EM & Spectral Clustering Algorithms

By

Charles Silkin

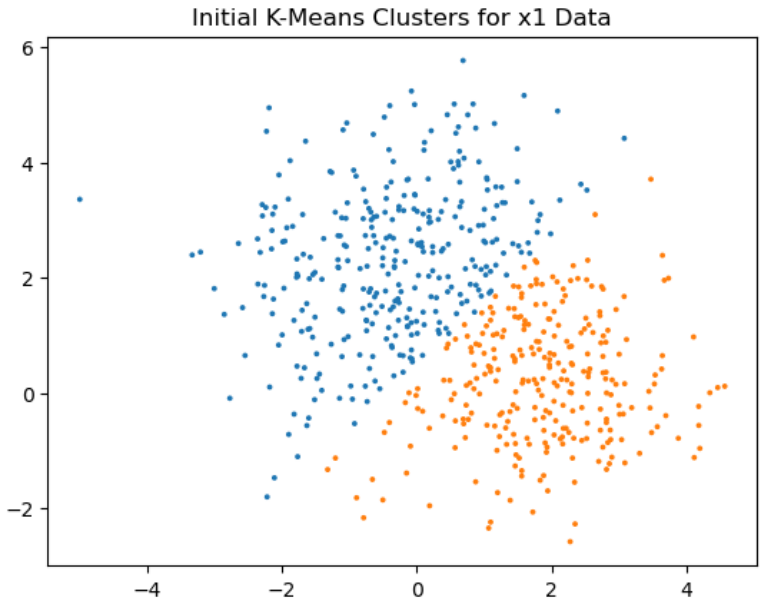
Contents

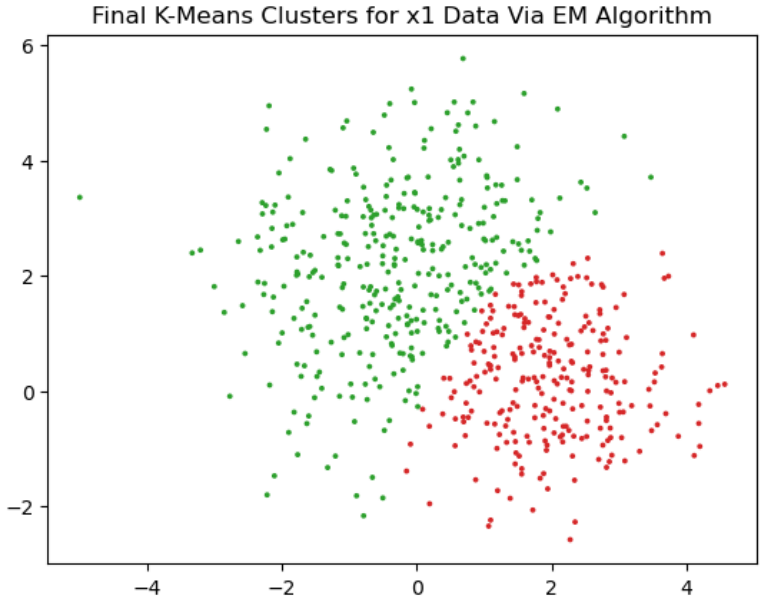
PROJECT INSTRUCTIONS

1. Implement the EM algorithm for clustering with multivariate Gaussian models. Initialize the algorithm with the K-means result (for example Matlab provides a `kmeans` function). Show the plot of the initial clusters in different colors. Show the plot the final clusters in different colors. For each question display two results (initial & final plot) obtained with two different random seeds.
    a. The dataset `x1` (2 clusters)
    b. The dataset `x2` (2 clusters)
    c. The dataset `pts` (3 clusters)
    d. The dataset `pts` (10 clusters)
2. Now implement the spectral clustering algorithm and repeat the questions a)-c) from problem 1. Be sure to tune the affinity matrix parameter σ to obtain a reasonable result. Use the same spectral clustering method for all three questions

