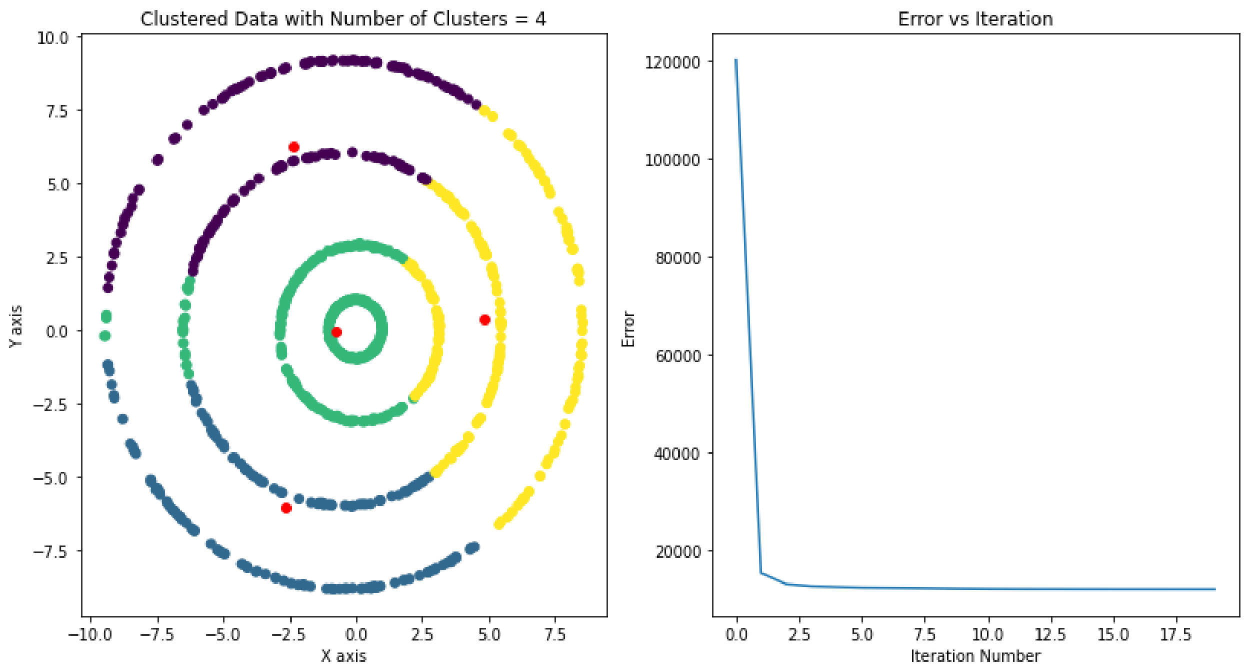
Polynomial kernal with d = 2 find the better clusters in the data.

Variance is high for d=2, Also we need 3 principle components for d=2, and 4 for d=3

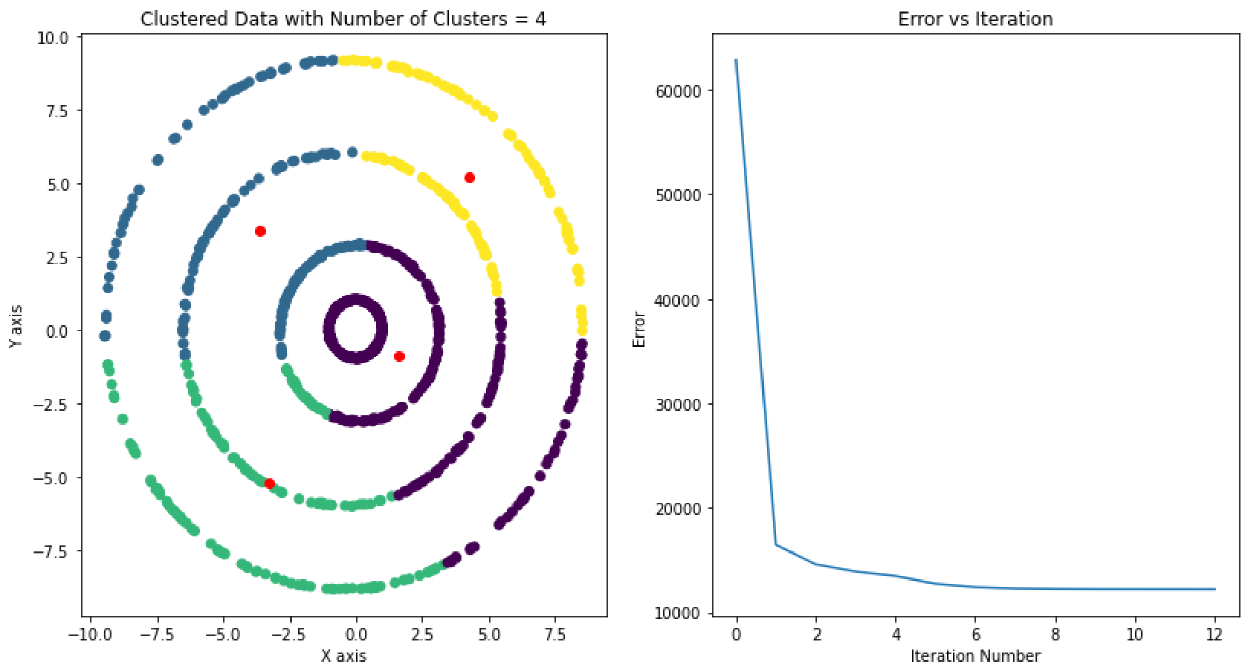
Q2:

