Project 2 - K-Means Algorithm Strategy 2

Project 2 - K-Means Algorithm Part 1

February 14, 2024

1 Project Summary

Strategy 2 of the k-means algorithm involves initializing the centroids such that they have the maximum distance from each other based off of the input data samples. The idea behind this method is that having the center vectors as far away from each other will eventually bring the center vectors closer to the middle of the clusters. Like in updating the centers in the original k-Means algorithm, the distance for the centroids is calculated by using the Euclidean distance method. After initializing the centroids, the k-means algorithm proceeds as usual. This method of initializing the centroids to have the maximum distance from each other is called k-Means++ algorithm.

2 Strategy 2

The centroids were first initialized by checking the furthest distance from the first centroid. After this, I generated the rest of the centroids by finding the maximum averages of all of the distances from the centroids and the sample data points. After generating the centroids, I used the same k-Means algorithm from part 1.

The loss values from this are shown as below. There appears to be small yet noticeable improvements using this strategy. Like in part 1, cluster 5 is when the loss values took a noticeable downturn.

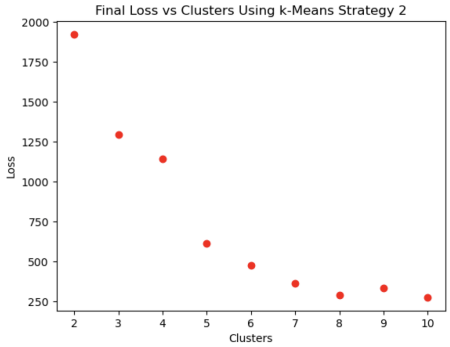

Figure 1: Enter Caption

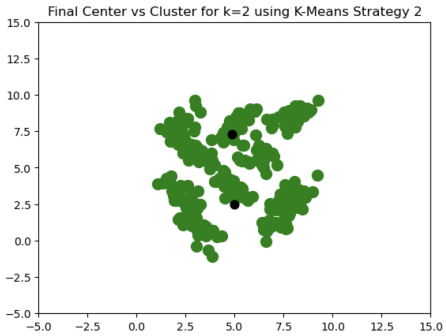

Figure 2: Enter Caption

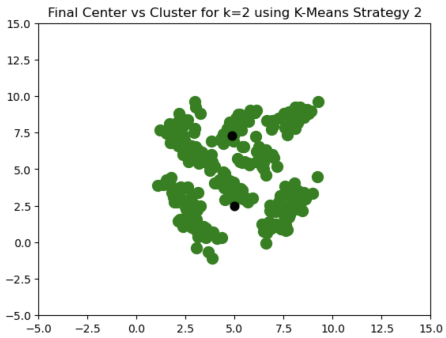

Figure 3: Enter Caption