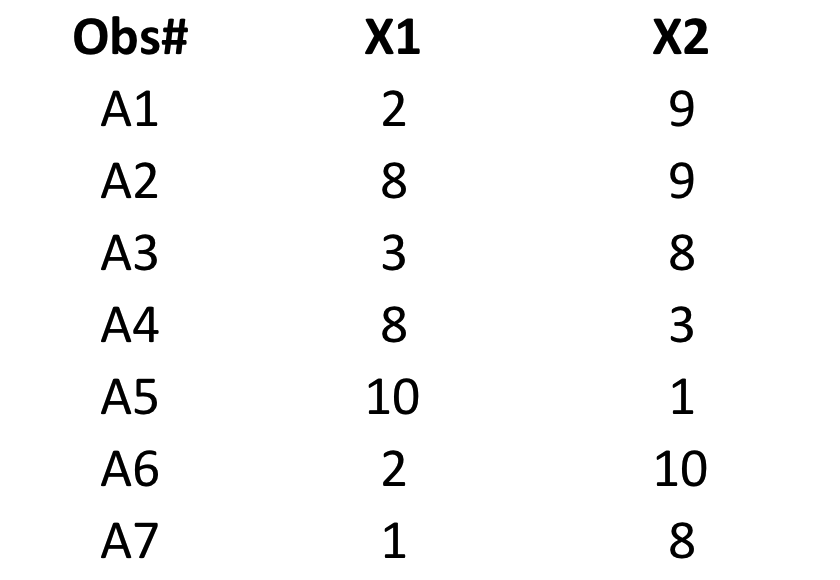
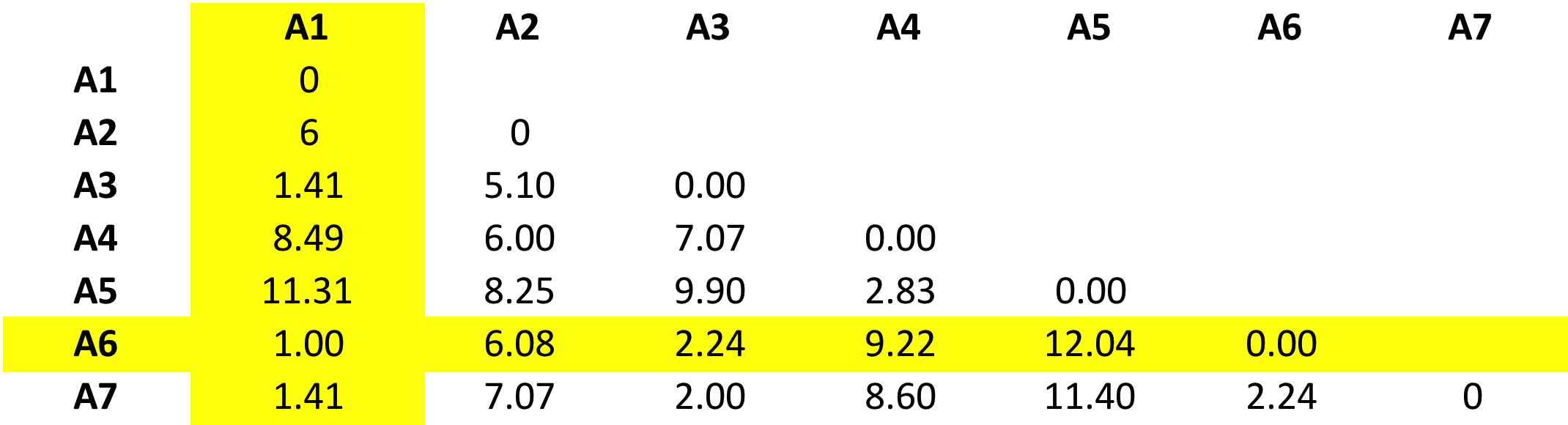
**Question 2: Hierarchical Clustering**



Step 1) Calculation of Euclidean distance (EUD) between pairs is shown in the matrix below:

Table 1:Matrix 1



Grouping together A1 and A6 into one cluster as their separation is minimum.

