Project 2

ENEE436 Foundation of Machine Learning

University of Maryland, College Park

Instructor: Joseph JaJa

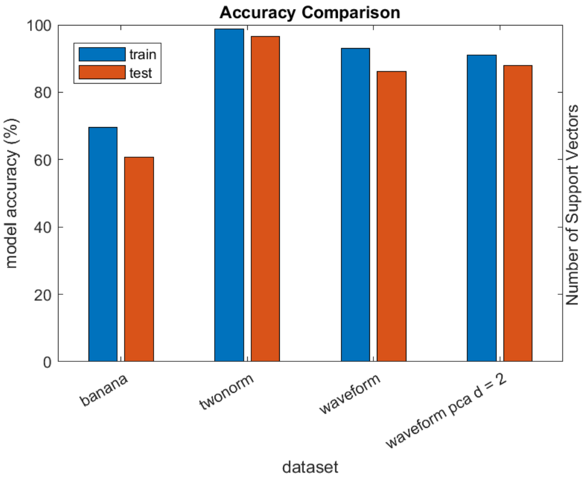
Yuchen Zhou

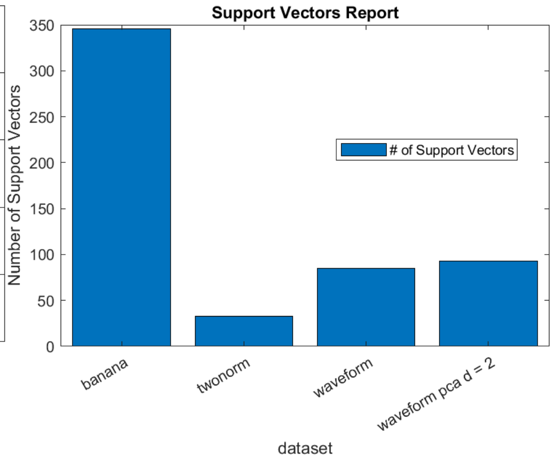
12/14/20

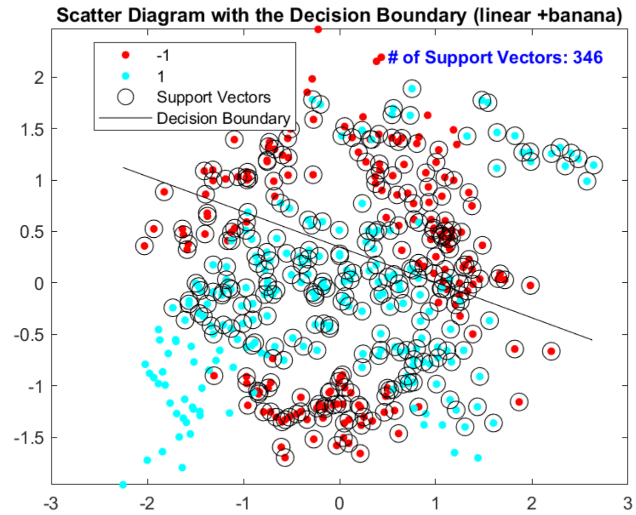
Table of Contents

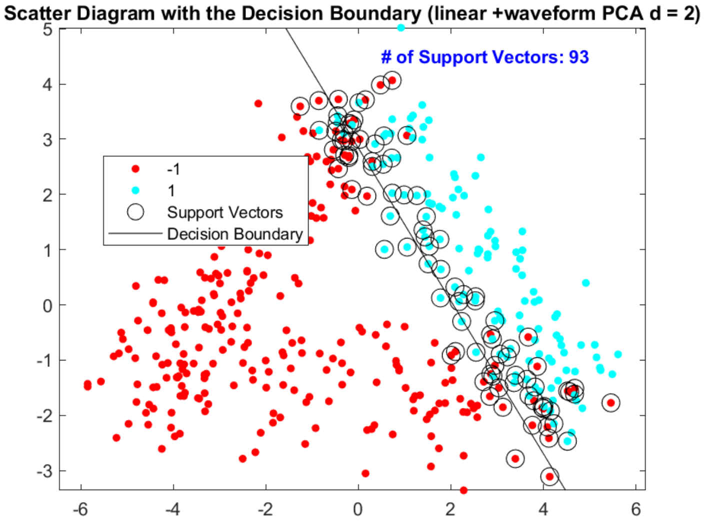
Task 1 – Support Vector Machine

Linear Kernel

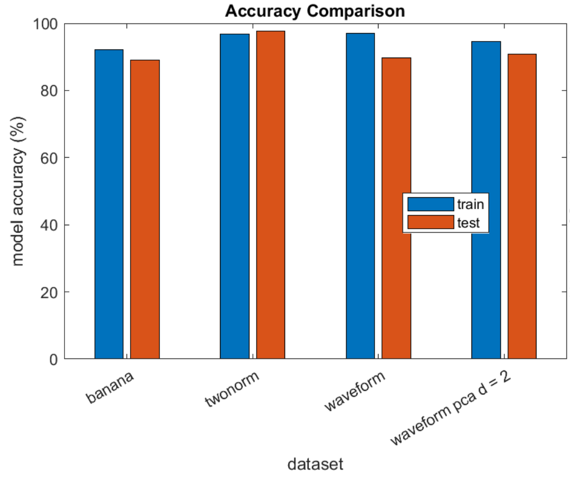


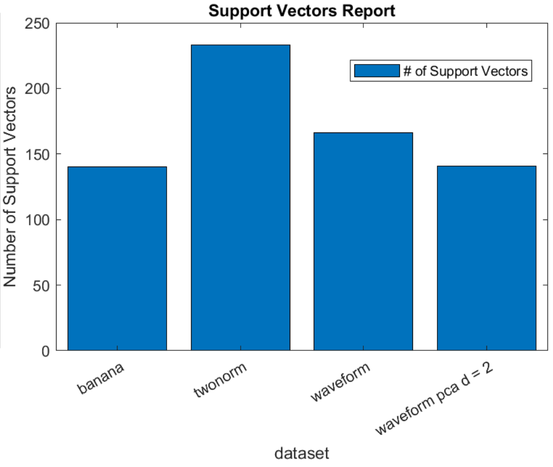


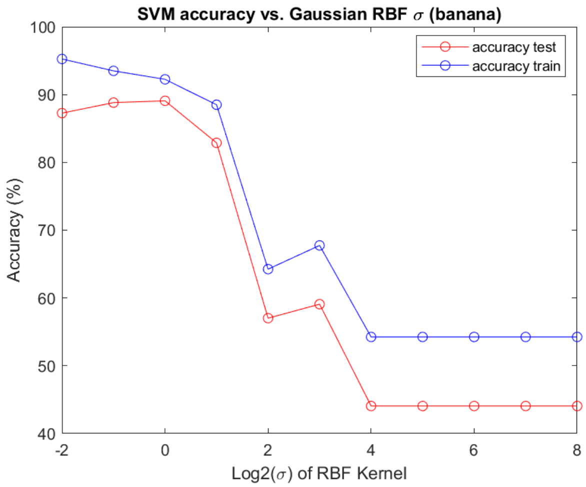


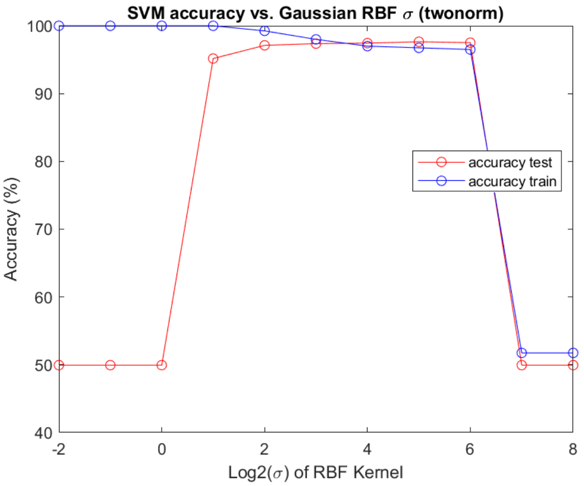


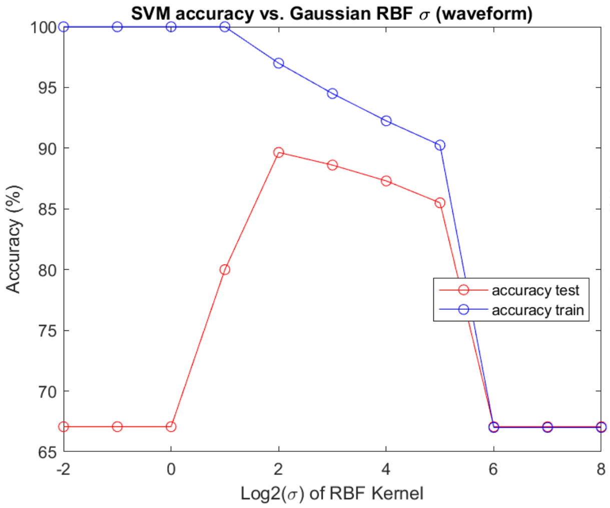
Gaussian RBF

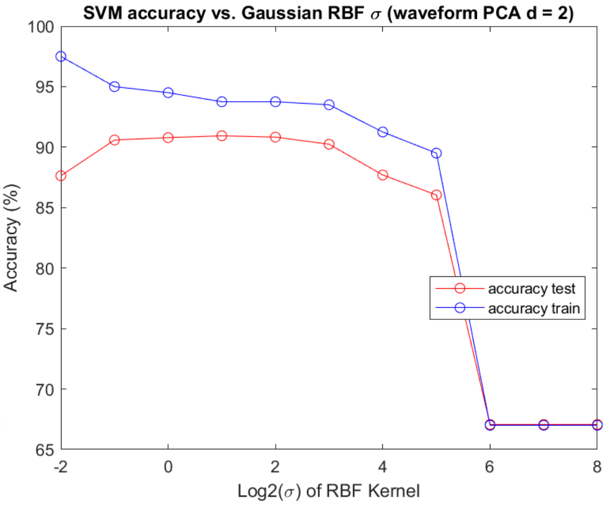


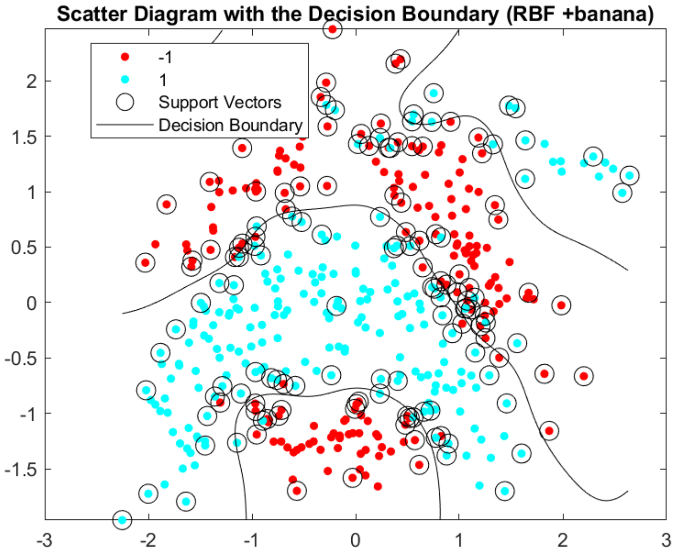


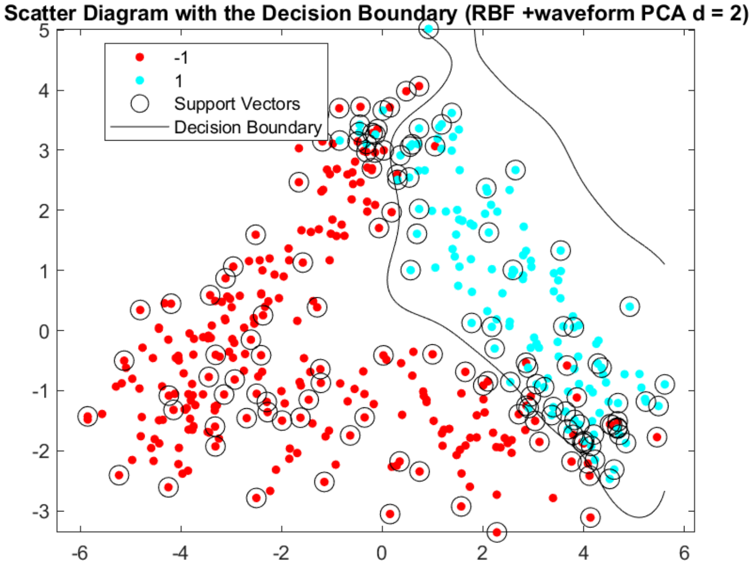












Optimal σ	
Dataset	σ
Banana	1
Twonorm	32
Waveform	4
Waveform PCA	1

Discussion:

Cross Validation Step (avoids overfitting):

1. Randomly shuffle the training data and fix it
