K-Means Clustering

By

Charles Silkin

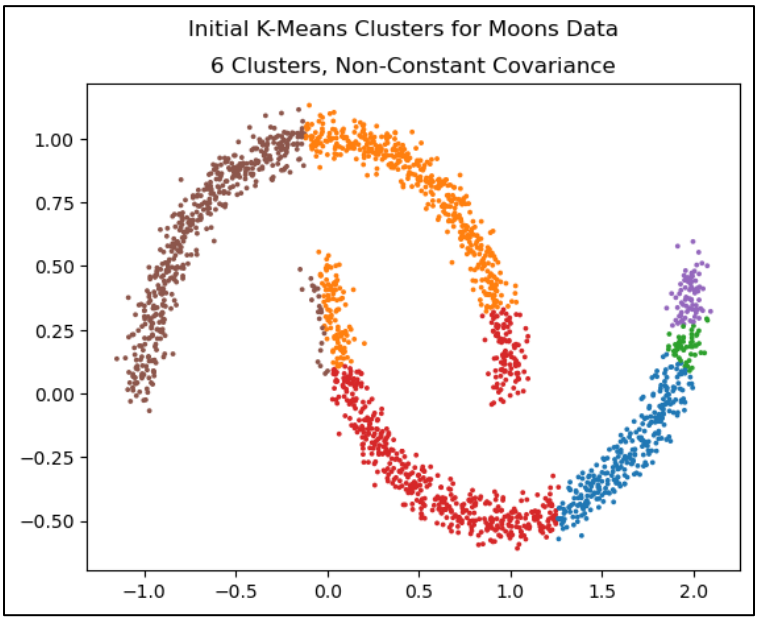
Contents

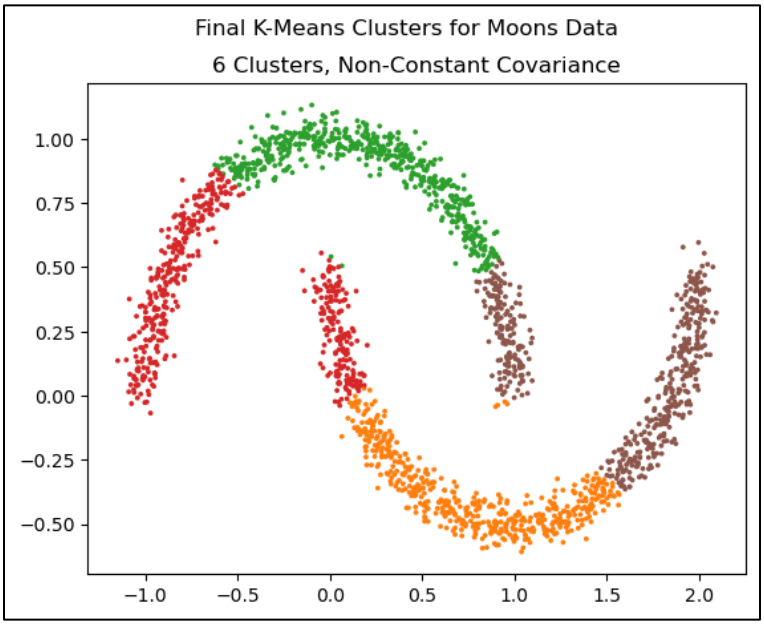
PROJECT INSTRUCTIONS

Implement the k-means algorithm for clustering with multivariate Gaussian models discussed in the LearningGM section. Initialize the algorithm with random observations as the k means and identity covariance matrix. 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 `moons` (6 clusters)
b) The dataset `moons` with standard k-means (with identity covariance matrices) (6 clusters)
c) The dataset `pts` (10 clusters)
d) The dataset `pts` (16 clusters)