Step 2) Now (A1 A6) forms a cluster,

Pairwise distance matrix is recomputed as below.

For calculation of a distance of a point from cluster (A1 A6), individual distances from points A1 and A6 are compared and maximum of the value is chosen(complete linkage).

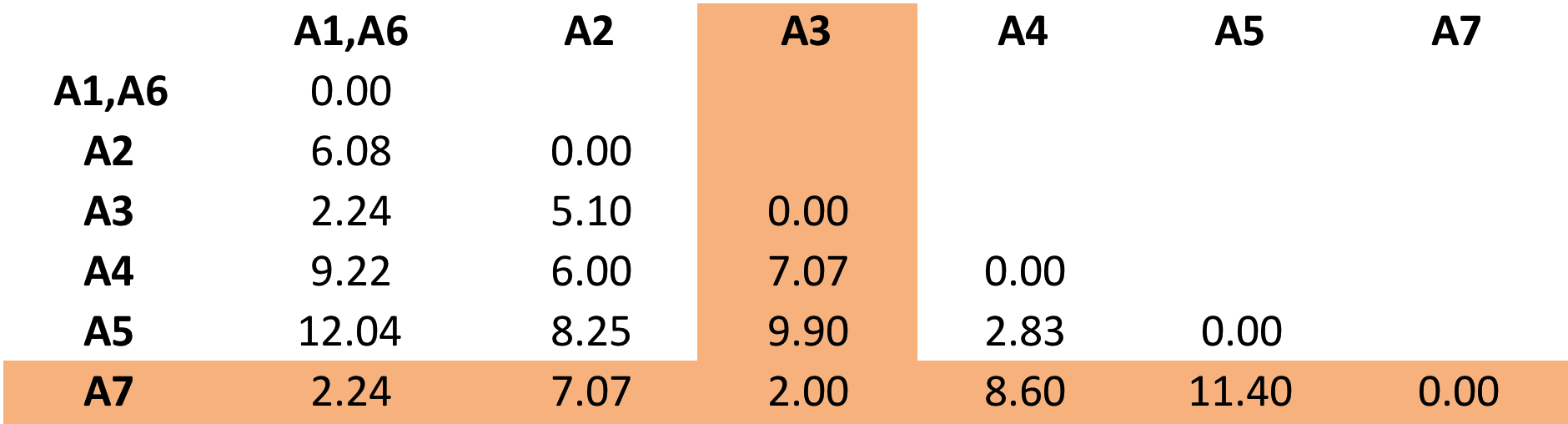
For example: the EUD between A2 and (A1 A6) is determined as follows:

From Table 1: EUD(A1~A2)[[1]](#footnote-1)=6 & EUD(A6~A2)=6.08🡪

For complete linkage, EUD ((A1,A6)~A2)=𝑀𝑎𝑥[6.00,6.08] =6.08

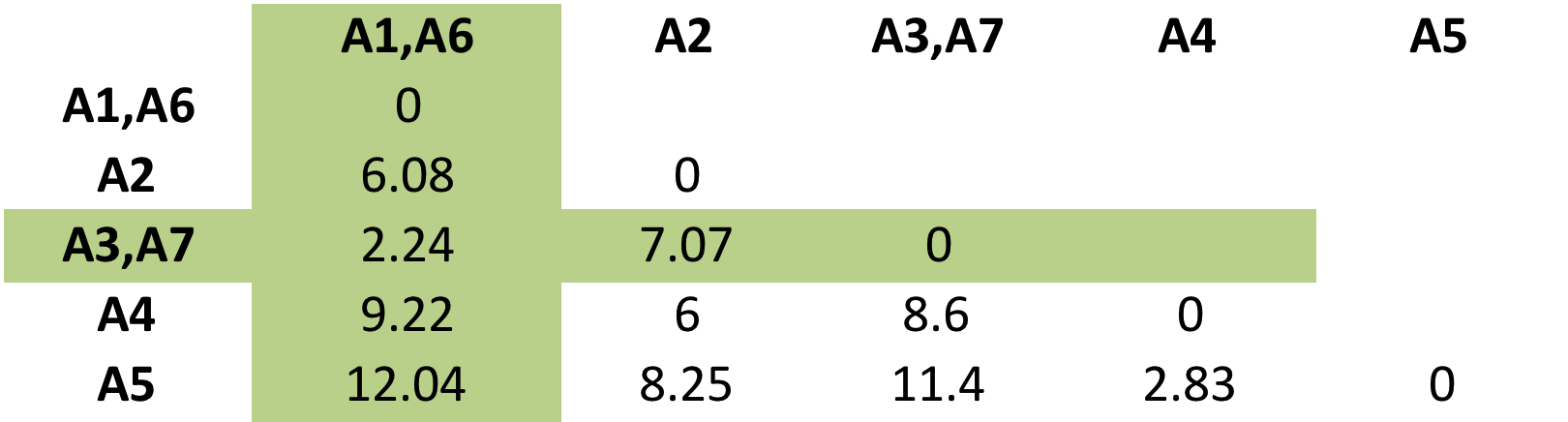
Now based on EUD’s a cluster (A3 A7) is generated as shown in Table 2 below.

Table 2:Matrix 2



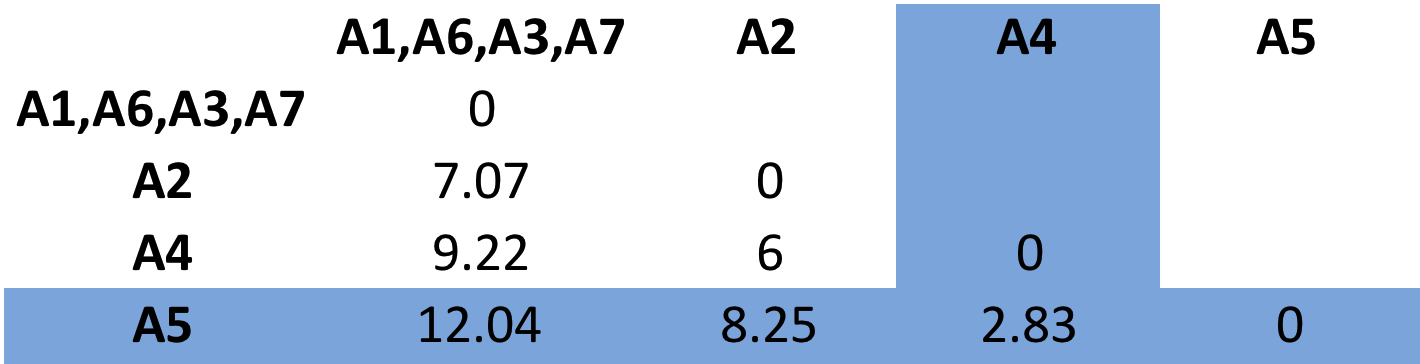
Step 3) Now the clusters (A1 A6) and (A3 A7) are merged as in Table 3.

Table 3:Matrix 3



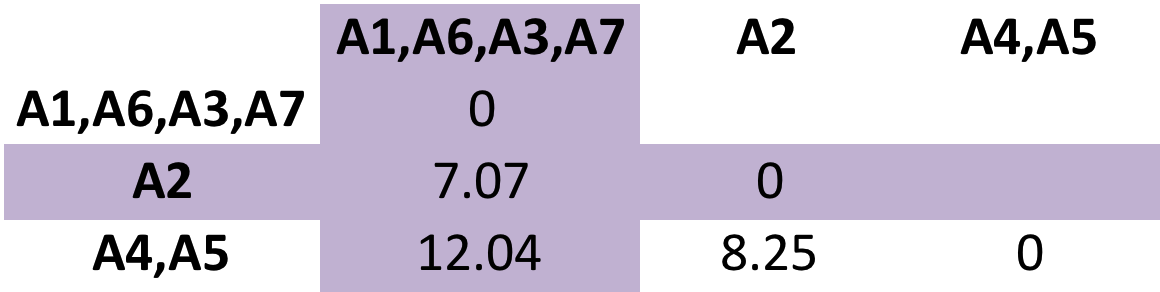
Step 4) Now by re-computing the matrix, another cluster is (A4 A5) is formed based on minimum EUD (Table 4).