2. Evenly divide the shuffled training data into 5 groups, {D1…...D5}
3. Select one σ from all target σ's
4. Select one group for testing and rest of the groups for training. Train a SVM model using an RBF kernel function with σ equal to the one selected from step 3, then record the accuracy of the testing group using the trained model.
5. Repeat step 4 until every group is used for testing. Calculate and record the average accuracy of selected σ
6. Go back to step 3. Stop until the average accuracy of all σ's is calculated.
7. Selected the σ that has the highest average accuracy as the optimal σ for the dataset

Banana Dataset:

Banana Dataset is clearly not linearly separable as the plot shown. Therefore, using the linear kernel will yield a SVM model that has a low accuracy (69.5% training, 60.57% testing). Using the RBF kernel creates a model that has a much higher accuracy and much decisive boundary as the plot shown (92.25% training, 89.08 % testing).

PCA Waveform:

Principle Component Analysis on the Waveform dataset is working effectively, even when using a linear function. The dimensional-reduced dataset is more linearly separable as the accuracy is increased compared to the accuracy of the original dataset (86.26% original vs. 87.93% reduced). With high accuracies in both cases, indicating that both sets are very close to linear separable. Just like using the linear kernel, the RBF kernel works slightly better on the dimensional-reduced dataset (89.65% original vs. 90.78% reduced). The decision boundary selected by using the RBF kernel works more accurate in classification.

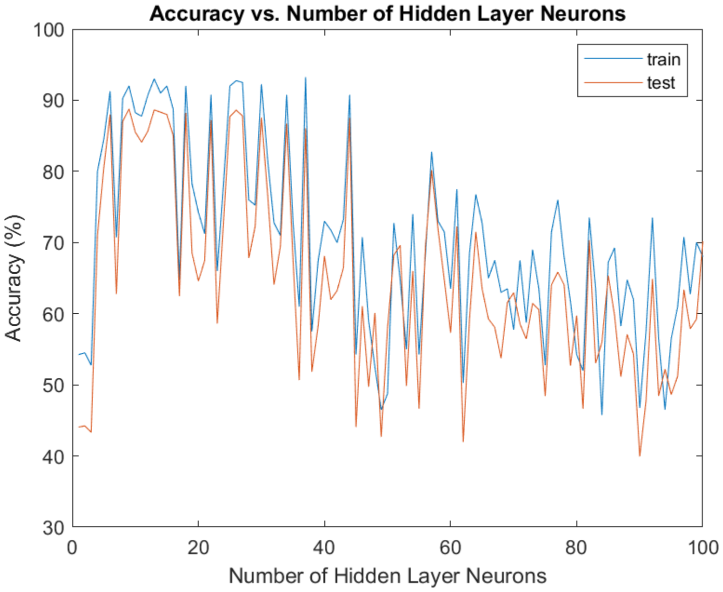
Task 2 – Neural Network

All Neuron Networks has N-X-2 numbers of neurons.

N = Number of features

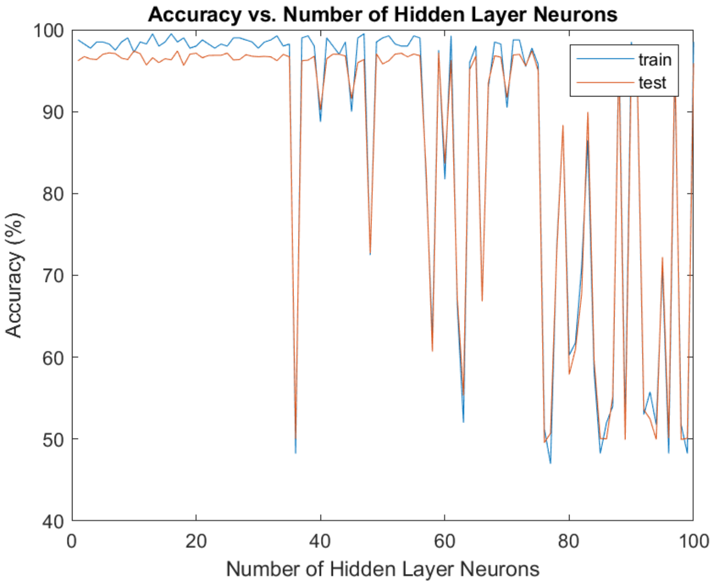
X = Number of Hidden Layer Neurons (to be tuned)