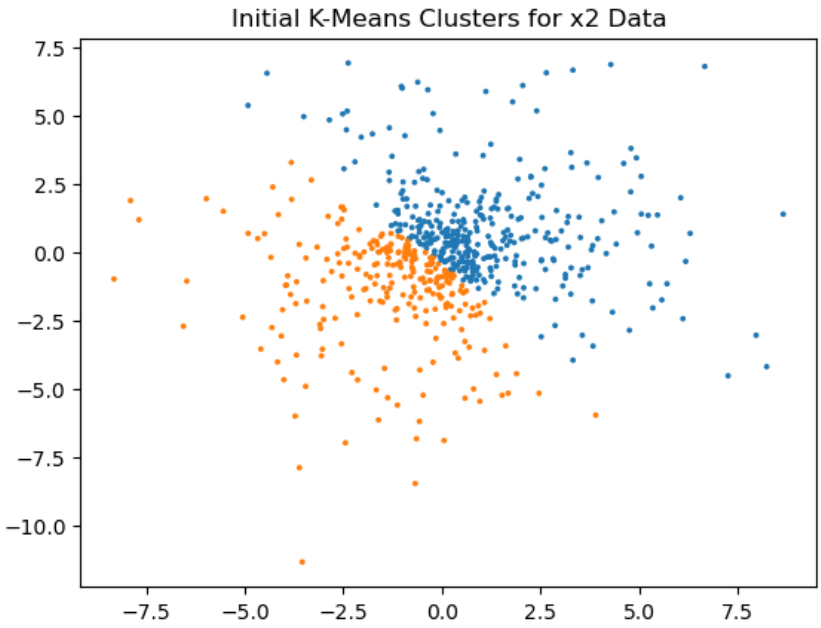
PROBLEM 1 (EM ALGORITHM)

