**ASSIGNMENT 4**

**Name: Jain Samkitkumar Hasmukhlal**

**Roll No.: 20 (BEIT)**

**K MEANS CLUSTERING**

> dim(iris)

[1] 150 5

> names(iris)

[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"

> str(iris)

'data.frame': 150 obs. of 5 variables:

$ Sepal.Length: num 5.1 4.9 4.7 4.6 5 5.4 4.6 5 4.4 4.9 ...

$ Sepal.Width : num 3.5 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 ...

$ Petal.Length: num 1.4 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 ...

$ Petal.Width : num 0.2 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 ...

$ Species : Factor w/ 3 levels "setosa","versicolor",..: 1 1 1 1 1 1 1 1 1 1 ...

> attributes(iris)

$names

[1] "Sepal.Length" "Sepal.Width" "Petal.Length" "Petal.Width" "Species"

$row.names

[1] 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22

[23] 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44

[45] 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66

[67] 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88

[89] 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110

[111] 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132

[133] 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150

$class

[1] "data.frame"

> iris[1:5,]

Sepal.Length Sepal.Width Petal.Length Petal.Width Species

1 5.1 3.5 1.4 0.2 setosa

2 4.9 3.0 1.4 0.2 setosa

3 4.7 3.2 1.3 0.2 setosa

4 4.6 3.1 1.5 0.2 setosa

5 5.0 3.6 1.4 0.2 setosa

> newiris <- iris

> newiris$Species <- NULL

> (kc <- kmeans(newiris, 3))

K-means clustering with 3 clusters of sizes 38, 50, 62

Cluster means:

Sepal.Length Sepal.Width Petal.Length Petal.Width

1 6.850000 3.073684 5.742105 2.071053

2 5.006000 3.428000 1.462000 0.246000

3 5.901613 2.748387 4.393548 1.433871

Clustering vector:

[1] 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2

[45] 2 2 2 2 2 2 3 3 1 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 3 1 3 3 3 3 3 3 3 3 3 3

[89] 3 3 3 3 3 3 3 3 3 3 3 3 1 3 1 1 1 1 3 1 1 1 1 1 1 3 3 1 1 1 1 3 1 3 1 3 1 1 3 3 1 1 1 1

[133] 1 3 1 1 1 1 3 1 1 1 3 1 1 1 3 1 1 3

Within cluster sum of squares by cluster:

[1] 23.87947 15.15100 39.82097

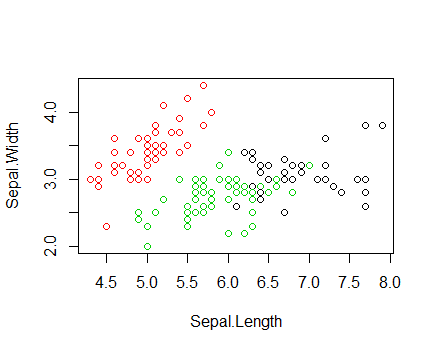
(between\_SS / total\_SS = 88.4 %)

Available components:

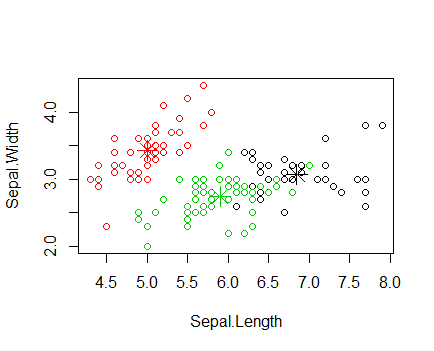
[1] "cluster" "centers" "totss" "withinss" "tot.withinss" "betweenss"

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

|  |
| --- |
| > table(iris$Species, kc$cluster)    1 2 3  setosa 0 50 0  versicolor 2 0 48  virginica 36 0 14  > plot(newiris[c("Sepal.Length", "Sepal.Width")], col=kc$cluster) |
|  |
| |  | | --- | |  | |



> points(kc$centers[,c("Sepal.Length", "Sepal.Width")], col=1:3, pch=8, cex=2)



**HIERARCHICAL CLUSTERING**

> install.packages("C:/Users/Solaris lap/Downloads/e1071\_1.6-7.zip", repos = NULL, type = "win.binary")

Installing package into ‘C:/Users/Solaris lap/Documents/R/win-library/3.2’

(as ‘lib’ is unspecified)

package ‘e1071’ successfully unpacked and MD5 sums checked

> library(e1071)

> idx <- sample(1:dim(iris)[1], 40)

> irisSample <- iris[idx,]

> irisSample$Species <- NULL

> hc <- hclust(dist(irisSample), method="ave")

> plot(hc, hang = -1, labels=iris$Species[idx])