Dataset 1: Banana



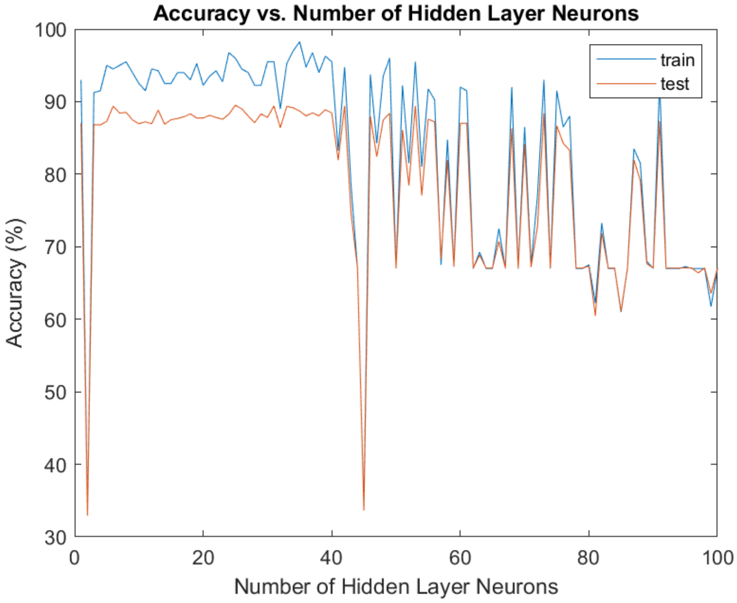
Optimal X = 17, Testing Accuracy = 88.76%

Dataset 2: Twonorm



Optimal X = 6, Testing Accuracy = 97.63%

Dataset 3: Waveform



Optimal X = 26, Testing Accuracy = 89.87%

Discussion:

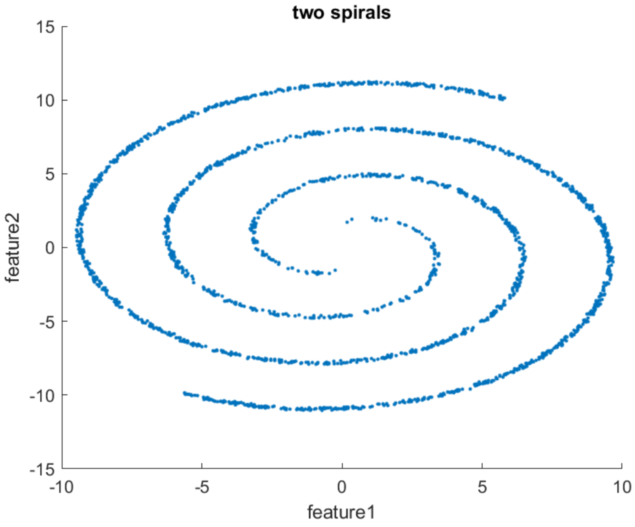
Neuron Network works very well on data that is not linearly separable (banana for example). When SVM uses the linear kernel on data that is not linearly separable, the accuracy of the model is significantly reduced. However, when the SVM model selects the RBF kernel, which resolves the problem of linear separation, there is no remarkable differences between the output of SVM and neuron network.

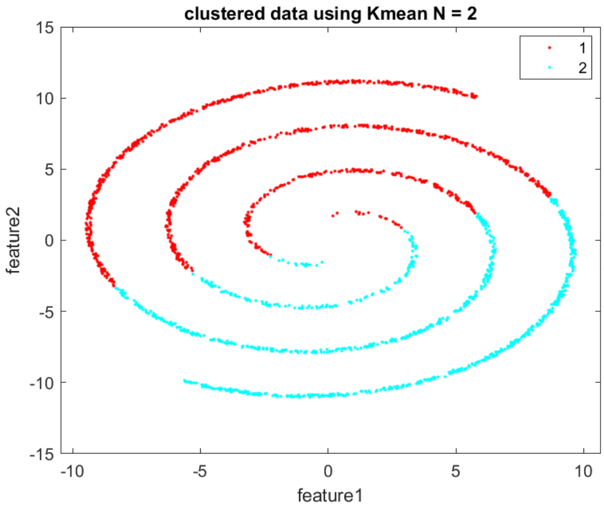
Dataset	Optimized SVM accuracy (BRF)	Optimized NN accuracy
Banana	89.08	88.76
Twonorm	97.64	97.63
Waveform	89.65	89.87