Q2(i):

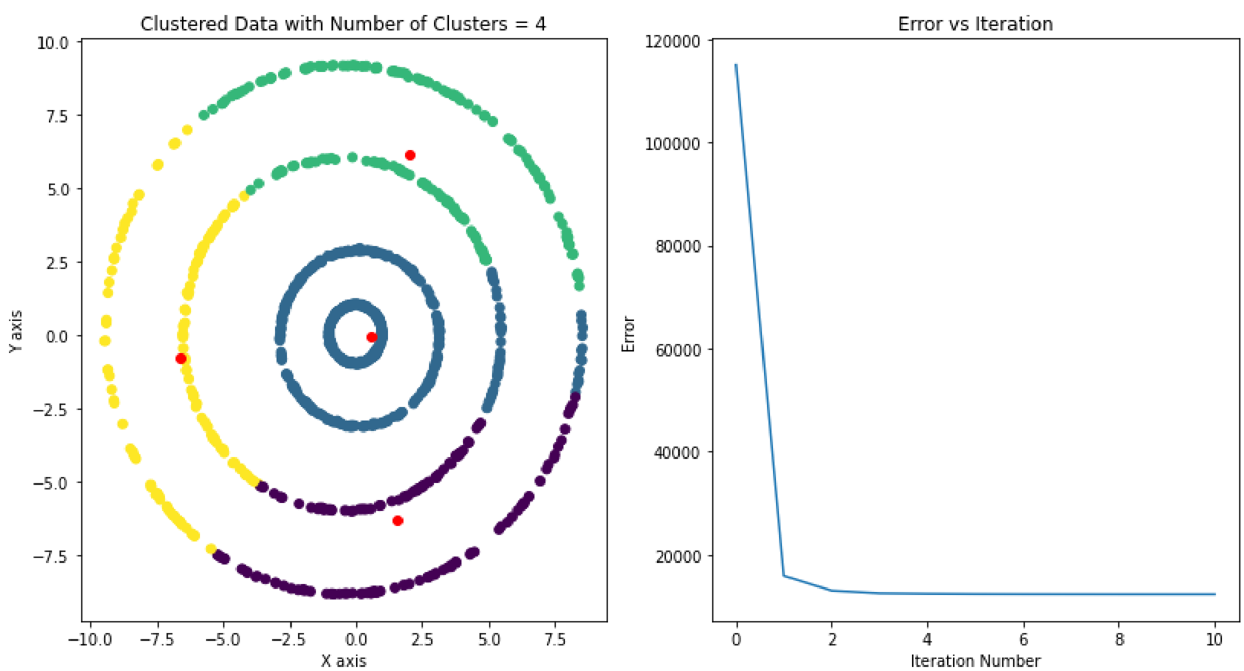
Run 1:



