

E0270 : Assignment 2

Dhaval Chavda (20973)

April 12, 2023

1 K-Means Clustering

K-means clustering is a popular unsupervised machine learning algorithm used for clustering similar data points into K distinct groups or clusters. The goal is to group similar data points together and separate dissimilar points into different groups.

1.1 Methodology

1. First randomly select K data points from the dataset that serves as the initial centroids of K clusters.
2. Find the distance for each data points to K centroids, based on the smallest distance data points assigned to the cluster.
3. Update the centroids of all clusters based on the mean of the newly assigned data points.
4. Repeat this process until the centroids of clusters remain the same or we reach the maximum number of iterations.

1.2 Clustered Image



Figure 1: Original Image



Figure 2: No. of Clusters = 2



Figure 3: No. of Clusters = 5



Figure 4: No. of Clusters = 10



Figure 5: No. of Clusters = 20



Figure 6: No. of Clusters = 50

1.3 Mean Squared Error

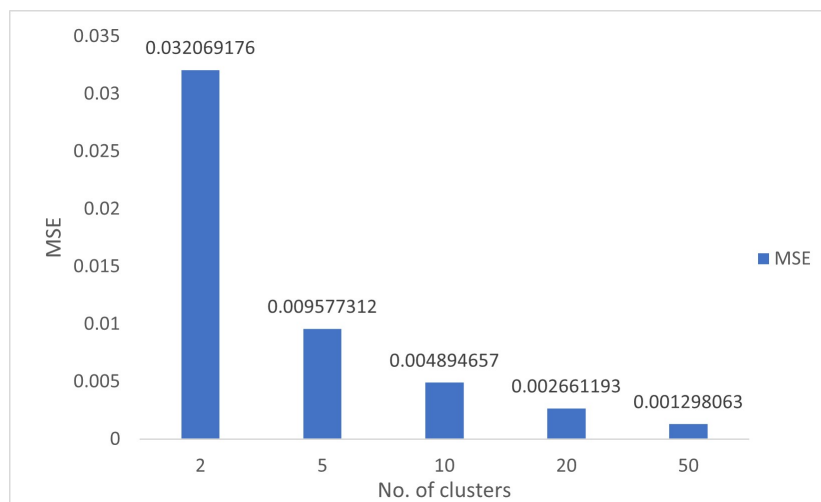


Figure 7: No. of Clusters = 50