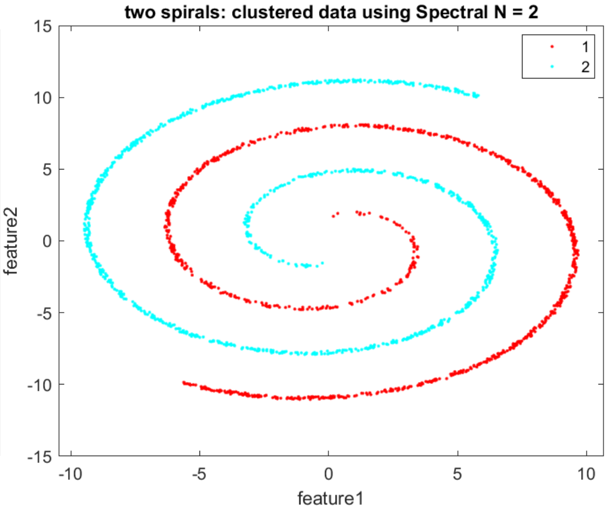
Task 3 – Unsupervised Clustering

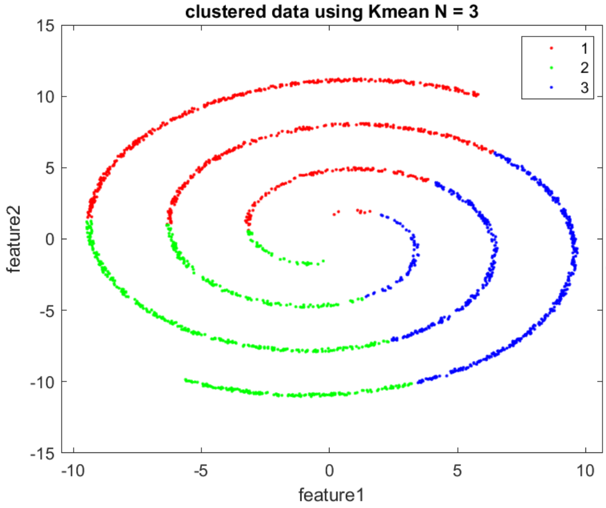
A. K means and Spectral Clustering

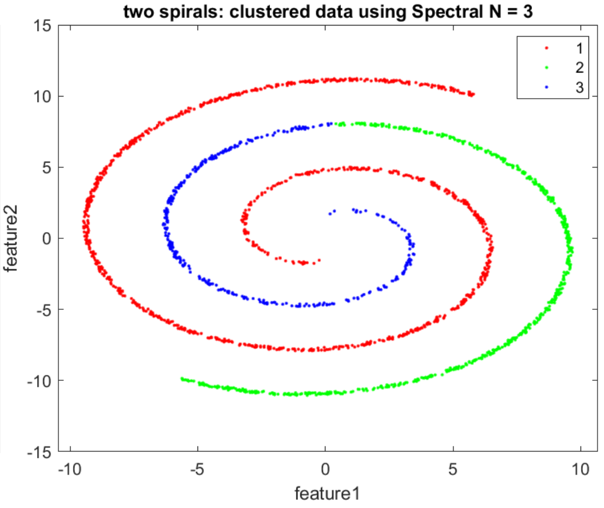
Dataset 1: Two Spirals

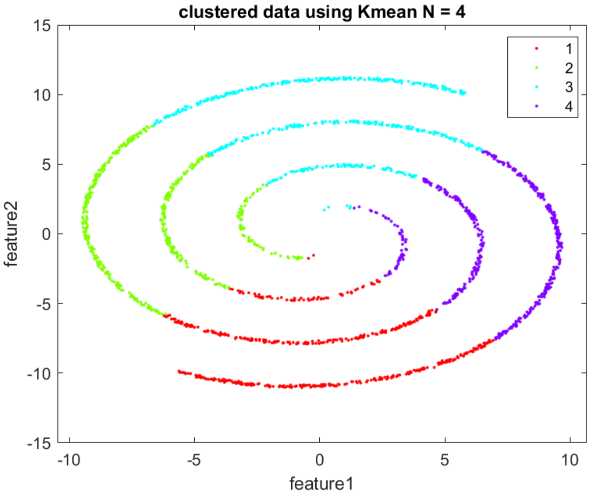


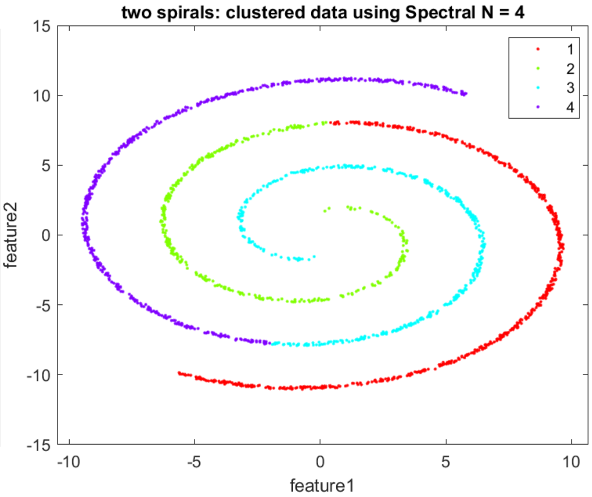




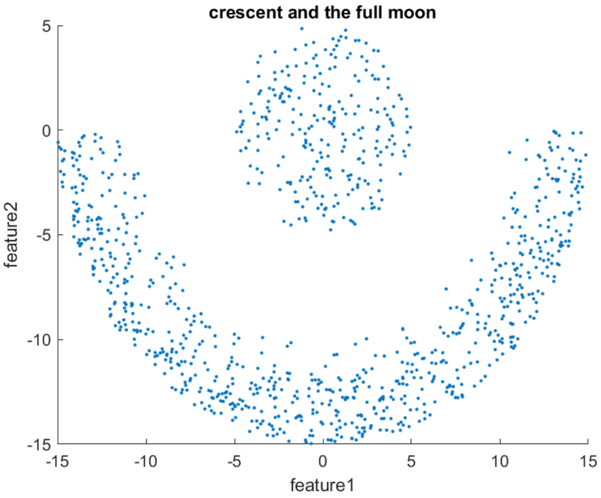


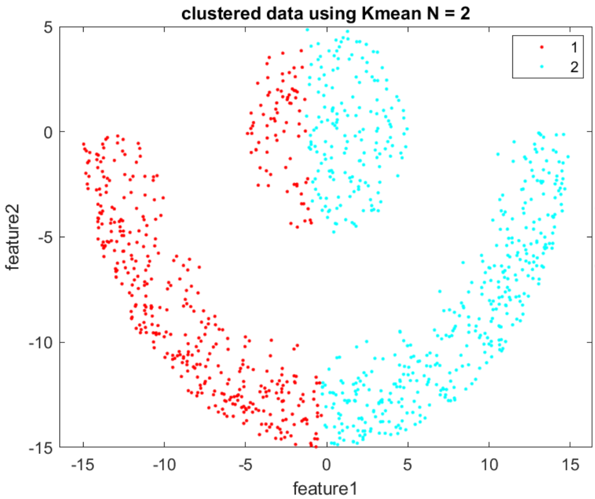


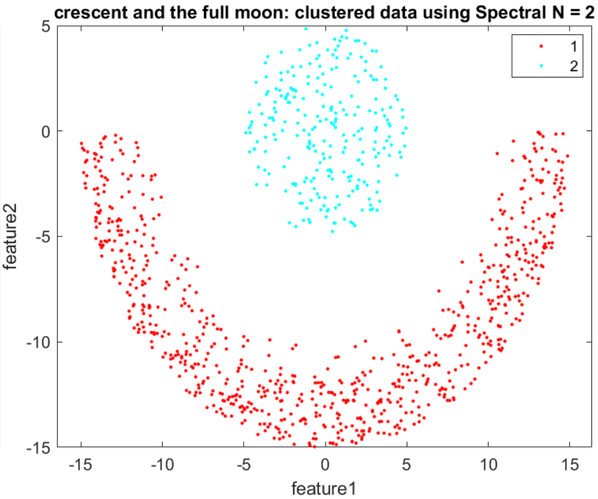


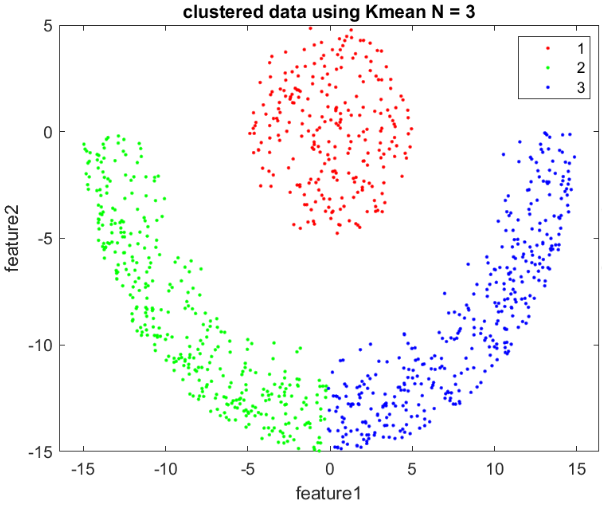


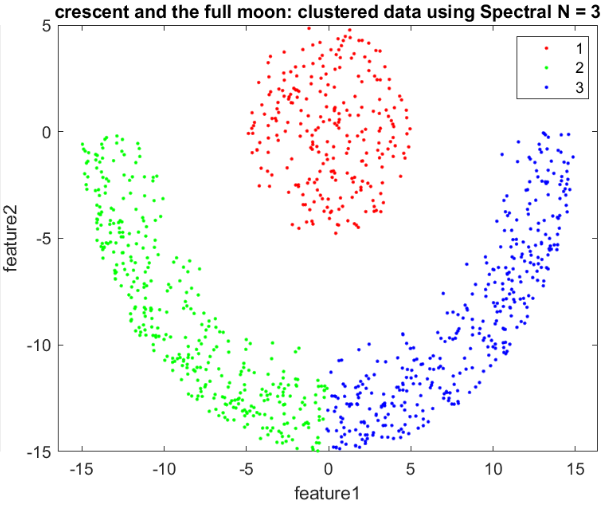
Dataset 2: Crescent and the full moon

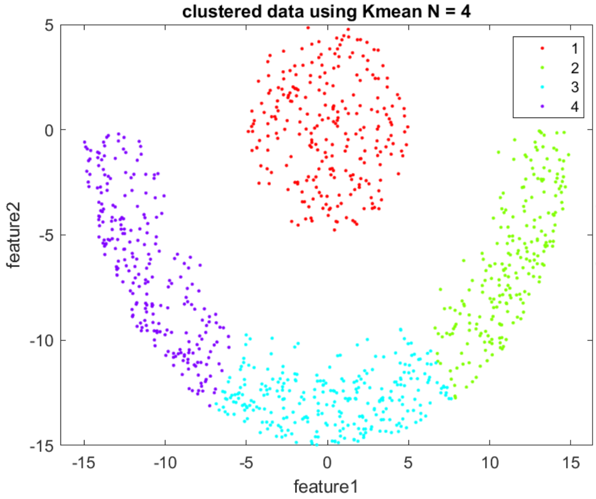


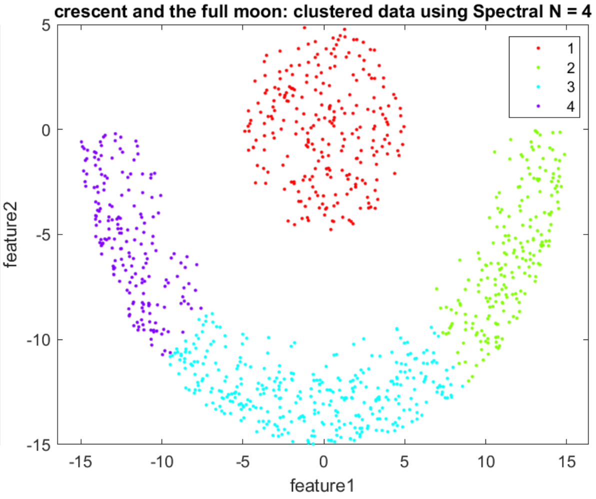

clustered data using Kmean N = 2