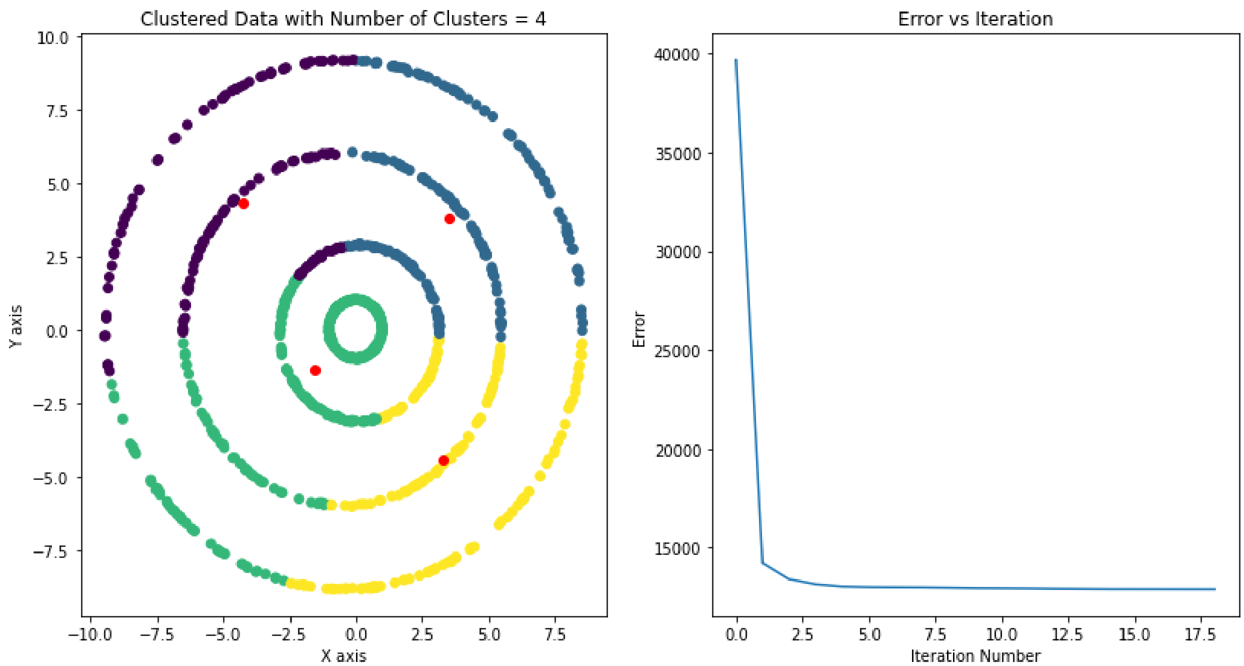
Run 2:


Clustered Data with Number of Clusters = 4
Y axis
X axis
Error vs Iteration
Error
Iteration Number

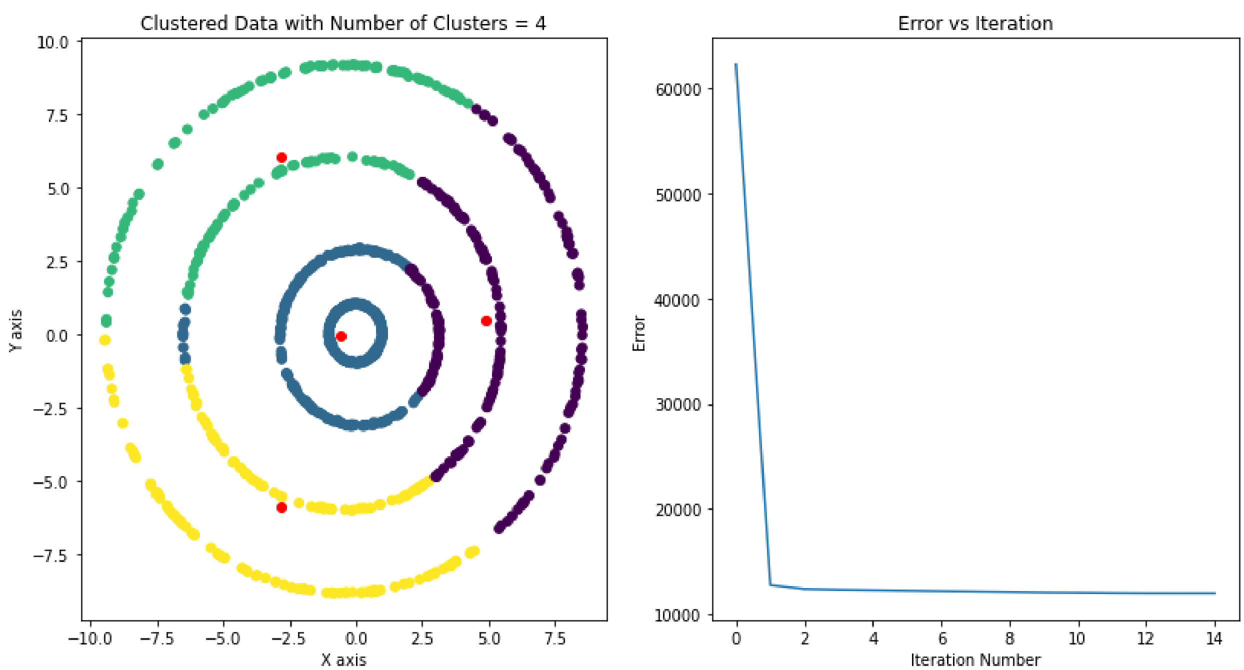
Run 3:


Clustered Data with Number of Clusters = 4
Y axis
X axis
Error vs Iteration
Error
Iteration Number

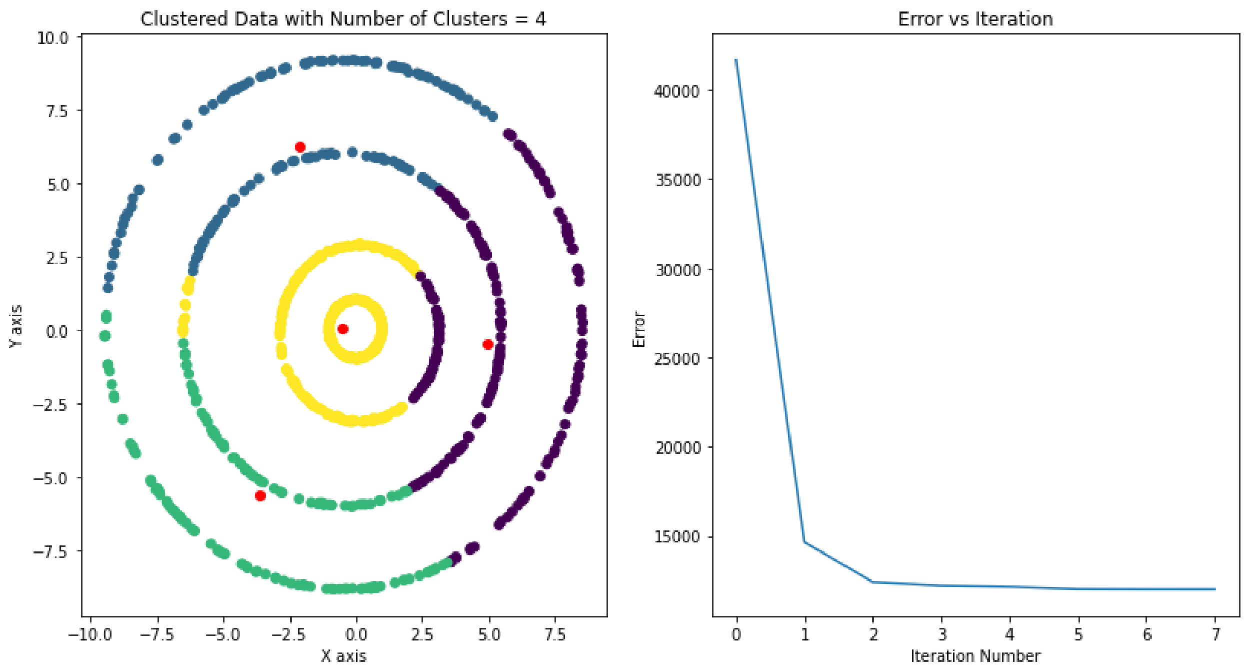
Run 4:


Clustered Data with Number of Clusters = 4
Y axis
X axis
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
Error vs Iteration
Error
Iteration Number
40000
35000
30000
25000
20000
15000
12.5
15.0
17.5

Run 5:


Clustered Data with Number of Clusters = 4
Y axis
X axis
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
Error vs Iteration
Error
Iteration Number
60000
50000
40000
30000
20000
10000
0
2
4
6
8
10
12
14