Table 4: Matrix 4



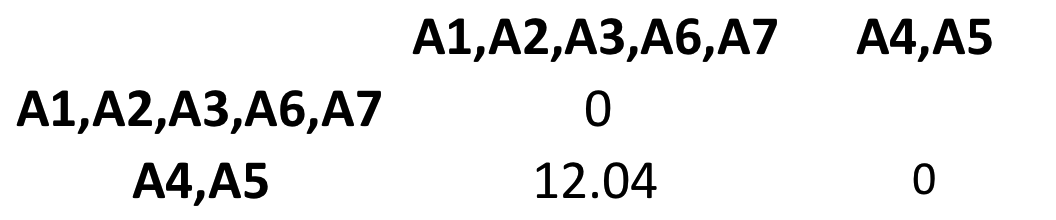
Step 5) A2 is merged with cluster (A1, A6, A3, A7) as per Table 5.

Table 5:Matrix 5

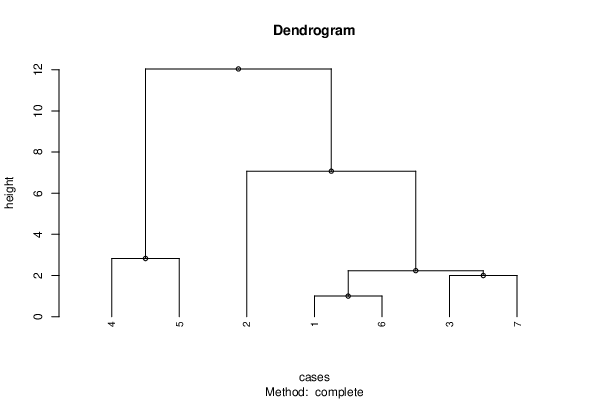


Step 6) All clusters are merged into a tree (Table 6).

Table 6:Matrix 6



**Dendrogram for Question 1**



Cut from here

\*the line shows the cut location for getting exactly two clusters.

Cluster Height (unit)

(A1, A6) 1

(A3, A7) 2

(A1, A6, A3, A7) 2.23

(A4, A5) 2.82

(A2, A1, A6, A3, A7) 7.07

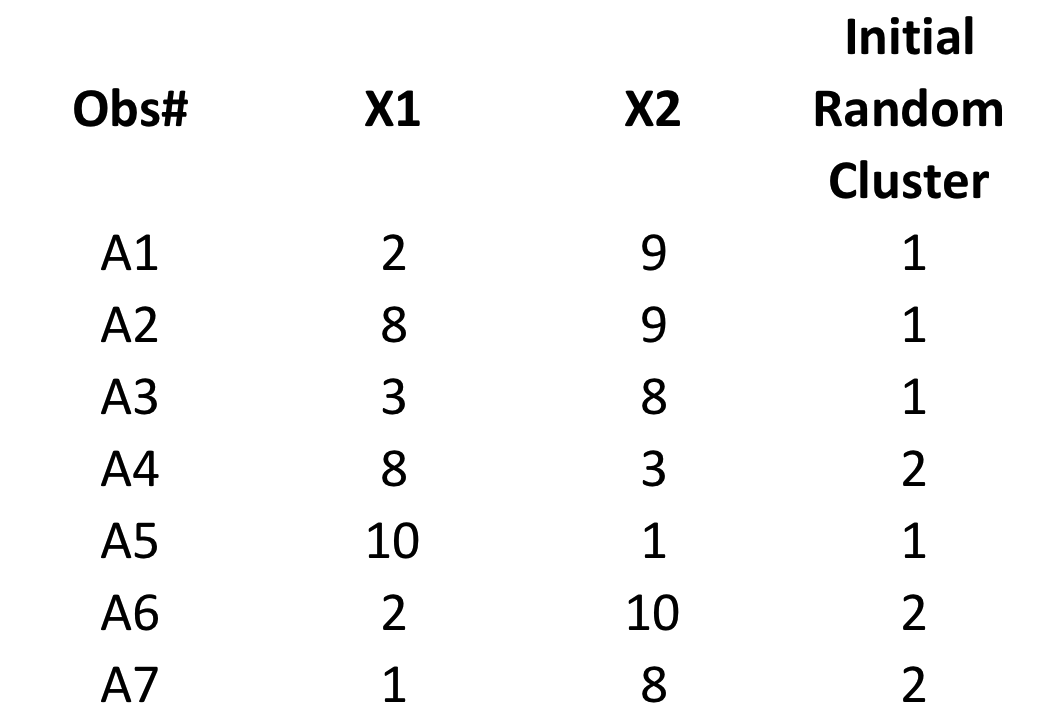
(A1, A2, A3, A4, A5, A6) 12.04

**Question 1: K Means Clustering, k=2**

***ITERATION 1***

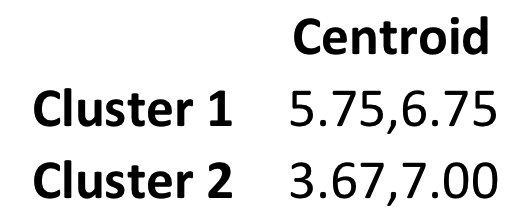
Step 1) Random assignment of clusters to the observations:

Table 7:Random Assignment of Clusters



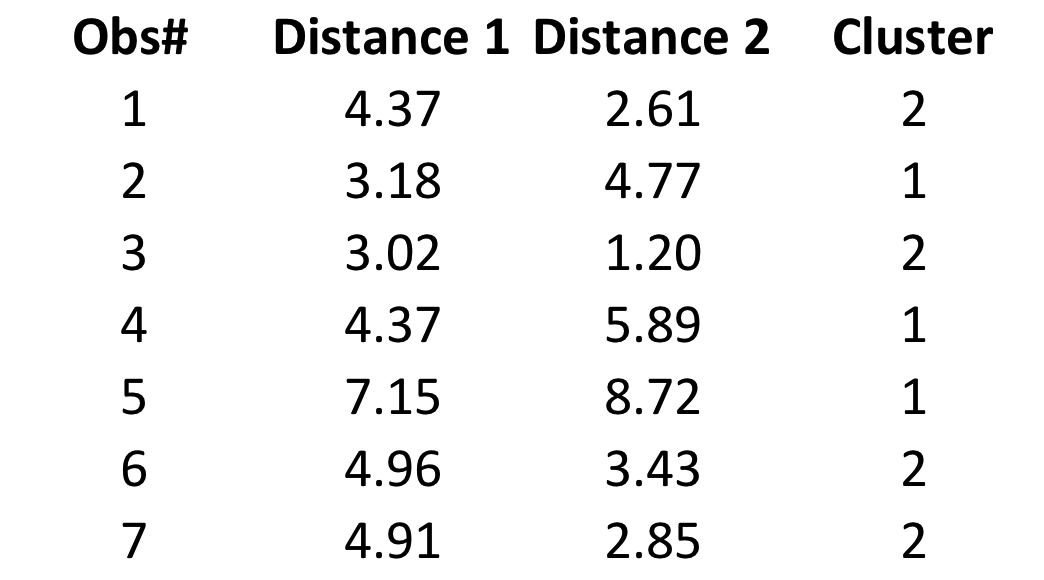
Step 2) Calculate the centroid of each cluster:

=(X1+X2+X3….Xp)/p where p is the number of observations in each cluster.



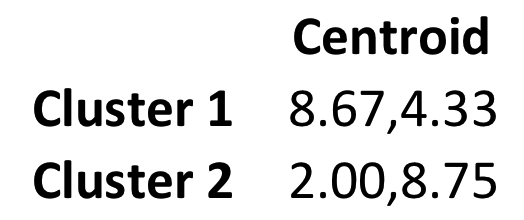
Step 3) Calculate the Euclidean distance of each observation from their two centroids. The distances from the centroids are labeled as Distance 1 and Distance 2. Compare the both EUD for an observation and assign it to the cluster where EUD to the cluster centroid minimum (Table 8).

Table 8:Assignment based on minimum EUD



***ITERATION 2)***

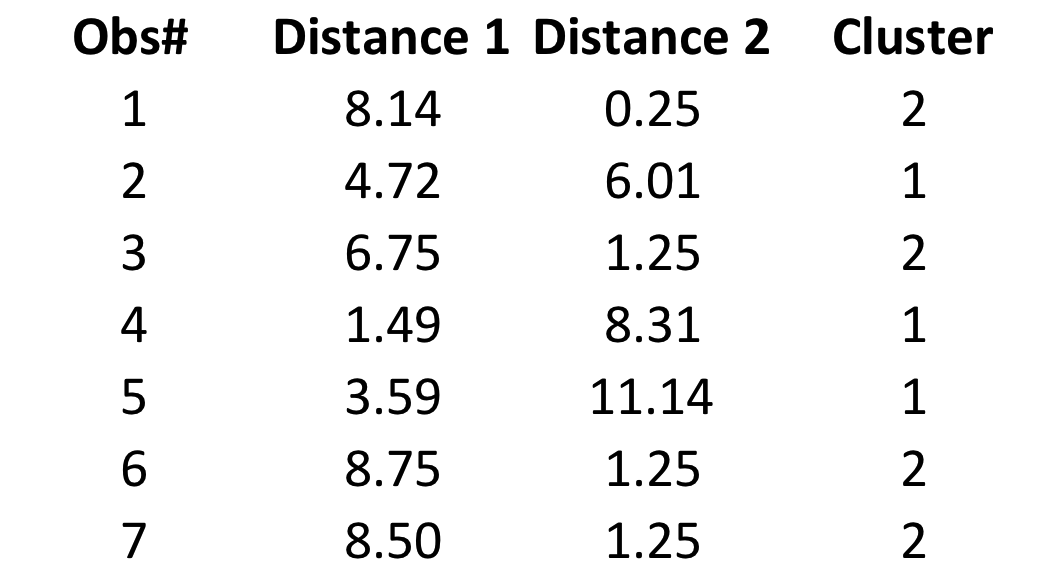
Step 1) Recalculate the centroid for the assignment in Table 8.



Step 2)

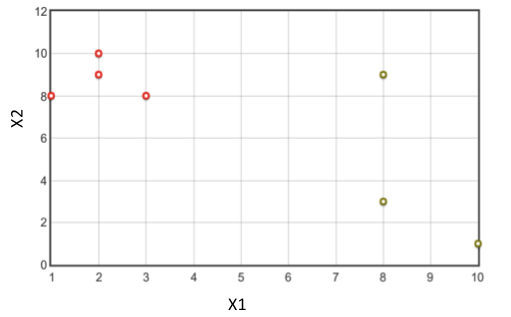
Calculate the EUD from the cluster centroids for each observation.

