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: 21/08/20 | Date of Submission: |
| Grade : |  |

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

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

setwd("C:/Users/Naman/Desktop/DM/Experiment 6")

x = c(0.4,0.22,0.35,0.26,0.08,0.45)

y = c(0.53,0.38,0.32,0.19,0.41,0.3)

data = data.frame(x, y)

d = dist(data)

h1 = hclust(d, method="single")

plot(as.dendrogram(h1))

h2 = hclust(d, method="complete")

plot(as.dendrogram(h2))

abline(h=0.3,col="red")

abline(h=0.1,col="red")

h3 = hclust(d,method="average")

plot(as.dendrogram(h3))

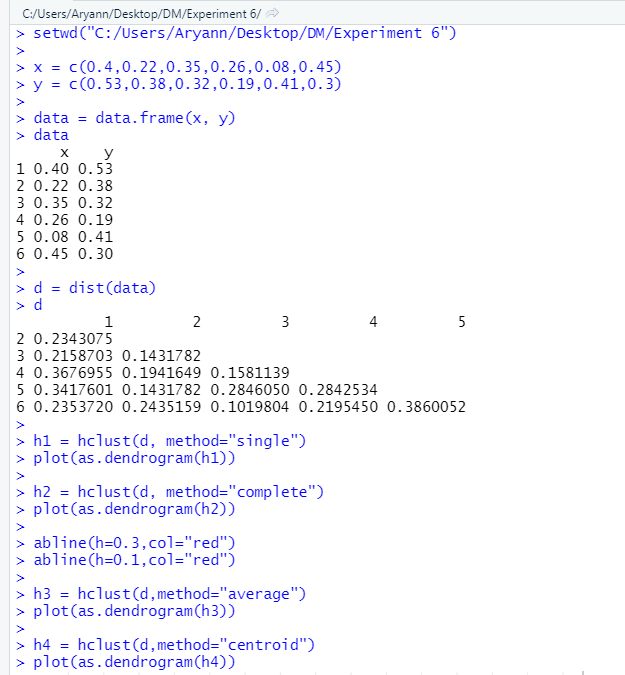
h4 = hclust(d,method="centroid")

plot(as.dendrogram(h4))