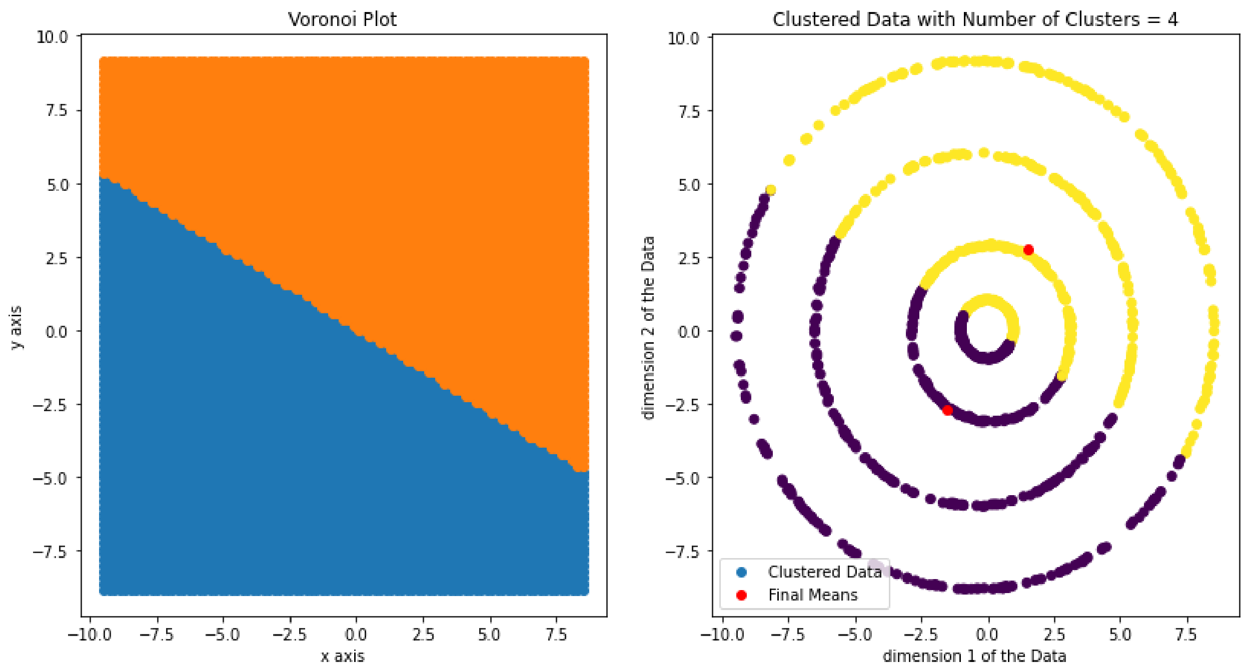
Run 6:



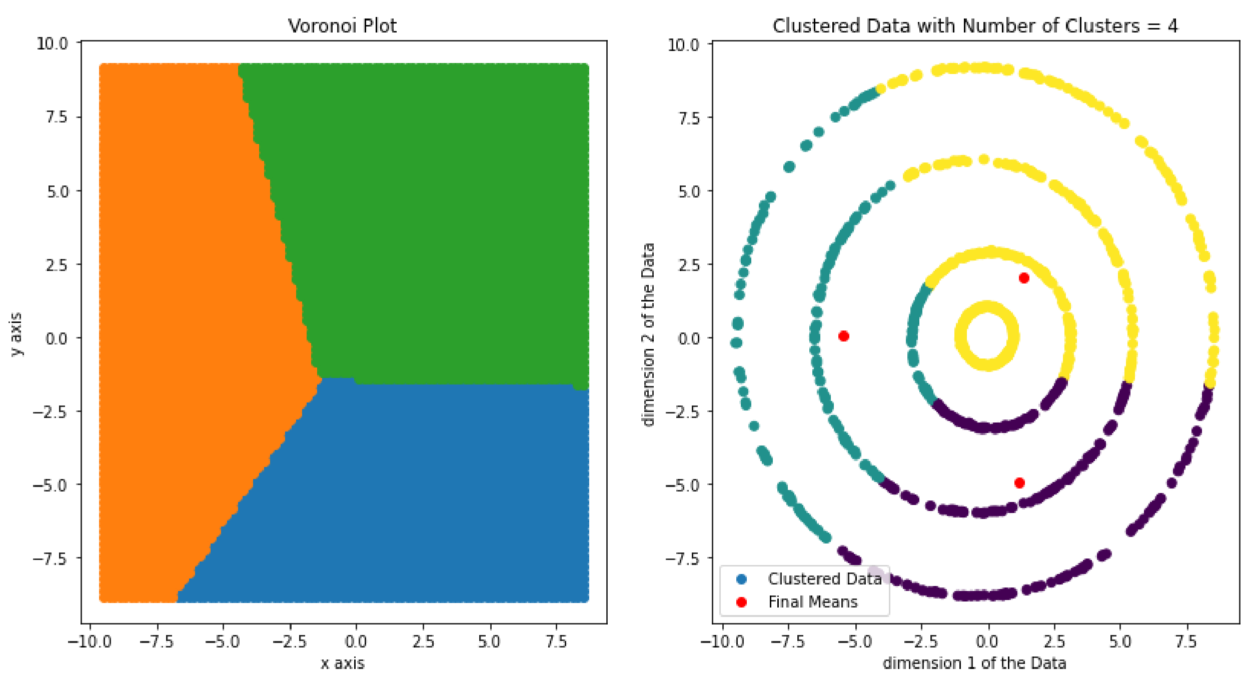
Clustered Data with Number of Clusters = 4
Y axis
X axis
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
Error vs Iteration
Error
Iteration Number
40000
35000
30000
25000
20000
15000
0
1
2
3
4
5
6
7

Q2(ii):

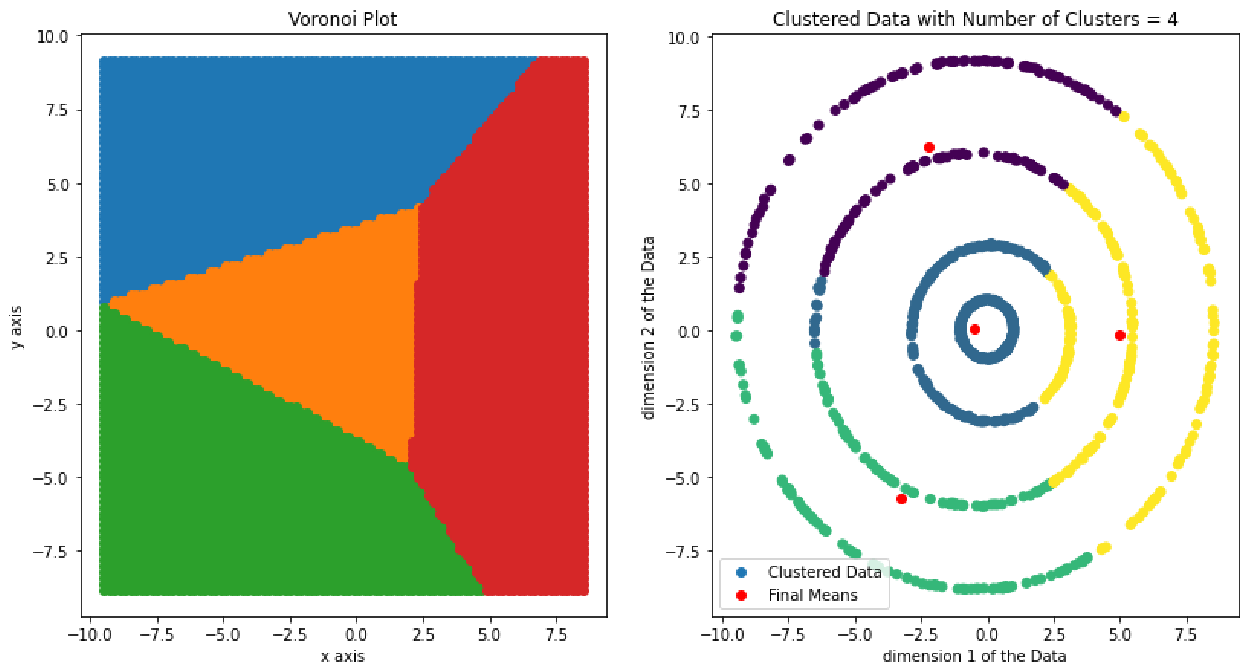
K=2


Voronoi Plot
y axis
x axis
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
Clustered Data with Number of Clusters = 4
dimension 2 of the Data
dimension 1 of the Data
Clustered Data
Final Means