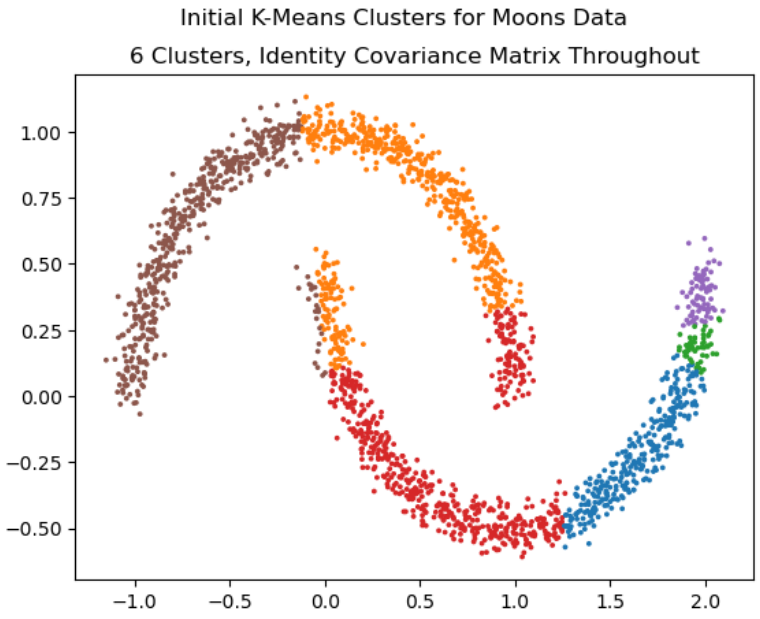
PART A

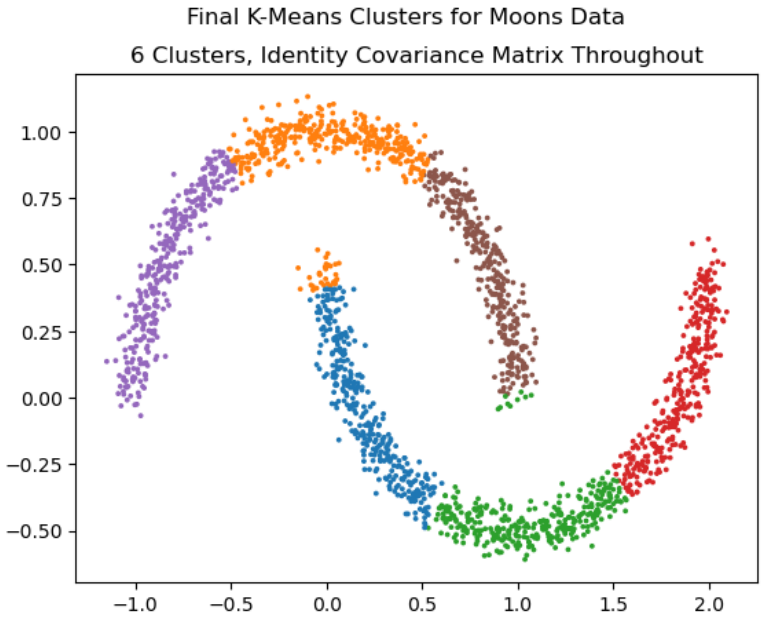




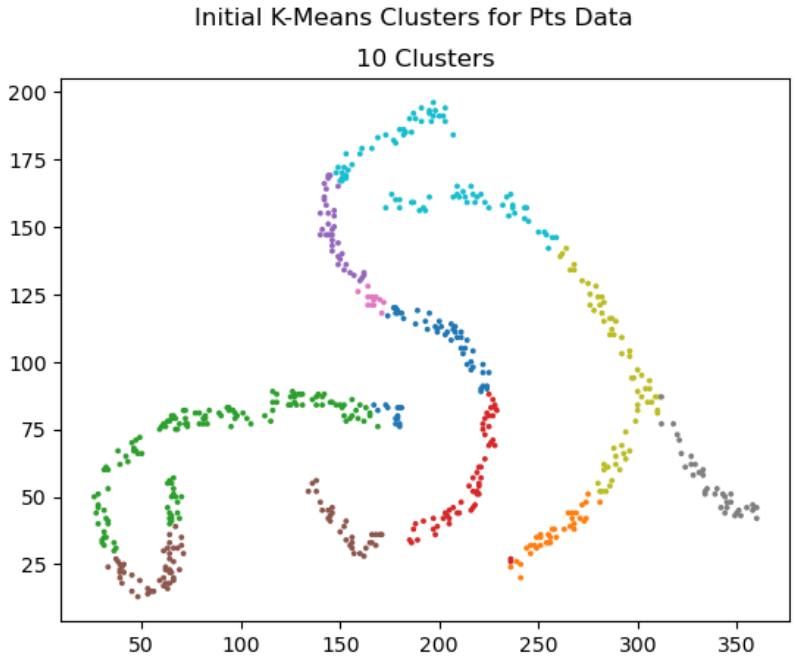
NOTE: The final K-means cluster should have 6 clusters, but the algorithm for the random seed I chose converged with only 4 cluster assignments.

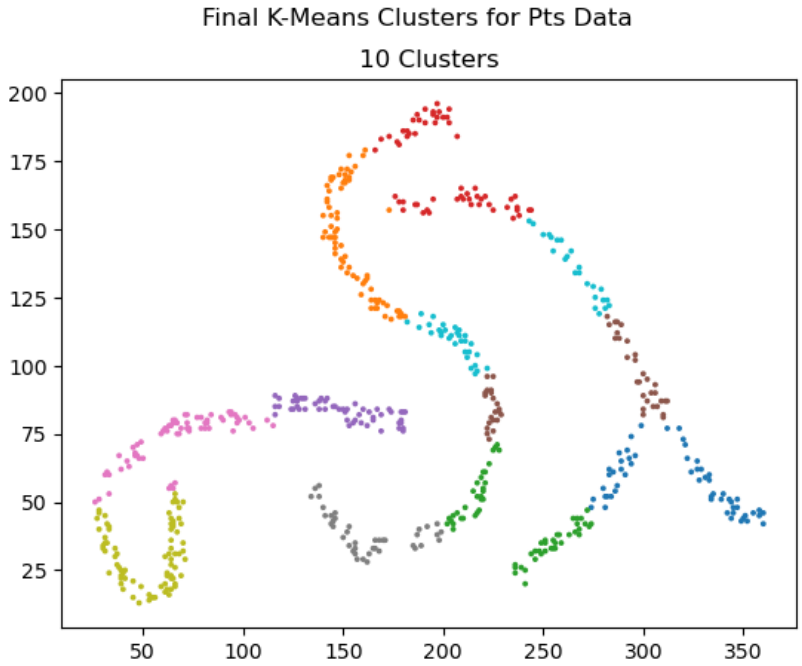
PART B


Initial K-Means Clusters for Moons Data
6 Clusters, Identity Covariance Matrix Throughout
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