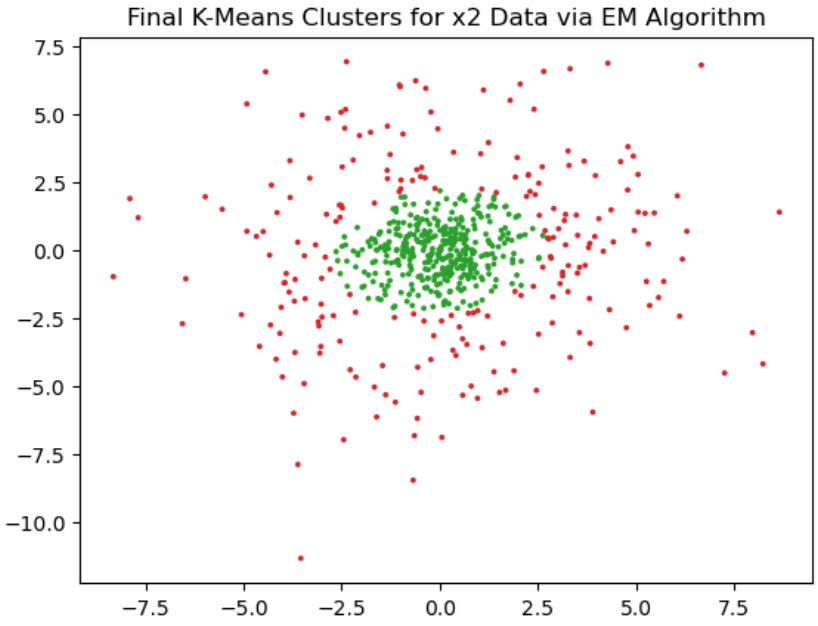
PART A



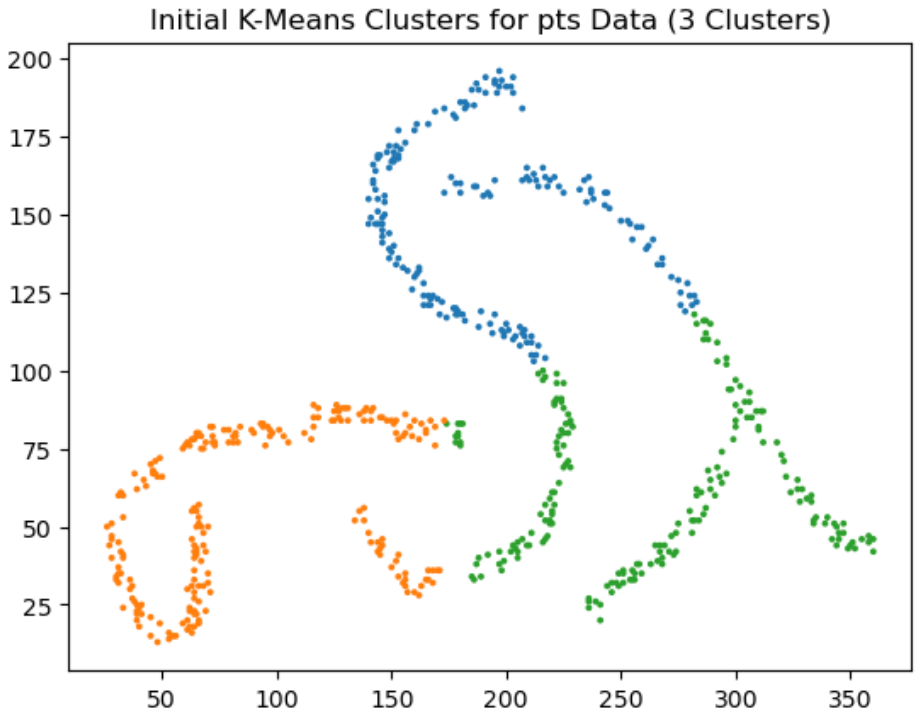


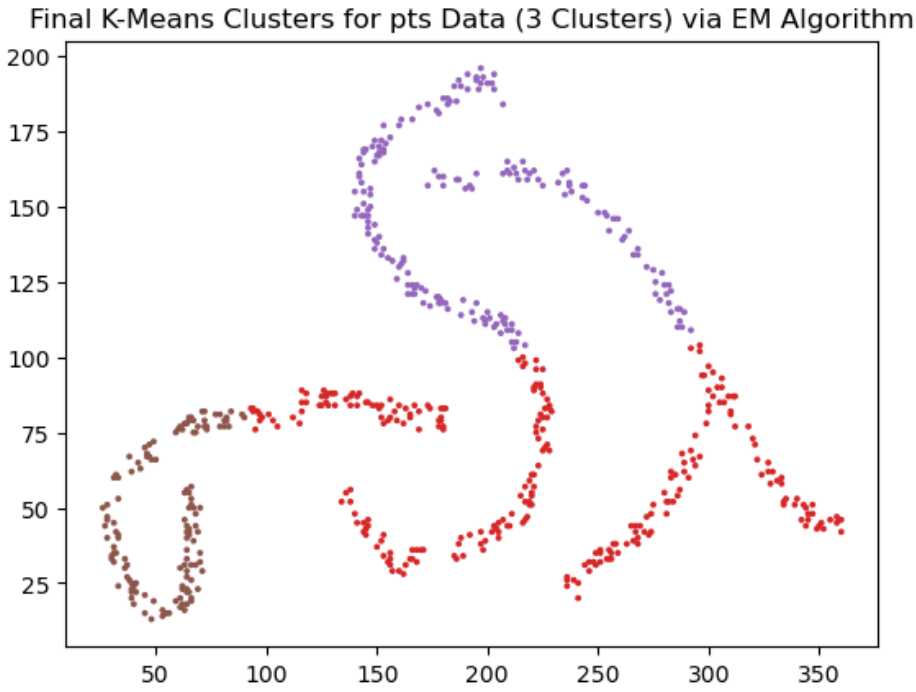
PART B


Initial K-Means Clusters for x2 Data
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