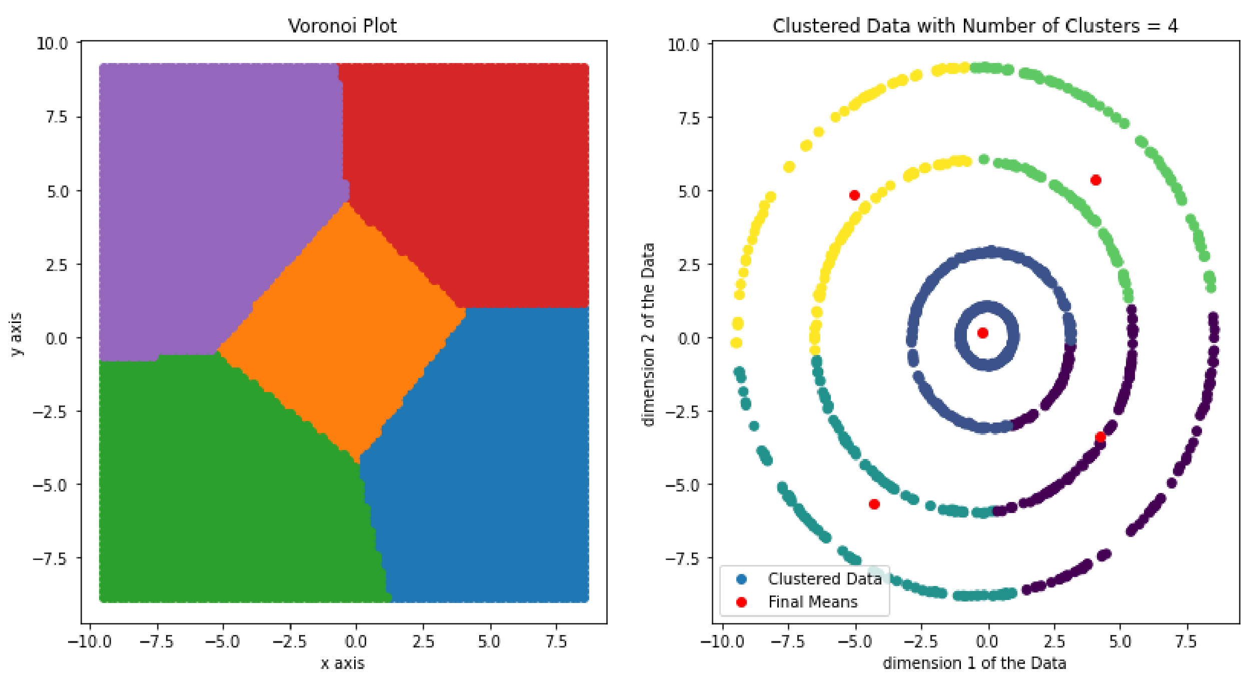
K=3


Voronoi Plot
y axis
x axis
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
Clustered Data with Number of Clusters = 4
dimension 2 of the Data
dimension 1 of the Data
Clustered Data
Final Means

K=4


Voronoi Plot
y axis
x axis
Clustered Data with Number of Clusters = 4
dimension 2 of the Data
dimension 1 of the Data
Clustered Data
Final Means