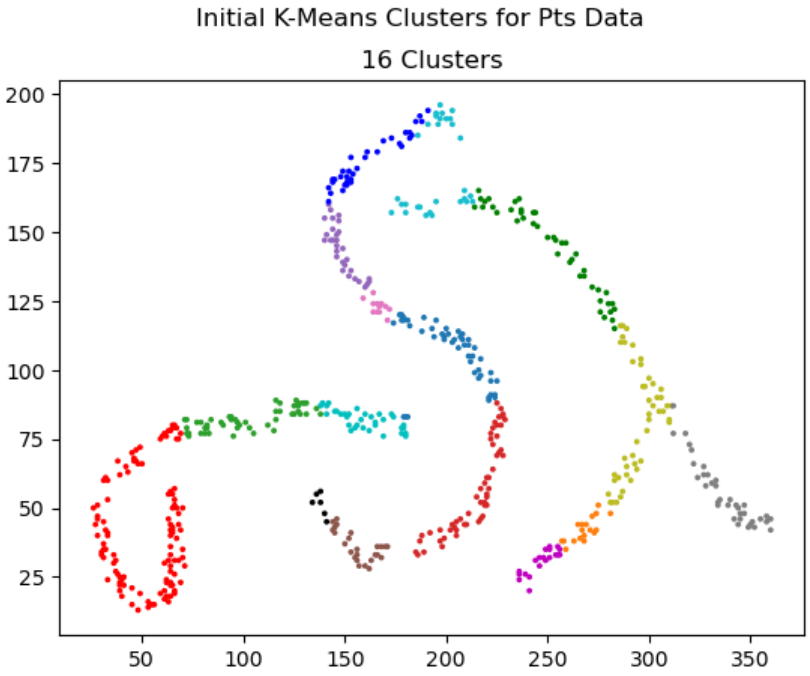

Final K-Means Clusters for Moons Data
6 Clusters, Identity Covariance Matrix Throughout
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0

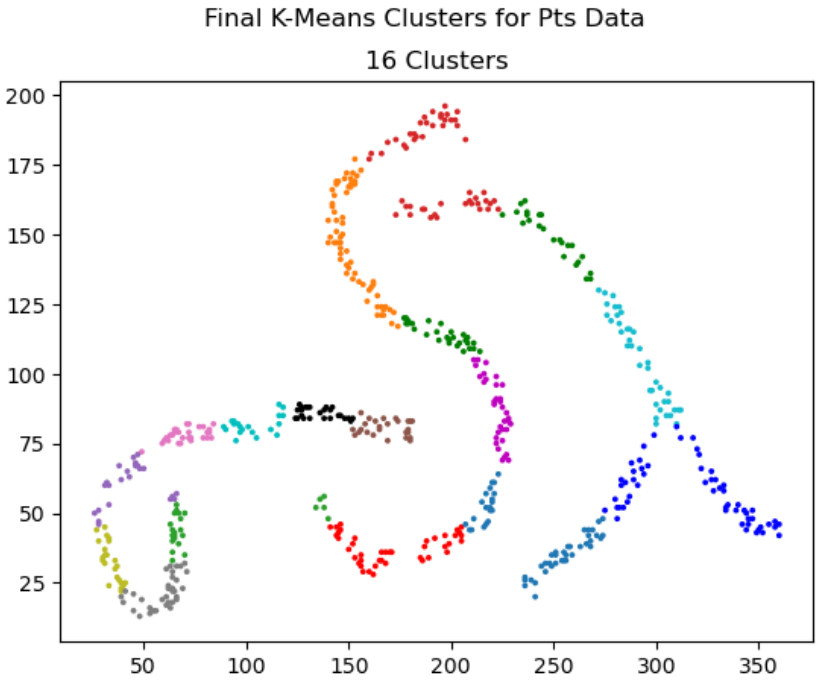
PART C


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
200
175
150
125
100
75
50
25
50
100
150
200
250
300
350

PART D


Initial K-Means Clusters for Pts Data
16 Clusters
200
175
150
125
100
75
50
25
50
100
150
200
250
300
350


Final K-Means Clusters for Pts Data
16 Clusters
200
175
150
125
100
75
50
25
50
100
150
200
250
300
350

REFERENCES

1. https://www.kdnuggets.com/2020/06/centroid-initialization-k-means-clustering.html
2. https://pytorch.org/tutorials/index.html
3. https://sites.google.com/view/paztronomer/blog/basic/python-colors
4. https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html
5. https://towardsdatascience.com/clustering-out-of-the-black-box-5e8285220717