Rituraj Anand 19BEC1089

CSE3506: Essentials of Data Analytics

Ex-08: Hierarchical clustering

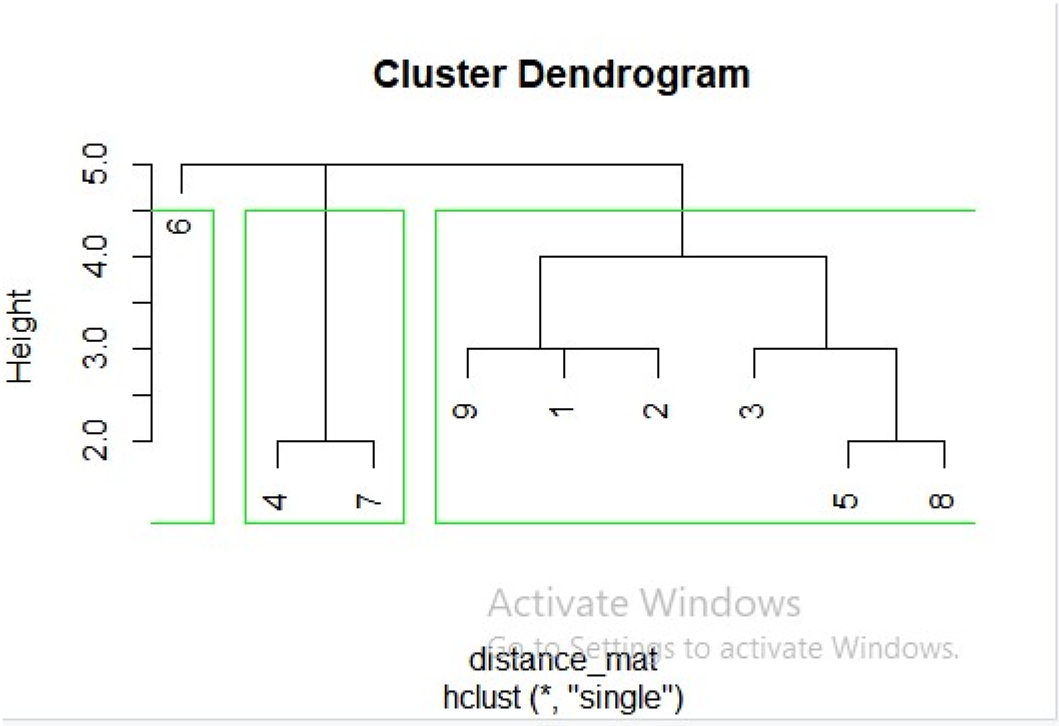
Aim: to perform hierarchical clustering using manhattan and euclidean distance with both single and complete linkage methods.