crescent and the full moon: clustered data using Spectral N = 2

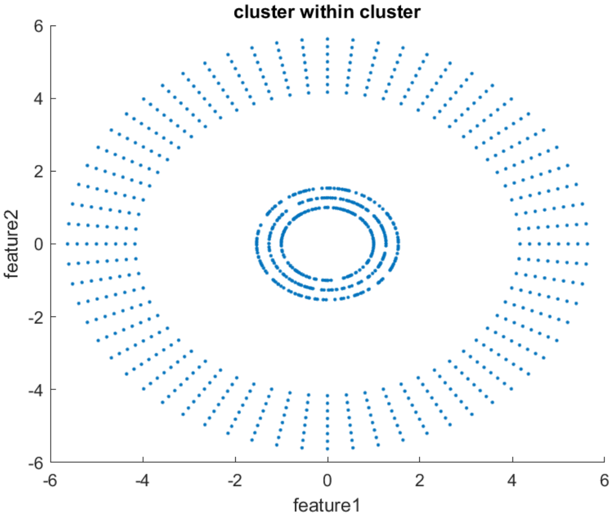

clustered data using Kmean N = 3

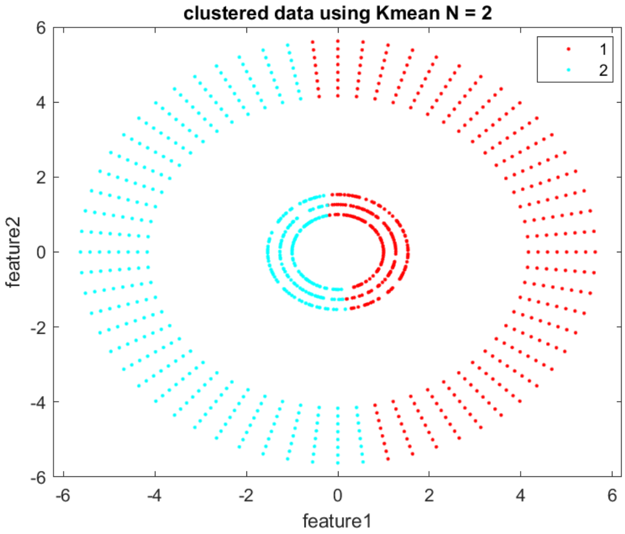

crescent and the full moon: clustered data using Spectral N = 3

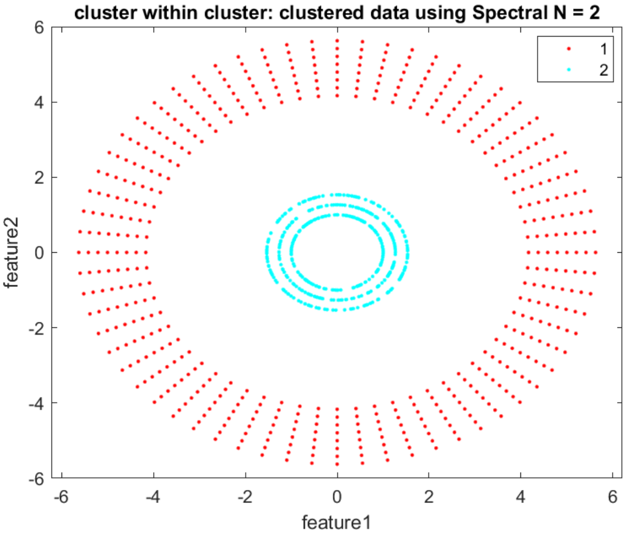

clustered data using Kmean N = 4

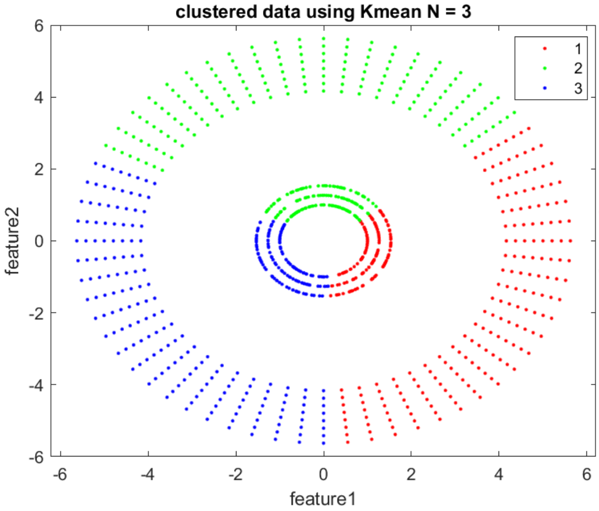

crescent and the full moon: clustered data using Spectral N = 4

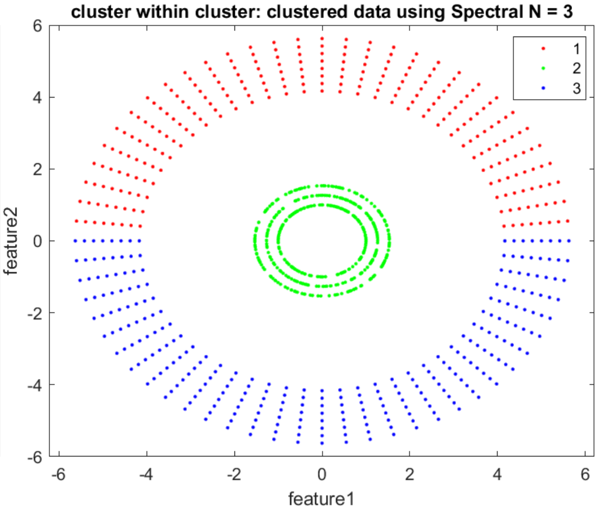
Dataset 3: Cluster within cluster

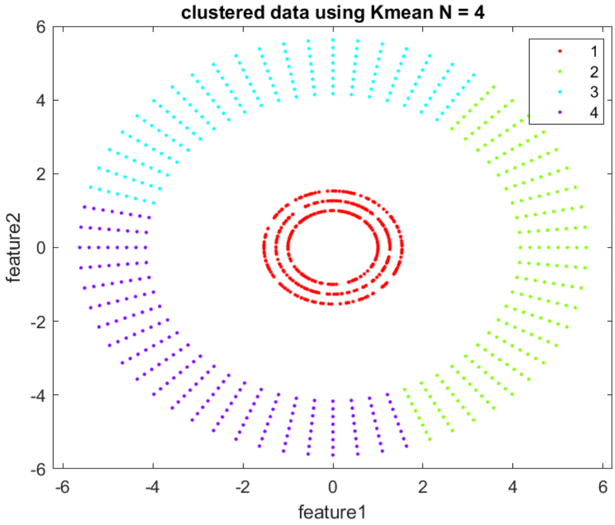


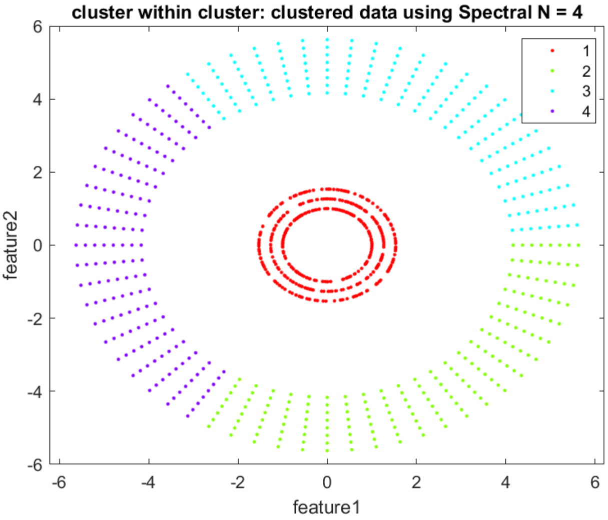












Discussion:

K-means is ideal for discovering globular clusters that all members of each cluster are near each other (in the Euclidean sense). As we can see from these data, each cluster is not gathered together. Therefore, the clustering drawn from K-means is not ideal here. It does not reach the most optimized clusters (not convex). The Spectral method is a graph-clustering technique, where we don't cluster data points directly in their native data space but instead using a similarity matrix that defines the similarity between every pair of data. Data points in spectral clustering are mapped into a connected graph and clusters are found by partitioning this graph into subgraphs (clusters) based on its spectral decomposition and similarity threshold level.

