**THE SPARKS FOUNDATION**

**TASK 1-**

**PREDICTION USING**

**SUPERVISED MACHINE**

**LEARNING**

BY-SIMRAN MOHATA

> #importing data

> data <- read.csv("C:\\Users\\user\\Desktop\\Sparks\\student\_score.csv")

> #Displaying the first 6 values of the data

> head(data)

Hours Scores

1 2.5 21

2 5.1 47

3 3.2 27

4 8.5 75

5 3.5 30

6 1.5 20

> View(data)

> #For obtaining summary statistics

> summary(data)

Hours Scores

Min. :1.100 Min. :17.00

1st Qu.:2.700 1st Qu.:30.00

Median :4.800 Median :47.00

Mean :5.012 Mean :51.48

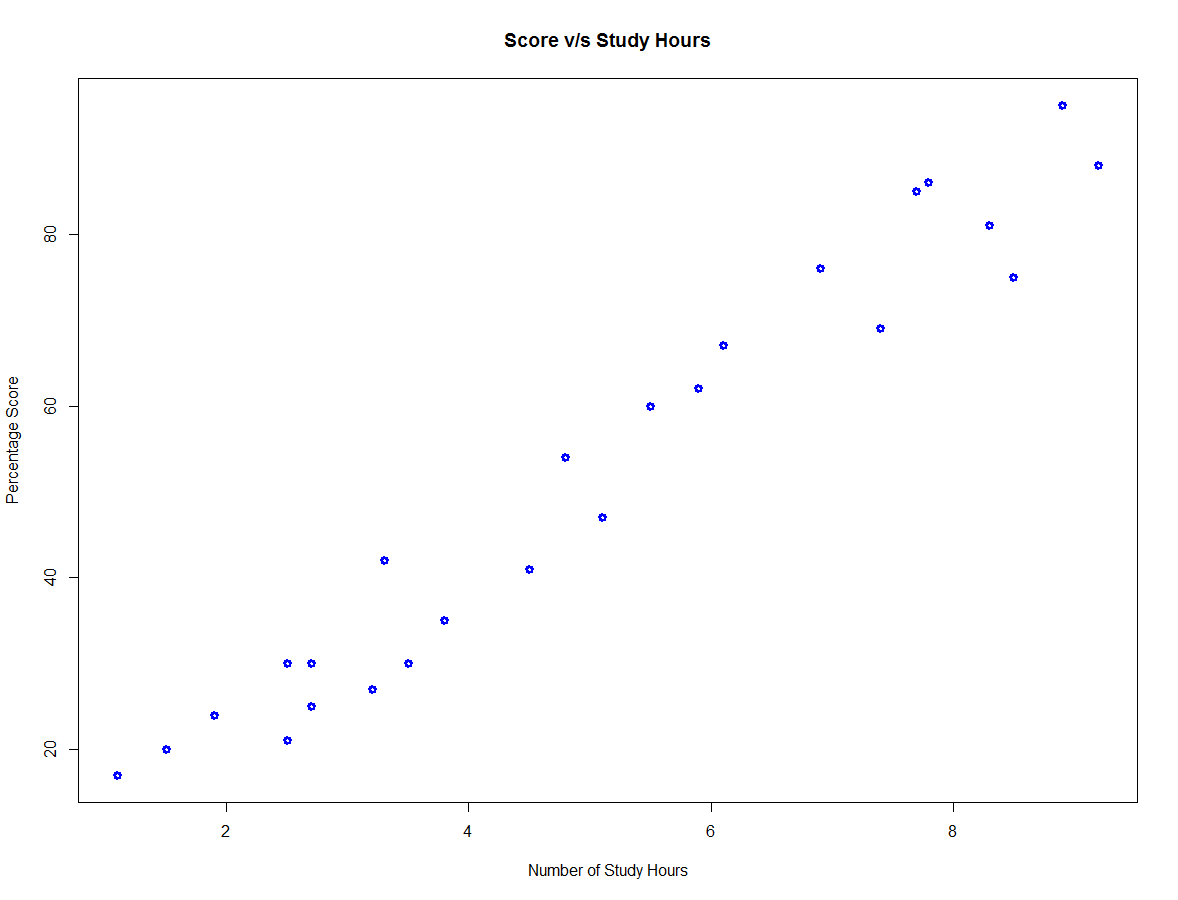
3rd Qu.:7.400 3rd Qu.:75.00

Max. :9.200 Max. :95.00

> #Plotting the given data

> plot(data$Hours,data$Scores,xlab="Number of Study Hours",ylab="Percentage Score",

+ main="Score v/s Study Hours",col="blue",lwd="3")



> #A linear regression model using "lm" function

> linear\_relation<- lm(Scores~Hours,data=data)

> linear\_relation

Call:

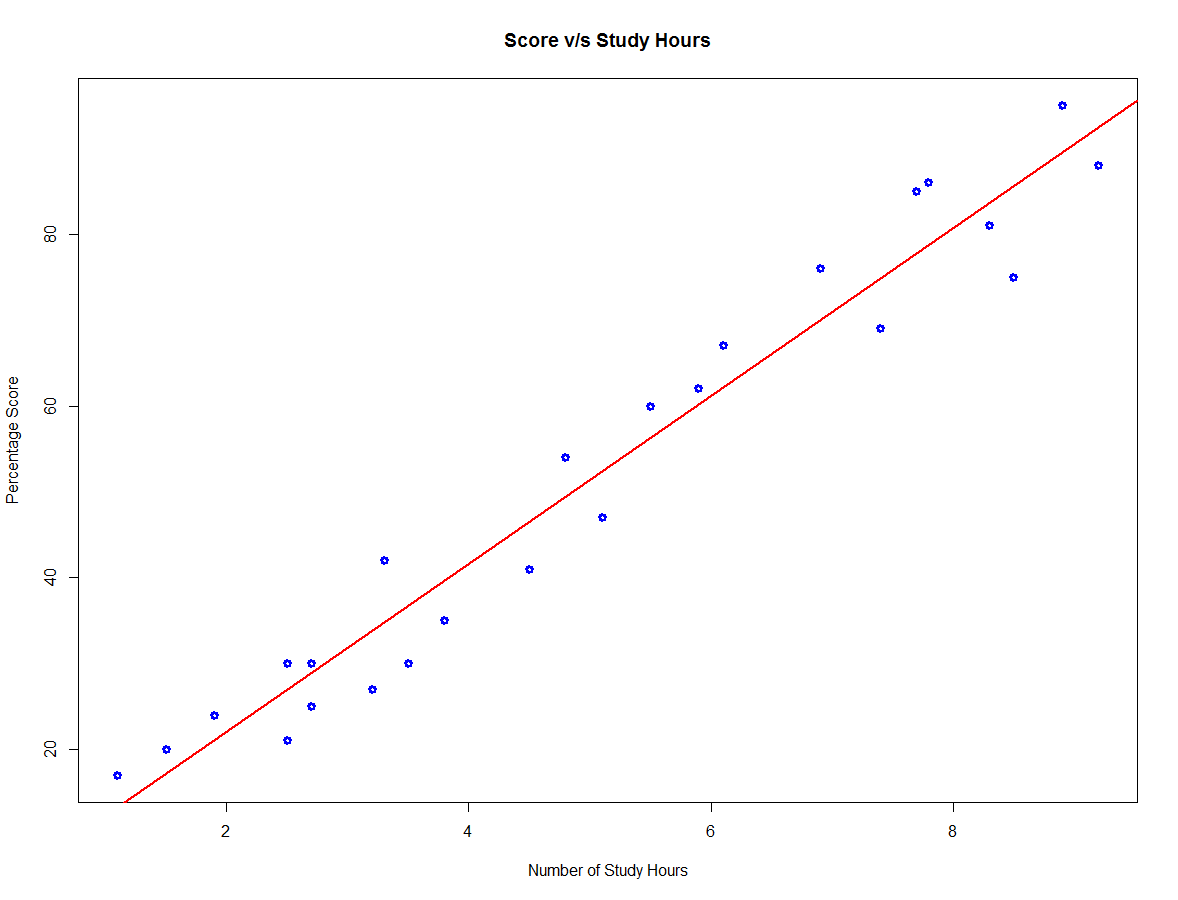
lm(formula = Scores ~ Hours, data = data)

Coefficients:

(Intercept) Hours

2.484 9.776

> abline(linear\_relation$coefficients,col="red",lwd="2")



> p<-predict(linear\_relation)

> data.frame(data$Scores,p)

data.Scores p

1 21 26.92318

2 47 52.34027

3 27 33.76624

4 75 85.57800

5 30 36.69899

6 20 17.14738

7 88 92.42106

8 60 56.25059

9 81 83.62284

10 25 28.87834

11 85 77.75736

12 62 60.16091

13 41 46.47479

14 42 34.74382

15 17 13.23706

16 95 89.48832

17 30 26.92318

18 24 21.05770

19 67 62.11607

20 69 74.82462

21 30 28.87834

22 54 49.40753

23 35 39.63173

24 76 69.93672

25 86 78.73494

> #Predicted score when the student studies for 9.25 hours

> predict(linear\_relation,data.frame(Hours=9.25))

1

92.90985

> #installing package MLmetrics

> install.packages("MLmetrics")

https://cran.rstudio.com/bin/windows/Rtools/

Installing package into ‘C:/Users/user/Documents/R/win-library/3.6’

(as ‘lib’ is unspecified)

Warning in install.packages :

package ‘MLmetrics’ is in use and will not be installed

> library("MLmetrics")

> #computes the average absolute difference between two numeric vectors

> MAE(data$Scores,p)

[1] 4.972805