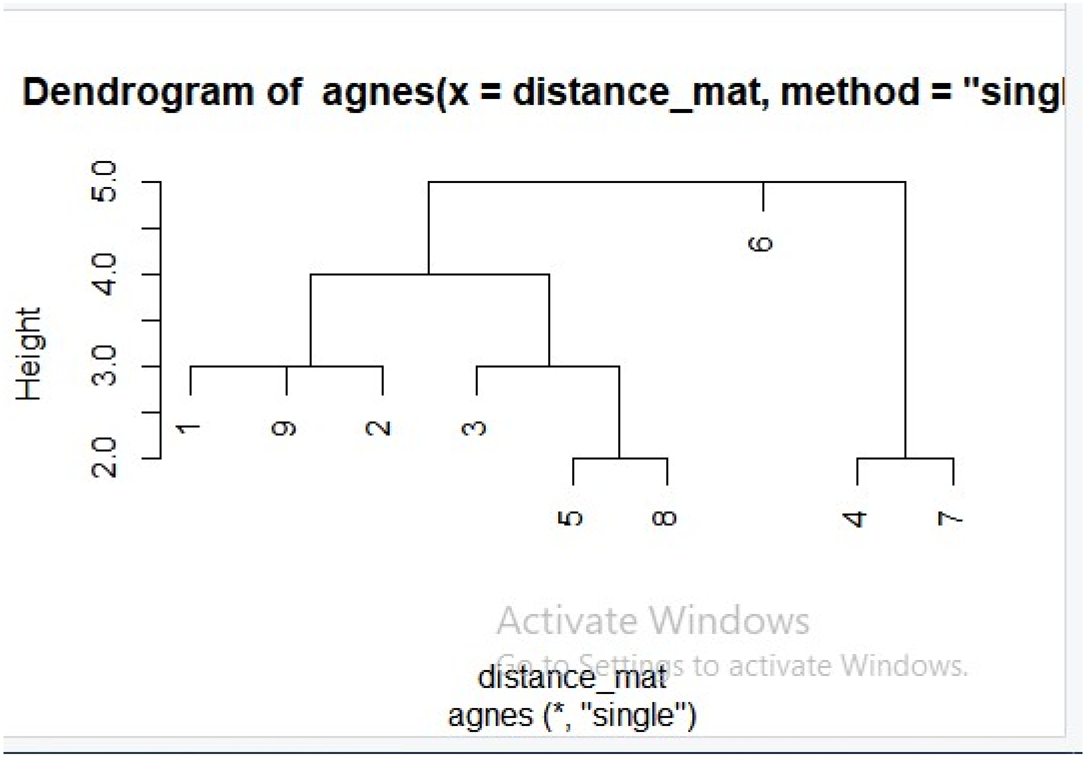
Tools: R studio

Method-1:

##Manhattan Distance ,single linkage method

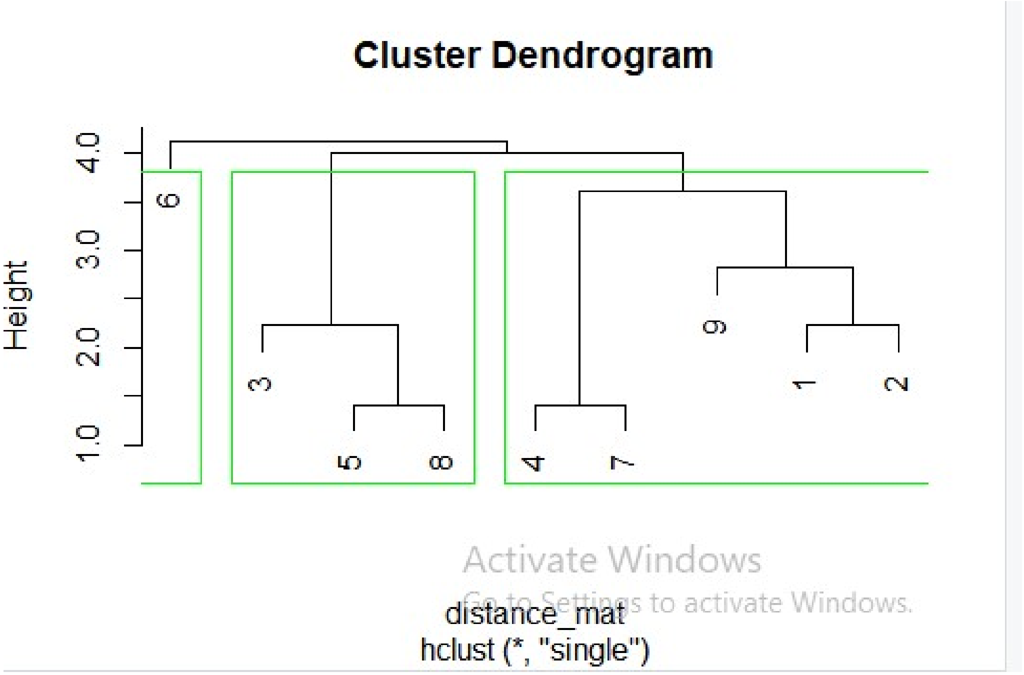
PLOTS:

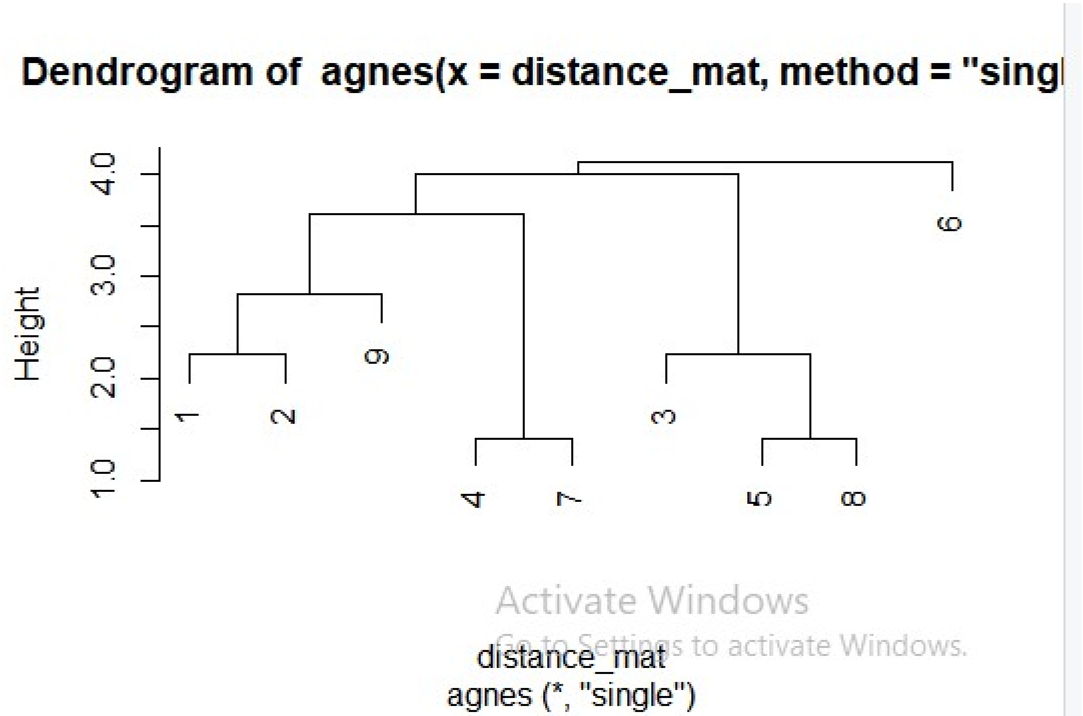

Cluster Dendrogram
Height
2.0
3.0
4.0
5.0
6
4
7
9
1
2
3
5
8
Activate Windows
Go to Settings to activate Windows.
distance_mat
hclust (*, "single")



Method-2:

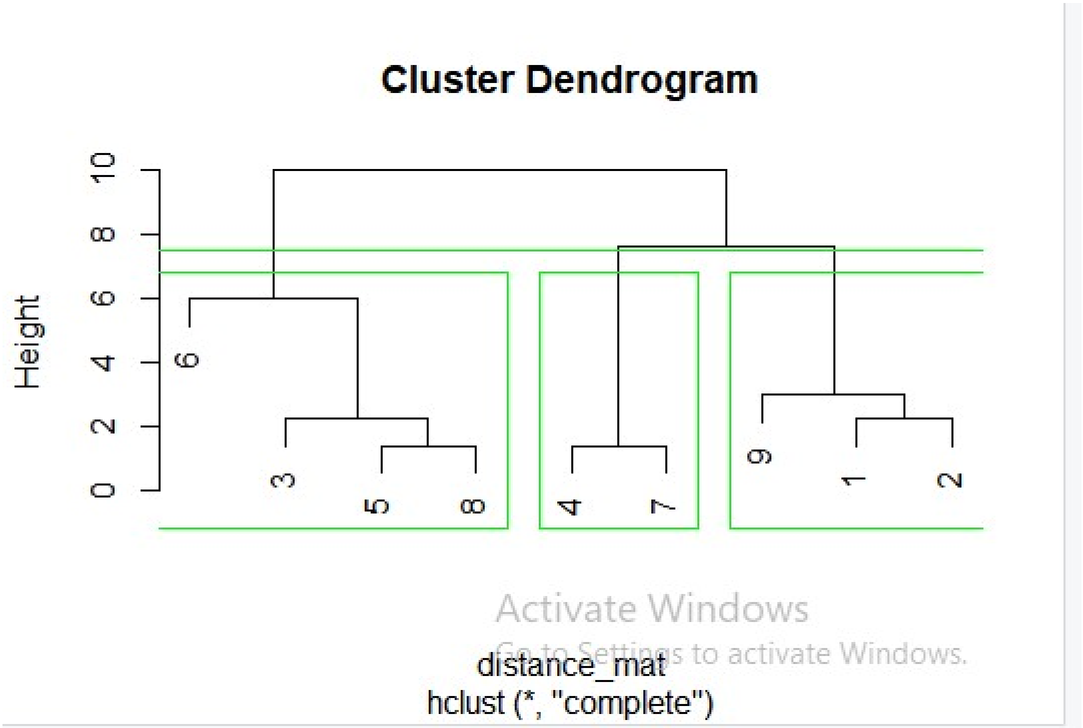
##Euclidean Distance ,single linkage method

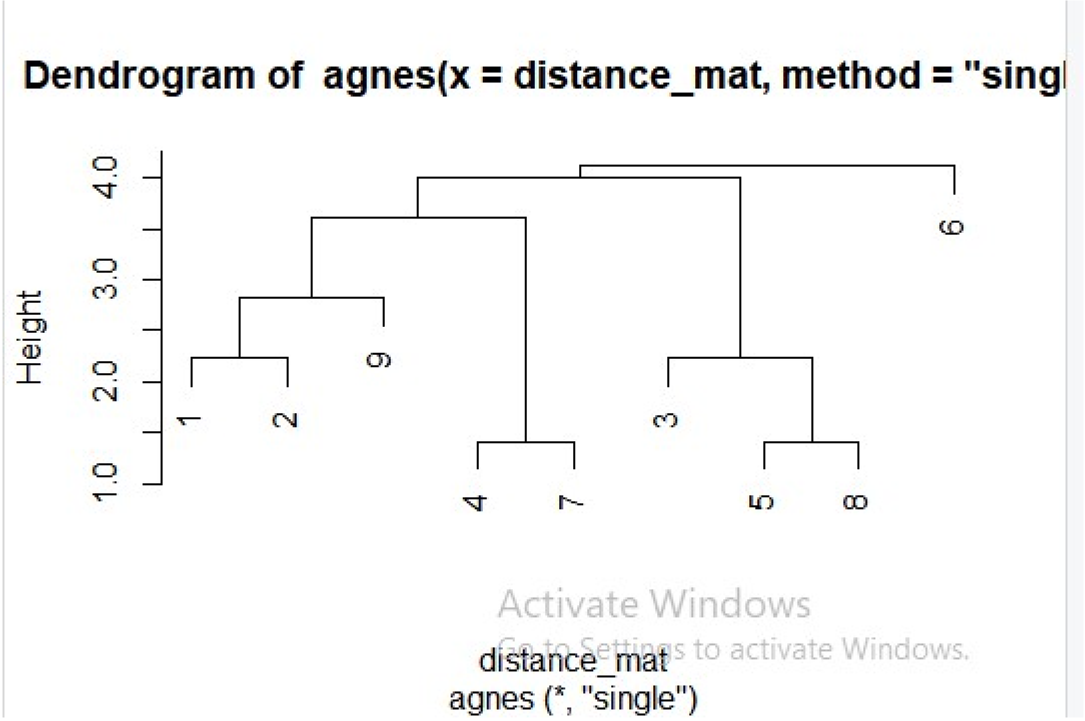




Method-3:

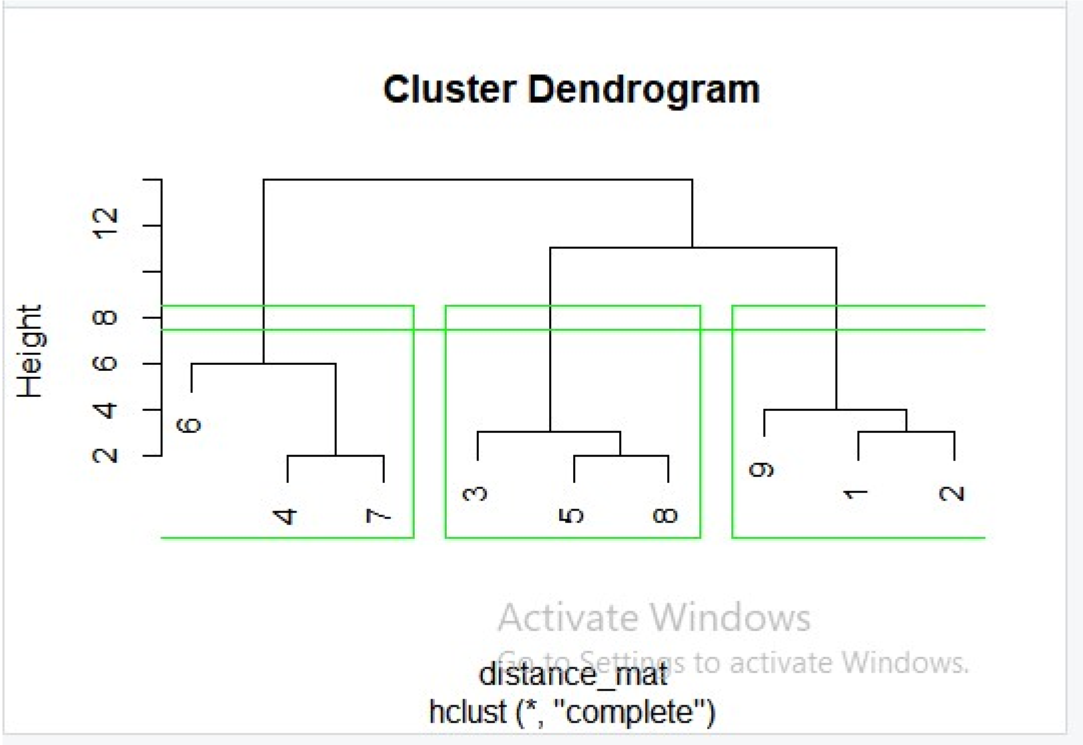
##Euclidean Distance ,complete linkage method

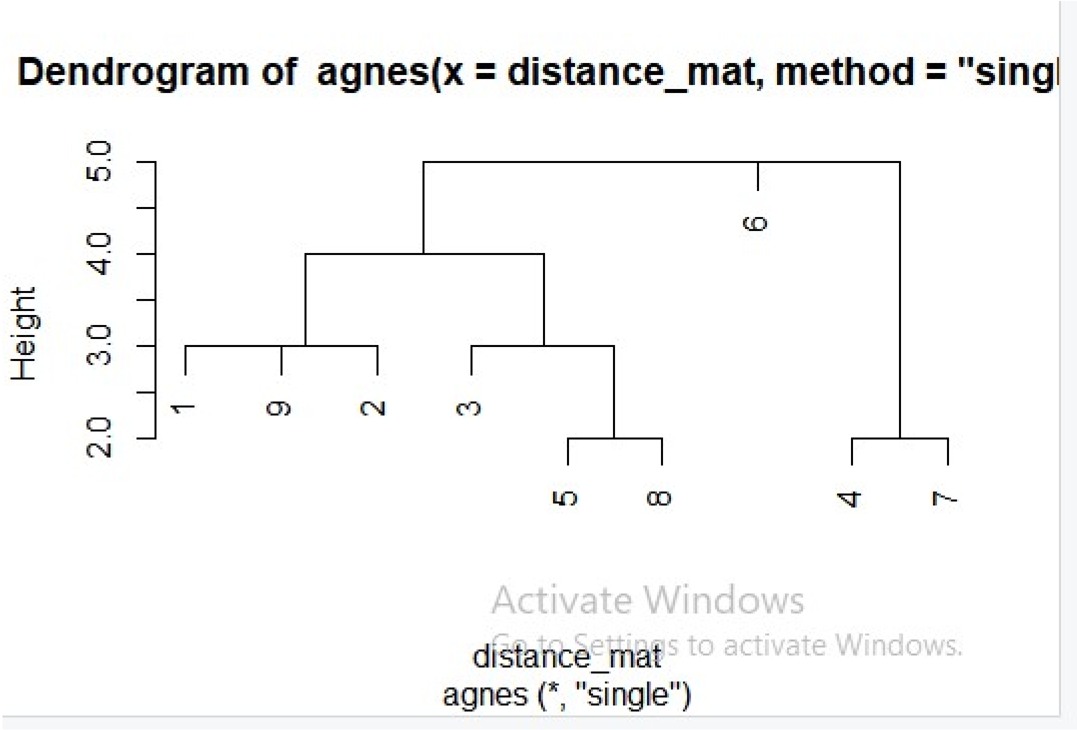




Method-4:

##Manhattan Distance ,complete linkage method





Inference:

Here we saw, the difference between dendograms when different methods were chosen.

In case of single linkage, the manhattan distance, we check for farthest points and vice versa.

In case of complete linkage, the manhattan distance, we check for closest points and vice versa.

#ANNEX

#Code

```
rm(list=ls())

library(cluster)
library(dplyr)

v2=c(1,2,1,7,2,7,8,3,4)
v1=c(1,3,9,3,7,9,4,8,1)

df=data.frame(v1,v2)
#DO FOR ALL POSSIBLE COMBN
#manhattan, euclidean
#single, complete

distance_mat =dist(df,method = 'manhattan')
distance_mat
```

```
hierar_cl=hclust(distance_mat,method= "single")
hierar_cl
plot(hierar_cl)
#choosing no of clusters
#cutting tree by height
abline(h=7.5 ,col ="green")
#cutting tree by no of clusters
fit = cutree(hierar_cl, k=3)
fit
table(fit) # frq of occurence
rect.hclust(hierar_cl, k=3 ,border = "green")








## eucliodian ,single


rm(list=ls())


library(cluster)
library(dplyr)
```

```
v2=c(1,2,1,7,2,7,8,3,4)

v1=c(1,3,9,3,7,9,4,8,1)


df=data.frame(v1,v2)

#DO FOR ALL POSSIBLE COMBN

#manhattan, euclidean

#single, complete


distance_mat =dist(df,method = 'euclidean')

distance_mat


hierar_cl=hclust(distance_mat,method= "single")

hierar_cl


plot(hierar_cl)


#choosing no of clusters

#cutting tree by height


abline(h=7.5 ,col ="green")


#cutting tree by no of clusters


fit = cutree(hierar_cl, k=3)

fit
```

```
table(fit) # frq of occurence


rect.hclust(hierar_cl, k=3 ,border = "green")


#------------------------------------------

Hierar_agnes<-agnes(distance_mat,method="single")

pltree(Hierar_agnes)
```