MAS8404 Practical Report 2

Mansoor Hamid

11/9/2019

### Question 1:

#### Part(a)

genexpr\_data=read.csv("C:/Users/User/Documents/RWorkSpace/data/Ch10Ex11.csv",header=TRUE)  
#Check dimesion of the loaded data  
dim(genexpr\_data)

## [1] 999 40

#transpose once to correct the data  
genexpr\_data = t(genexpr\_data)

##### Creating correlation distance matrix

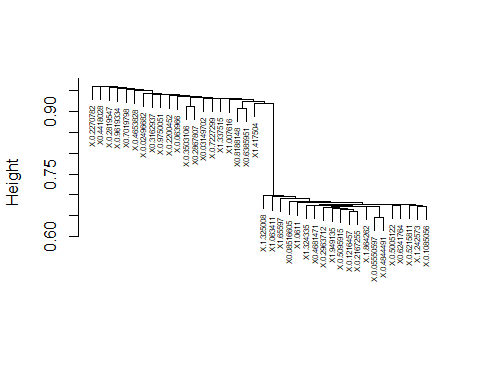
#Transpose the matrix in order to use cor function  
genexpr\_data\_transposed = t(genexpr\_data)  
  
#Compute the inter-observation correlations  
r\_obs = cor(genexpr\_data\_transposed)  
#Create the distance matrix  
d\_cor = 1-r\_obs  
  
#convert the distance matrix to dis object  
d\_cor = as.dist(d\_cor)

##### Perform hierarchical clustering using hclust with metod specified as “single” to indicate single-linkage

hc\_c = hclust(d\_cor, method = "single")

##### Plot the dendrogram

plot(hc\_c, cex=0.5,main="",sub="",xlab="")



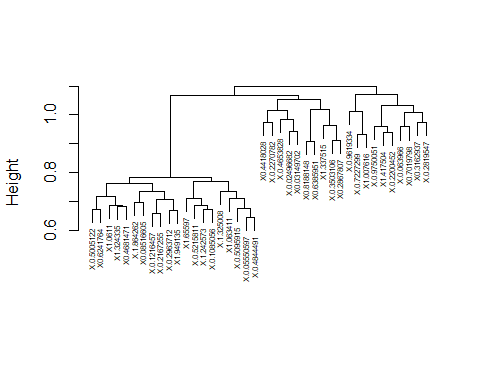
#### Part(b)

##### Perform hierarchical clustering using hclust with metod specified as “complete” and “average”

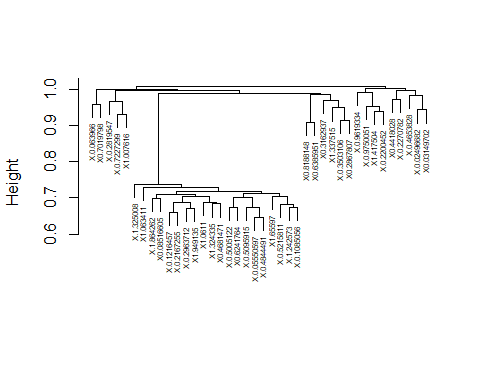
hc\_c\_complete = hclust(d\_cor, method = "complete")  
hc\_c\_average = hclust(d\_cor, method = "average")

##### Plot the dendrograms

plot(hc\_c\_complete, cex=0.5,main="",sub="",xlab="")



plot(hc\_c\_average, cex=0.5,main="",sub="",xlab="")



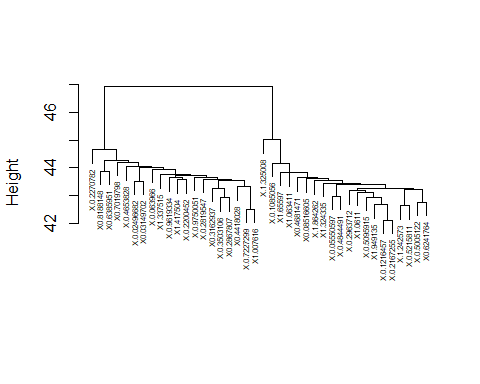
#### Part(c)

###### Determine Euclidean distance and perform hierarchical clustering with single-linkage

genexpr\_data\_euc = dist(genexpr\_data)  
hc\_c\_euc\_single = hclust(genexpr\_data\_euc, method = "single")

##### Plot the dendrogram

plot(hc\_c\_euc\_single, cex=0.5,main="",sub="",xlab="")



### Question 2:

#### Part(a)

#Scale the data  
Kmax = 10  
genexpr\_data\_scaled = scale(genexpr\_data)