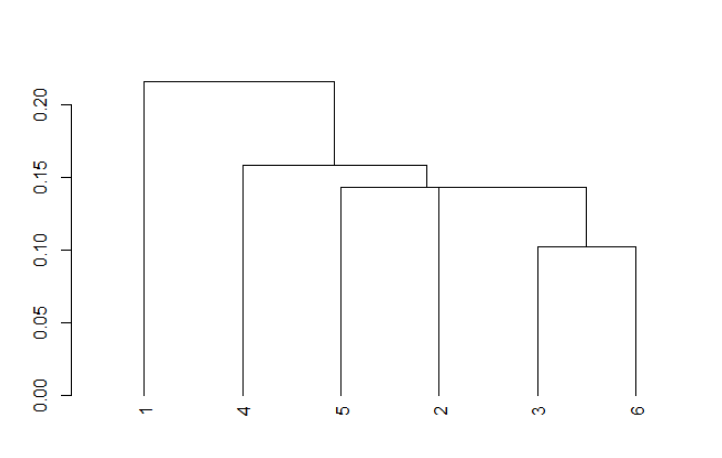
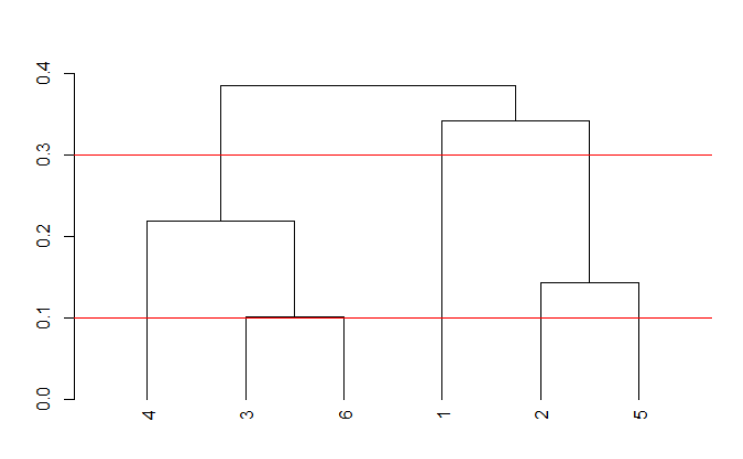
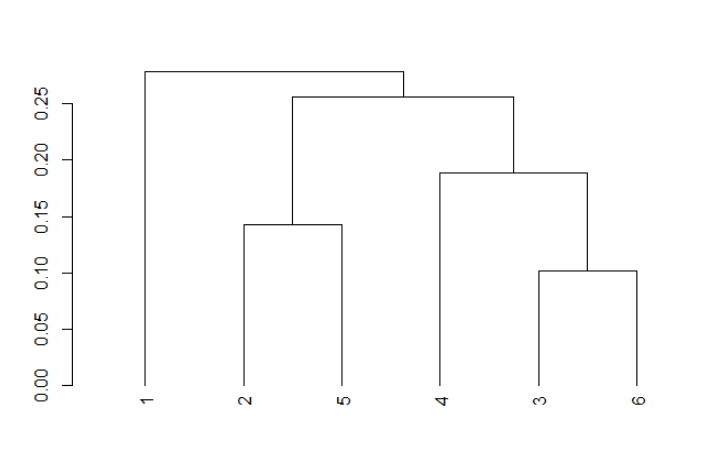
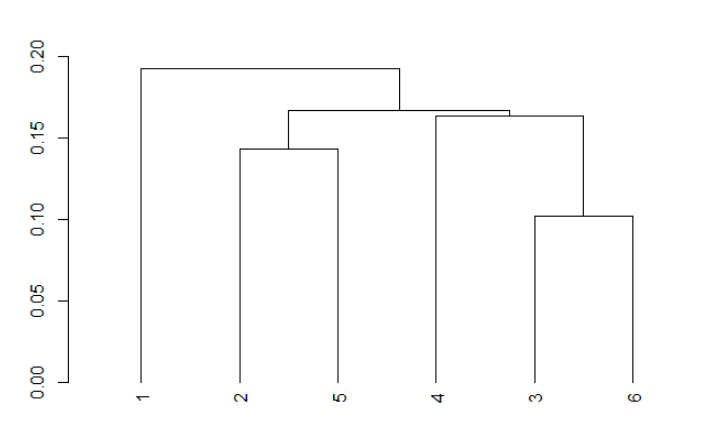
**B.2 Input and Output:**

***(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:**

****

**Output Clusters:**

**B.3 Observations and learning:**

Hierarchical clustering is an alternative approach which builds a hierarchy from the bottom-up, and doesn’t require us to specify the number of clusters beforehand.

The algorithm works as follows:

1. Put each data point in its own cluster.
2. Identify the closest two clusters and combine them into one cluster.
3. Repeat the above step till all the data points are in a single cluster.

Once this is done, it is usually represented by a dendrogram like structure.

There are a few ways to determine how close two clusters are:

* Complete linkage clustering: Find the maximum possible distance between points belonging to two different clusters.
* Single linkage clustering: Find the minimum possible distance between points belonging to two different clusters.
* Mean linkage clustering: Find all possible pairwise distances for points belonging to two different clusters and then calculate the average.
* Centroid linkage clustering: Find the centroid of each cluster and calculate the distance between centroids of two clusters.

All 4 methods have been used and dendrograms have been plotted for them.

**B.4 Conclusion:**

* Understood the concept of Data Mining by implementing some data mining algorithm.
* Understood the various Clustering techniques in Mining.
* Understood Clustering using Agglomerative Algorithm.