**THE SPARKS FOUNDATION**

**TASK 1-**

**PREDICTION USING**

**SUPERVISED MACHINE**

**LEARNING**

BY-SIMRAN MOHATA

> #Storing Iris data in variable "data"

> data<- iris[,-5]

> #Displaying the first 6 records

> head(data)

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

1 5.1 3.5 1.4 0.2

2 4.9 3.0 1.4 0.2

3 4.7 3.2 1.3 0.2

4 4.6 3.1 1.5 0.2

5 5.0 3.6 1.4 0.2

6 5.4 3.9 1.7 0.4

> # function to compute total within-cluster sum of square

> f<- function(x)

+ {

+ within<- kmeans(data,x)$tot.withinss

+ }

> # Compute and plot wss for x = 1 to x = 15

> x<- 1:15

> plot(x,sapply(x,f),type="b",lwd="2")

> #K means clustering with 3 clusters

> model1<- kmeans(data,3)

> model1

K-means clustering with 3 clusters of sizes 21, 96, 33

Cluster means:

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

1 4.738095 2.904762 1.790476 0.3523810

2 6.314583 2.895833 4.973958 1.7031250

3 5.175758 3.624242 1.472727 0.2727273

Clustering vector:

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

[47] 3 1 3 3 2 2 2 2 2 2 2 1 2 2 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

[93] 2 1 2 2 2 2 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

[139] 2 2 2 2 2 2 2 2 2 2 2 2

Within cluster sum of squares by cluster:

[1] 17.669524 118.651875 6.432121

(between\_SS / total\_SS = 79.0 %)

Available components:

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

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

> y<- iris$Species

> table(y,model1$cluster)

y 1 2 3

setosa 17 0 33

versicolor 4 46 0

virginica 0 50 0

> #Visualizing the results

> plot(data$Sepal.Length,data$Sepal.Width,col=model1$cluster,lwd=6,xlab="Sepal Length",

+ main="Kmeans Clustering

+ Data-Iris",

+ ylab="Sepal Width")

> #Adding centroids to our graph

> points(model1$centers[,1:2],col="blue",lwd=10)

> legend("topright",c("Setosa","Versicolor","Virginica","Centroids"),

+ fill=c(unique(model1$cluster),"blue"))

