Fundamentals of Data Science

Semester B 20-21

Tutorial 8

1. Consider the following set of one-dimensional points: {6, 12, 18, 24, 30, 42, 48}.

   a. For each of the following sets of initial centroids (i) {18,45}, (ii) {15,40}, create two clusters by assigning each point to the nearest centroid, and then calculate the sum squared error for each set of two clusters after updating the centroids.

   b. Do both sets of centroids represent stable solutions, i.e., if the K-means algorithm is applied to this set of points using the given centroids as the starting centroids, would there be any change in the clusters generated ?

2. Suppose that we need to group the following eight points into three clusters:
   A:(2,10), B:(2,5), C(8,4), D(5,8), E(7,5), F(6,4), G(1,2), H(4,9)

   The distance function is Euclidean distance. Suppose initially we assign A, D and G as the prototype of the first, second and third cluster respectively. Use the K-means algorithm to find the three clusters and their respective centroids after the first iteration.