ASSIGNMENT-12

AKSHAY SRIRANGAM (AXS156630)

VISHNU KUMAR REDDY CHINAREDDY (VXC151130)

ANUPAMA VIJAYAN (AXV146030)

Cluster Analysis on Water Treatment:

a) Hierarchical clustering.

b) kmeans clustering.

## Hierarchical clustering

##Creating the model

>hc.out<-hclust(dist(a))

> hc.out

Call:

hclust(d = dist(a))

Cluster method : complete

Distance : euclidean

Number of objects: 380

##summary

> summary(hc.out)

Length Class Mode

merge 758 -none- numeric

height 379 -none- numeric

order 380 -none- numeric

labels 380 -none- character

method 1 -none- character

call 2 -none- call

dist.method 1 -none- character

##Extracting the clusters

>hc.clusters =cutree (hc.out ,4)

#summary

> summary(hc.clusters)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 1.000 1.000 1.442 2.000 4.000

#Plots: (these are included in the r code)

> plot(hclust (dist(a),method ="complete"),main="CompleteLink ", xlab ="", sub ="", ylab ="")

>plot(hclust (dist(a),method ="average"),main="averageLinkage ", xlab ="", sub ="", ylab ="")

## Kmeans Clustering

#Creating the cluster

>km.out=kmeans(a,4,nstart = 10)

Within cluster sum of squares by cluster:

[1] 618341733 617015080 611767209 848387861

(between\_SS / total\_SS = 85.1 %)

Available components:

[1] "cluster" "centers" "totss"

[4] "withinss" "tot.withinss" "betweenss"

[7] "size" "iter" "ifault"

#summary

> summary(km.out)

Length Class Mode

cluster 380 -none- numeric

centers 152 -none- numeric

totss 1 -none- numeric

withinss 4 -none- numeric

tot.withinss 1 -none- numeric

betweenss 1 -none- numeric

size 4 -none- numeric

iter 1 -none- numeric

ifault 1 -none- numeric

#Extracting the clusters

>km.clusters =km.out$cluster

#Summary

> summary(km.clusters)

Min. 1st Qu. Median Mean 3rd Qu. Max.

1.000 2.000 3.000 2.682 3.000 4.000

#Comparison

>table(km.clusters ,hc.clusters )

> table(km.clusters ,hc.clusters )

hc.clusters

km.clusters 1 2 3 4

1 0 23 16 0

2 0 104 0 0

3 176 0 0 0

4 58 0 0 3