CS 6375: Machine Learning

Assignment 5

K-Means Clustering for Image Compression

-Anushree Jumade

(axj174430)

An image comprises of several very small intensity values known as [Pixels](https://en.wikipedia.org/wiki/Pixel). In a colored image, each pixel has three channels: Red, Green and Blue intensity value for each pixel. This tends to make the size of a colored digital image very huge (3 values per pixel). So, these images can be stored as compressed images with lesser number of bits for intensity values and hence, lesser memory for storage.

K-means clustering is the optimization technique to find the ‘k’ clusters or groups in the given set of data points. The data points are clustered together on the basis of some kind of similarity. Initially, it starts with the random initialization of the ‘k’ clusters and then on the basis of some similarity (like euclidean distance metric), it aims to minimize the distance from every data point to the cluster center in each clusters. There are mainly two iterative steps in the algorithm: