Lab Exercise – 12: K-Means Clustering

NOTE:

* Prepare a PDF document and name the file as "Lab12_RegisterNo.pdf".
* PDF file should consist Question No, Code, and Result for each Question.
* File Should be headed with your Register number, Slot number, Lab Exercise number.

1. a. Develop k-Means Clustering algorithm to apply clustering on the following data objects referred by (x, y) pair: (k =3)

A1(2, 10), A2(2, 5), A3(8, 4), A4(5, 8), A5(7, 5), A6(6, 4), A7(1, 2), A8(4, 9)

Use Euclidian distance metric to determine closest centroid.

Algorithm for Reference:

Algorithm: k-means. The k-means algorithm for partitioning, where each cluster's center is represented by the mean value of the objects in the cluster.

Input:

- k: the number of clusters,
- D: a data set containing n objects.

Output: A set of k clusters.

Method:

(1) arbitrarily choose k objects from D as the initial cluster centers;
(2) **repeat**
(3)   (re)assign each object to the cluster to which the object is the most similar, based on the mean value of the objects in the cluster;
(4)   update the cluster means, that is, calculate the mean value of the objects for each cluster;
(5) **until** no change;

Note: Do not use any predefined packages.

b.
- Load IRIS data set (IRIS.csv)

- Remove Class Label column from IRIS data set

- Apply developed k-Means clustering in Question 1 on the unlabelled IRIS data set with k = 3.

- Plot the clusters using a scatter plot in such a manner so that each user should identify each cluster easily.