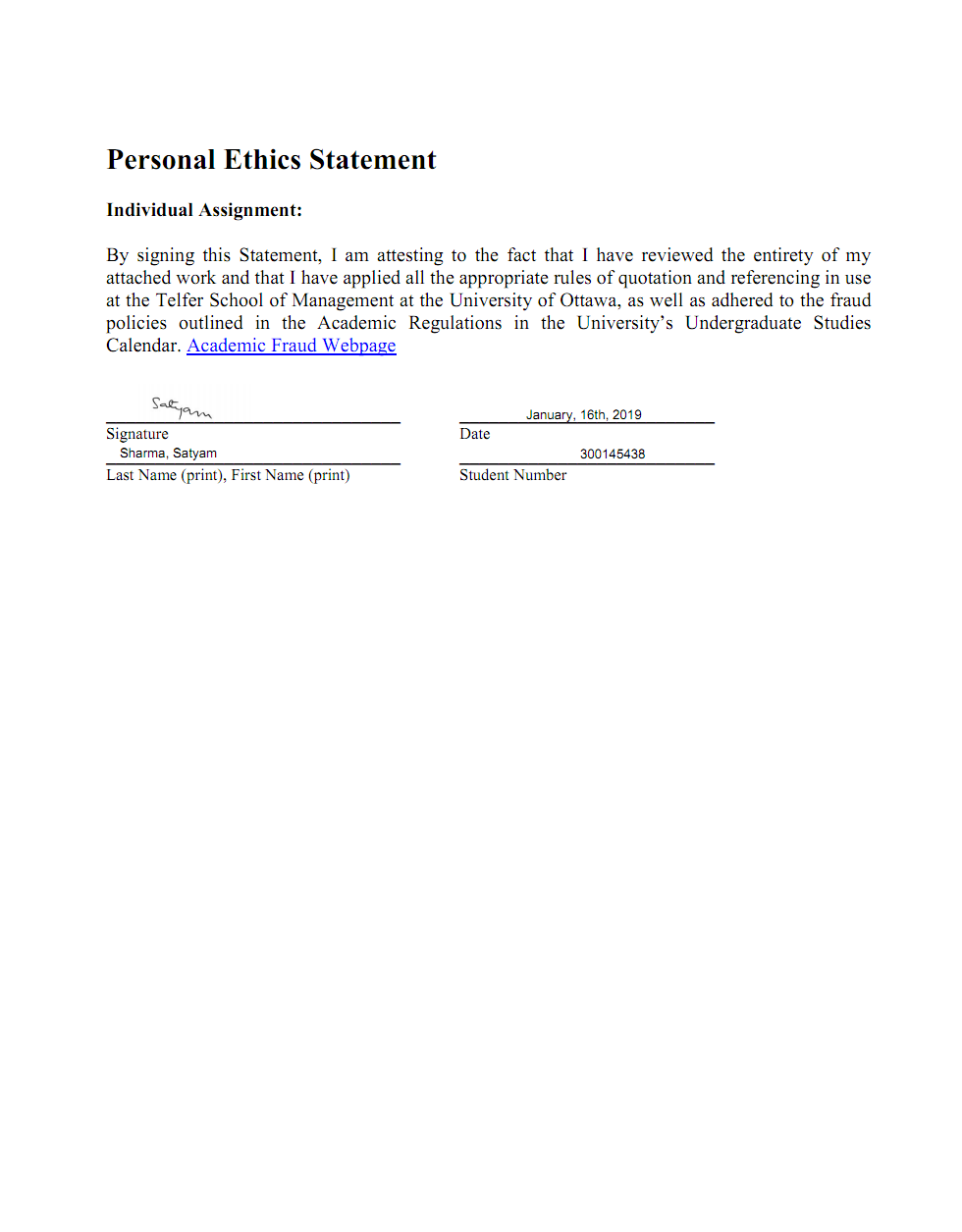
Table 9:Assignment to Clusters based on minimum EUD(iteration2)



By comparing tables 8 and 9, the composition of clusters remains the same implying the convergence of algorithm. Plot of the clusters is shown in Table 10

Table 10:K Means Clustering





March 18,2020

1. ~ denotes EUD between two objects [↑](#footnote-ref-1)