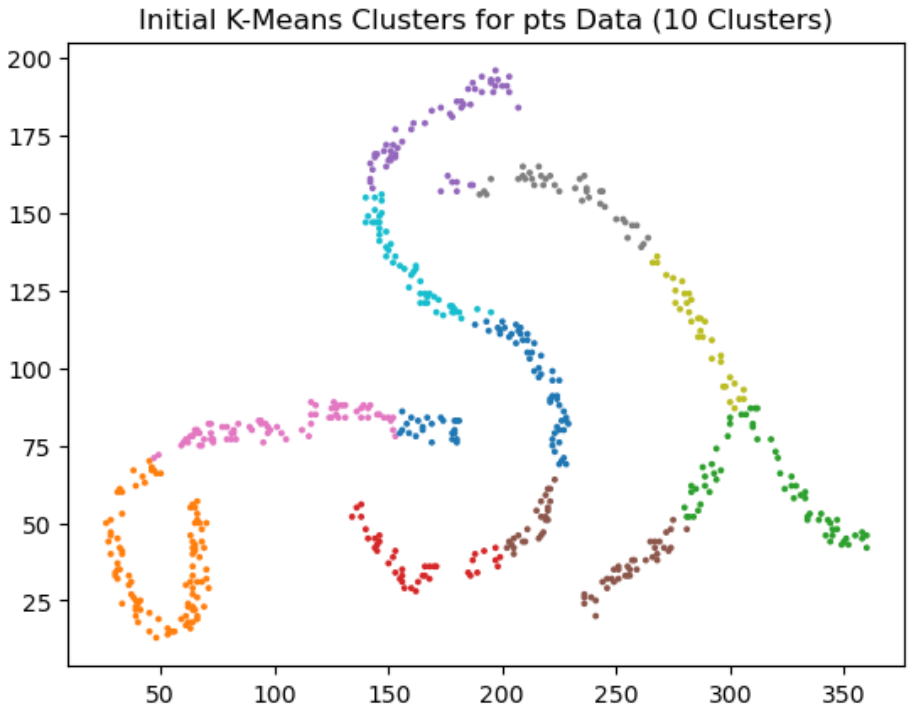

Final K-Means Clusters for x2 Data via EM Algorithm
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5

PART C


Initial K-Means Clusters for pts Data (3 Clusters)
200
175
150
125
100
75
50
25
50
100
150
200
250
300
350


Final K-Means Clusters for pts Data (3 Clusters) via EM Algorithm
200
175
150
125
100
75
50
25
50
100
150
200
250
300
350

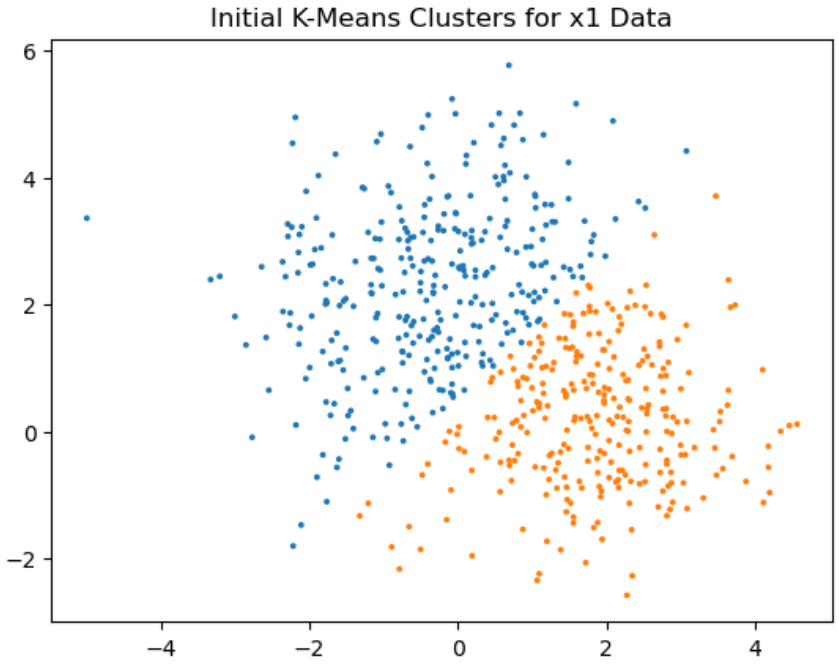
PART D

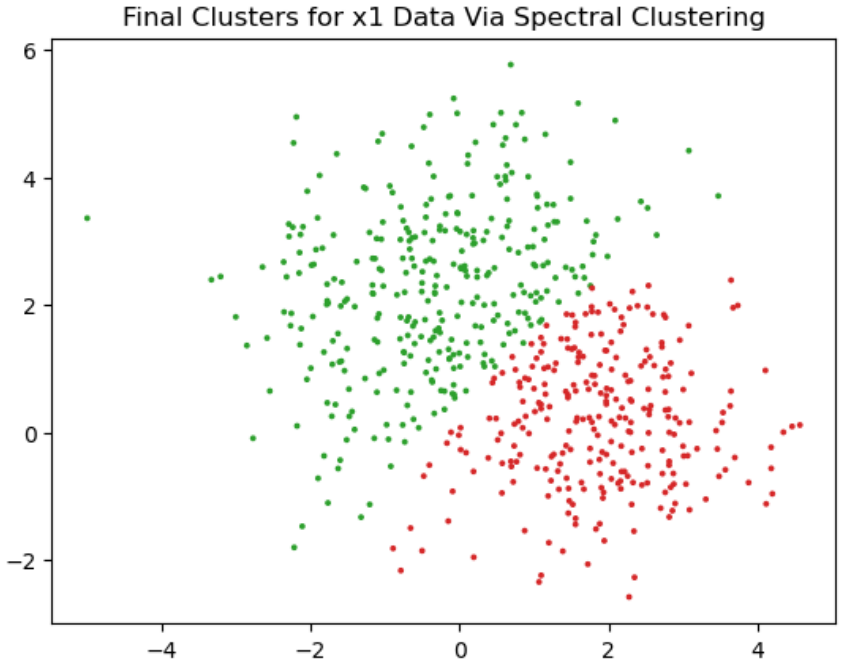

Initial K-Means Clusters for pts Data (10 Clusters)
200
175
150
125
100
75
50
25
50
100
150
200
250
300
350


