R Codes – Hierarchical Clustering

1. Installing Packages

install.packages(“clusters”)

install.packages(“xlsx”)

library(“xlsx”)

1. Review And Setting Dataset

-View(“dataset”)

-dataset$UNS *# to delete non-numerical attribute.*

-d <- dist(dataset,method = "euclidian",upper = TRUE) *# to calculate distances*

*between values each cases base on euclidian method.*

-round(d,2) *# to convert each value to two-digit numbers.*

1. Clustering Operations

-plot(hclust(d,method = "complete")) *# to create dendrogram tree base on complete method(complete method is provide clustering , combining remote points )*

-h <- hclust(d,method = "complete") *# to identify points where clusters are joined*

-h$height

i <- which.max(diff(h$height))

cut\_height <- (h$height[i] + h$height[i+1])/2 *# to determine the midpoint of the point where the greatest difference begins*

cut\_height

clusters <- cutree(h,h=cut\_height) # *to seperate each case to clusters*

clusters

par(mfrow=c(1,2)) *#  A vector of length 2, where the first argument specifies the number of rows and the second the number of columns of plots.*

plot(h)

abline(h=cut\_height,col="red",lty=2) *# cut tree which determined before (cut\_height)*