B. K means and Spectral Clustering Accuracy on Supervised Data Set

K means:

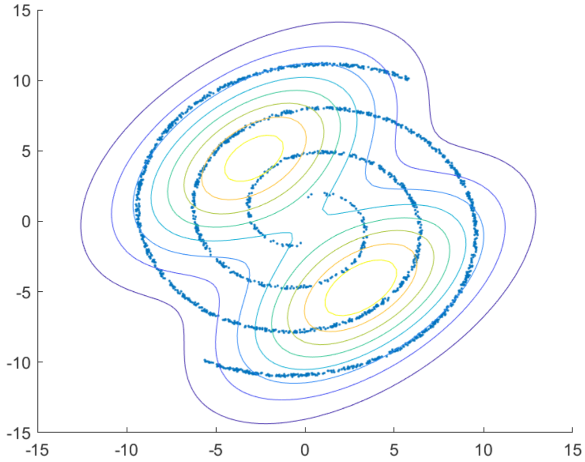
Error Rate = 19.75%

Spectral Clustering:

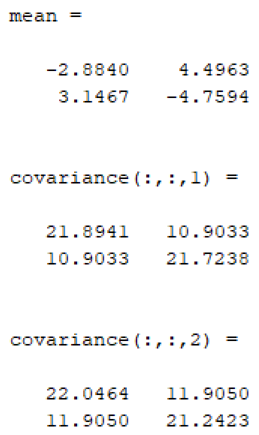
Error Rate = 25.25%

C. Gaussian Mixture Model

Dataset 1: Two Spirals

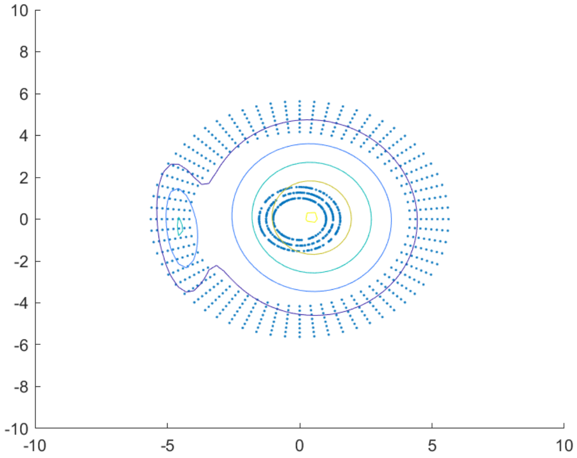


Each row of the mean matrix represents the mean of one Gaussian Distribution

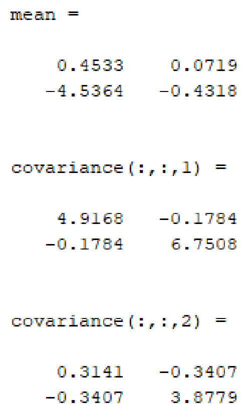


```
mean =

   -2.8840    4.4963
    3.1467   -4.7594


covariance(:,:,1) =

   21.8941   10.9033
   10.9033   21.7238


covariance(:,:,2) =

   22.0464   11.9050
   11.9050   21.2423
```

Dataset 2: Crescent and the full moon

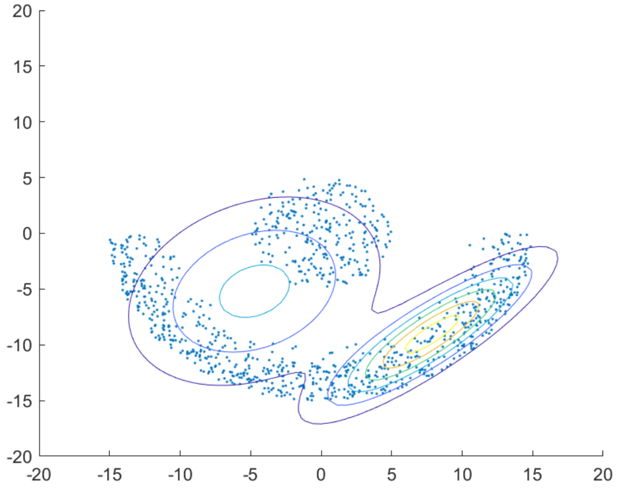


Each row of the mean matrix represents the mean of one Gaussian Distribution

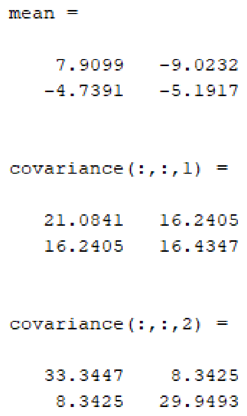


```
mean =

    0.4533    0.0719
   -4.5364   -0.4318


covariance(:,:,1) =

    4.9168   -0.1784
   -0.1784    6.7508


covariance(:,:,2) =

    0.3141   -0.3407
   -0.3407    3.8779
```

Dataset 3: Cluster within cluster



Each row of the mean matrix represents the mean of one Gaussian Distribution



```
mean =

    7.9099   -9.0232
   -4.7391   -5.1917


covariance(:,:,1) =

   21.0841   16.2405
   16.2405   16.4347


covariance(:,:,2) =

   33.3447    8.3425
    8.3425   29.9493
```