PART B

(PART B : TO BE COMPLETED BY STUDENTS)

***(Students must submit the soft copy as per following segments within two hours of the practical. The soft copy must be uploaded on the Blackboard or emailed to the concerned lab in charge faculties at the end of the practical in case the there is no Black board access available)***

|  |  |
| --- | --- |
| Roll No: B032 | Name: NAMAN GARG |
| Class: CS B | Batch: B2 |
| Date of Experiment: 14/08/20 | Date of Submission: |
| Grade: |  |

**B.1 Clustering Code written by student:**

***(Paste your clustering code completed during the 2 hours of practical in the lab here)***

setwd("C:/Users/Naman/Desktop/DM/Experiment 7")

D = c(2,4,10,12,3,20,30,11,25)

k = 2

c = 0

m1 = 3

m2 = 4

K11 = c()

K22 = c()

while (TRUE){

K1 = c()

K2 = c()

for (i in D) {

if(abs(i-m1) <= abs(i-m2)){

K1 = append(K1,i)

}else{

K2 = append(K2,i)

}

}

m1 = mean(K1)

m2 = mean(K2)

if(c!=0){

if(K11==K1 & K22==K2){

break

}

}

K11 = K1

K22 = K2

c = c+1

}

K11

K22

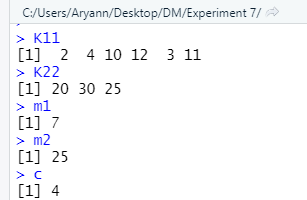
c

**B.2 Input and Output:**

***(Paste your program input and output in following format, If there is error then paste the specific error in the output part. In case of error with due permission of the faculty extension can be given to submit the error free code with output in due course of time. Students will be graded accordingly.)***

**Input Data:** {2,4,10,12,3,20,30,11,25}, k=2

**Output Clusters:** K1 = {2,3,4,10,11,12}, K2 = {20,30,25}, m1=7, m2=25



**B.3 Observations and learning:**

***(Students are expected to comment on the output obtained with clear observations and learning for each task/ sub part assigned)***

K-means clustering is a method of vector quantization that aims to partition n observations into k

clusters in which each observation belongs to the cluster with the nearest mean (cluster centers or

cluster centroid), serving as a prototype of the cluster. This results in a partitioning of the data

space. In the task solved above, we took a one-dimensional dataset and partitioned it into two

clusters. Initially two values from the dataset were assumed to be the means and clusters were

formed accordingly from the data by comparing Euclidean distance (absolute distance in this

case since the dataset is one-dimensional) of means with each value. The means were updated to

the actual mean of the two clusters formed. This entire process was repeated for several iterations

till there was no further change in the composition of the clusters.

**B.3 Conclusion:**

*(****Students must write the conclusion as per the attainment of individual outcome listed above and learning/observation noted in section B.3)***

After completing this lab, I am able to:

1. Understood the concept of Data Mining by implementing some data mining algorithm.
2. Understood the clustering and types of clustering.
3. Understood how to create clusters by K-means algorithm.

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*