Final K-Means Clusters for pts Data (10 Clusters) via EM Algorithm
200
175
150
125
100
75
50
25
50
100
150
200
250
300
350

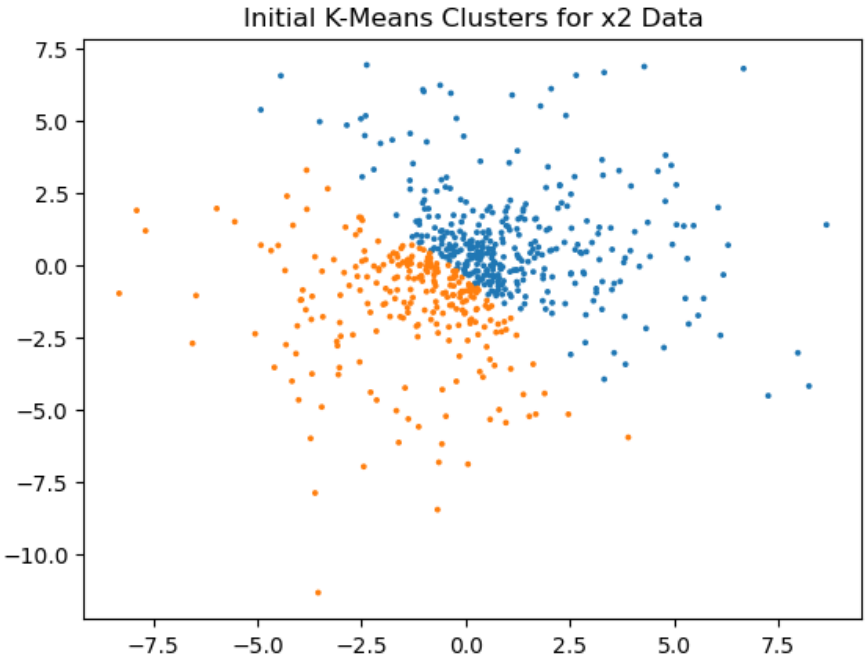
PROBLEM 2 (SPECTRAL CLUSTERING ALGORITHM)

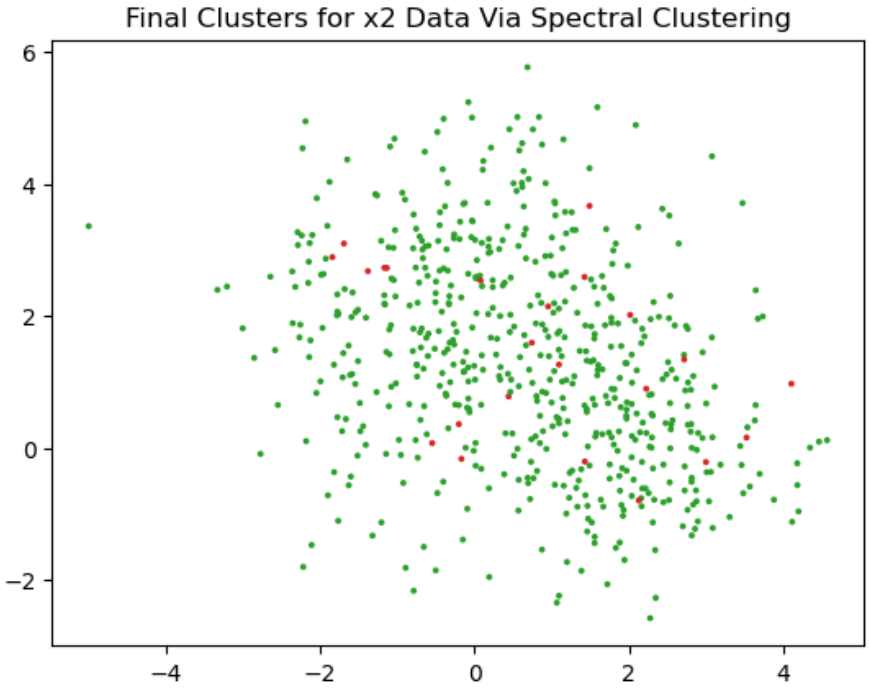
PART A



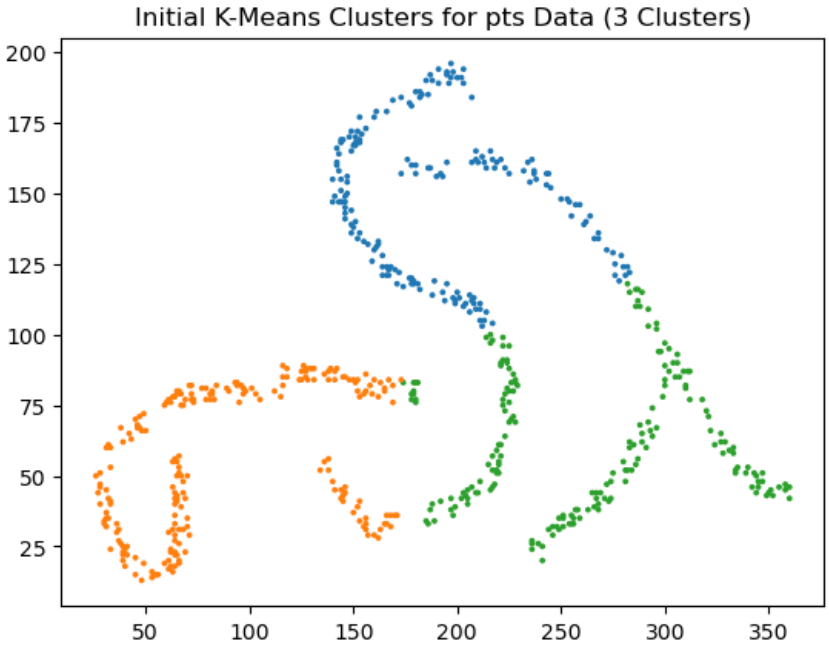


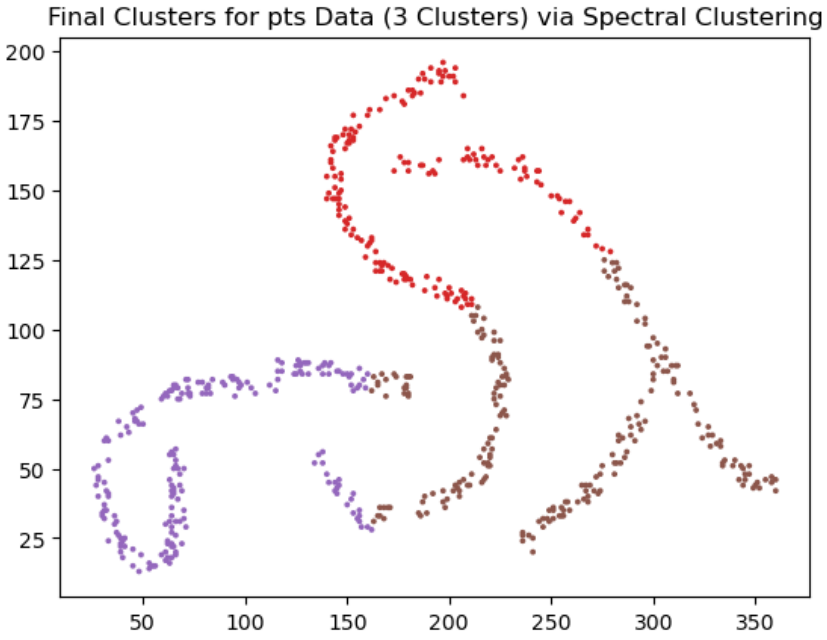
PART B


Initial K-Means Clusters for x2 Data
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


Final Clusters for x2 Data Via Spectral Clustering
6
4
2
0
−2
−4
−2
0
2
4

PART C


Initial K-Means Clusters for pts Data (3 Clusters)
200
175
150
125
100
75
50
25
50
100
150
200
250
300
350


Final Clusters for pts Data (3 Clusters) via Spectral Clustering
200
175
150
125
100
75
50
25
50
100
150
200
250
300
350

REFERENCES

1. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html
2. https://scikit-learn.org/stable/modules/generated/sklearn.mixture.GaussianMixture.html
3. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.SpectralClustering.html