Lab 4: Clustering

Student Yadder Aceituno
Student ID i6382150

Assignment 1

(a) Data Generation

Using the `make_blobs` function from sklearn, 4 clusters are generated with 300 points total, 2 features (X_1 and X_2) and a standard deviation of 0.6. The clusters are then visualized in a scatter plot in figure 1.

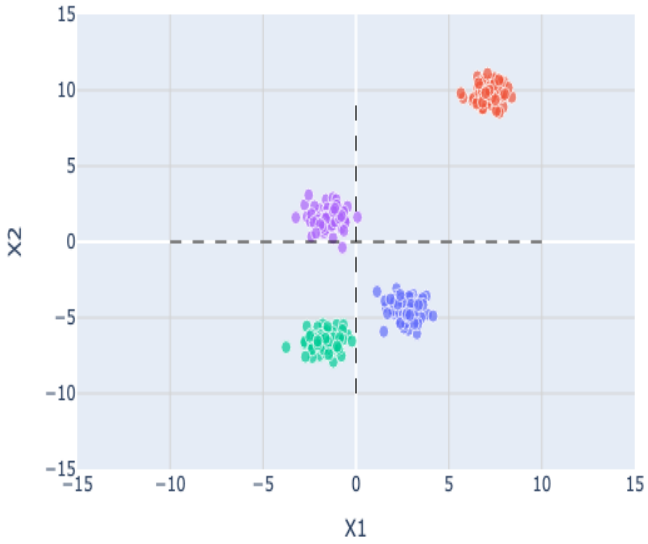

Figure 1: Clusters Generated

(b) K-means clustering with K = [1,2,3,4,5,6,7,8,9,10]

With the data generated in the previous step, the K-means clustering algorithm is applied with K = [1,2,3,4,5,6,7,8,9,10]. The results are visualized in Figure 2.

The figure shows when K $<$ 4, the centers of the clusters are not well defined with the data points generated. They are in the middle of all the clusters.

When K = 4, the centers provided by K-means are well positioned because the data points generated were created with 4 clusters.

When K $>$ 4, the centers provided by K-means are well positioned but it starts to create multiple centers for the same cluster.

Assignment 2

Assignment 3

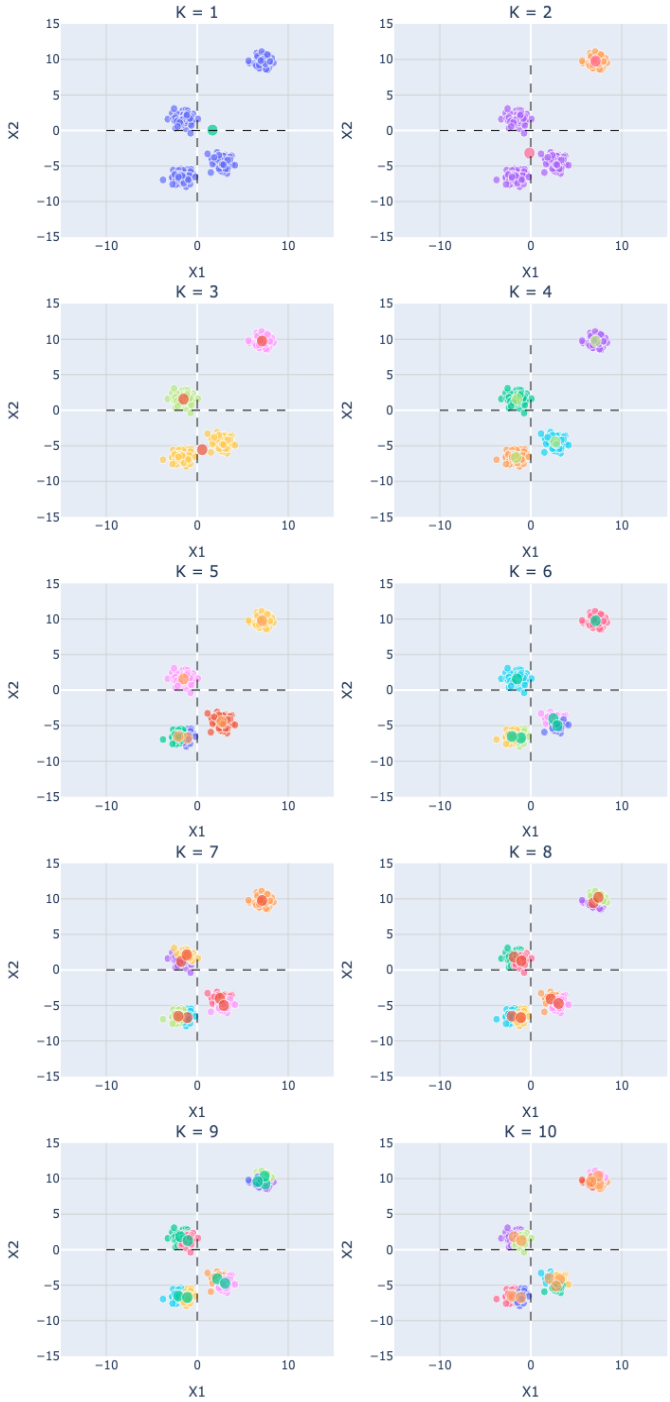

Figure 2: K-means Clustering