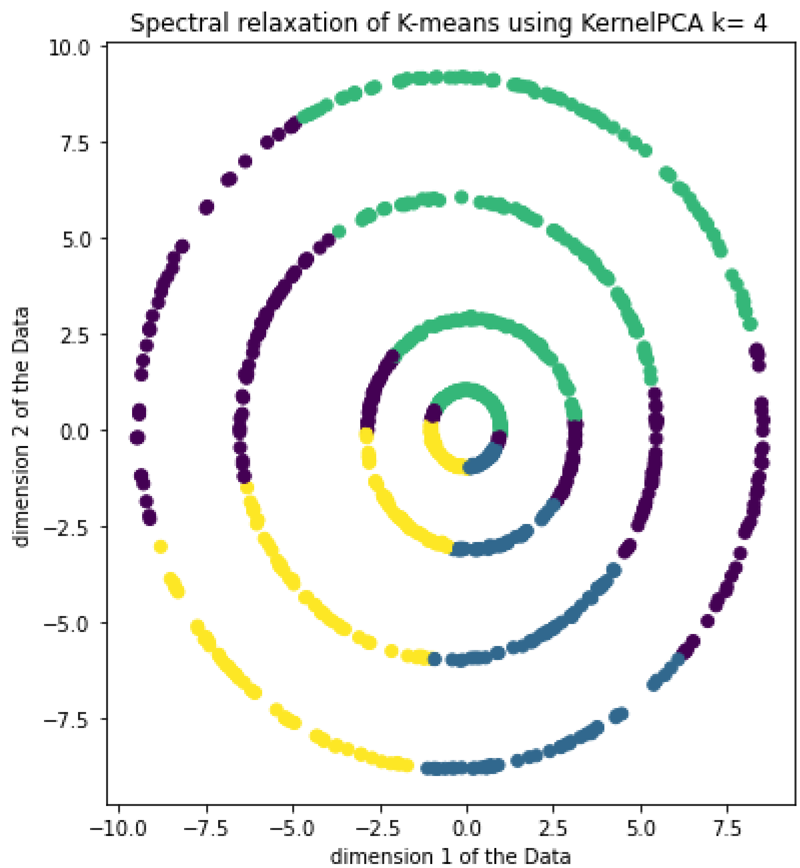
K=5


Voronoi Plot
y axis
x axis
Clustered Data with Number of Clusters = 4
dimension 2 of the Data
dimension 1 of the Data
Clustered Data
Final Means

Q2(iii):

Polynomial performs better than Exponential Kernel function

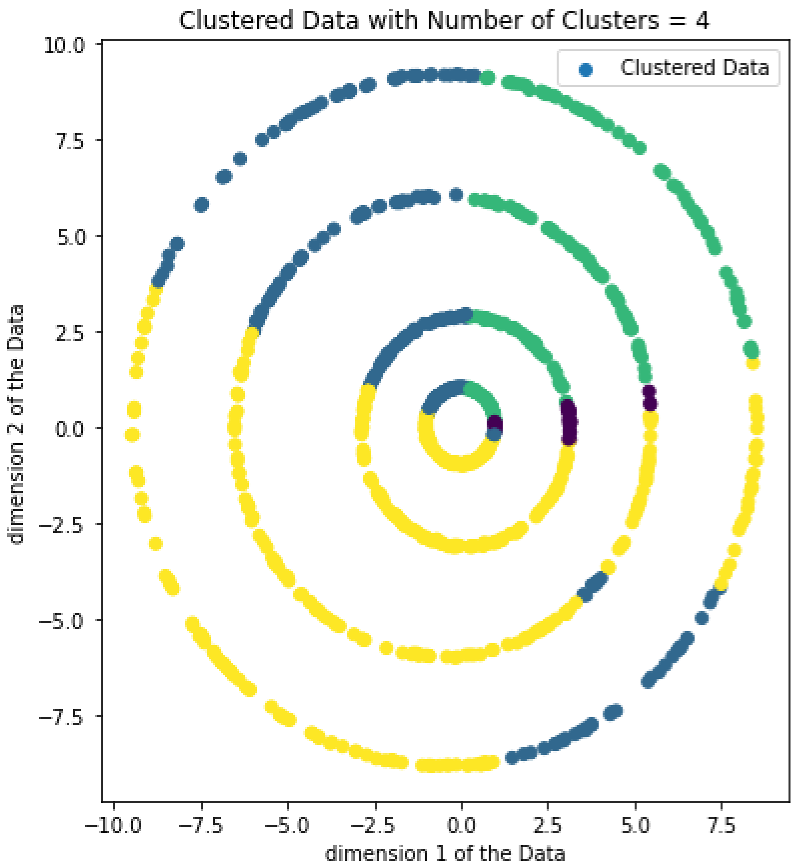
For d=2 in polynomial ,we get clearly distinct clusters



Q2(iv):

The results were visualised and it was seen that it was not as good as spectral clustering.

Because in given function clustered points are scattered across different dimensions



Clustered Data with Number of Clusters = 4
Clustered Data
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
dimension 2 of the Data
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
dimension 1 of the Data