TASK 1- PREDICTION USING SUPERVISED

MACHINE LEARNING

#Import the data in R

> student.data<-read.csv("C:\\Users\\User\\Desktop\\GRIP\\5\_6215462502226985478.csv")

>

#Check if you have imported the correct data by

checking the head

> head(student.data)

Hours Scores

1 2.5 21

2 5.1 47

3 3.2 27

4 8.5 75

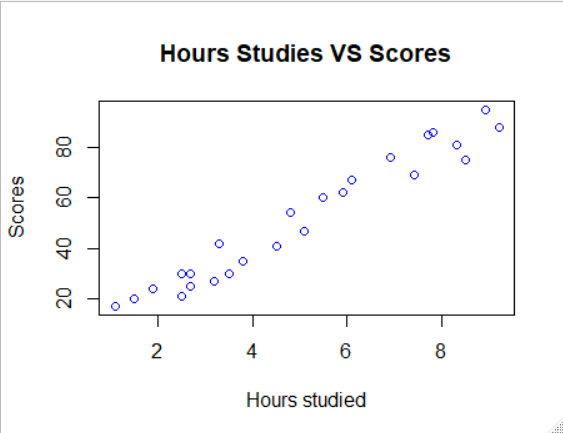
5 3.5 30

6 1.5 20

#Plot the Hours studied against the scores

> plot(student.data$Hours,student.data$Scores,xlab="Hours studied",

+ ylab="Scores",main="Hours Studies VS Scores",col="blue")



#Make a linear regression model using 'lm' function

> student.reg<-lm(Scores~Hours,student.data)

> summary(student.reg)

Call:

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

Residuals:

Min 1Q Median 3Q Max

-10.578 -5.340 1.839 4.593 7.265

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 2.4837 2.5317 0.981 0.337

Hours 9.7758 0.4529 21.583 <2e-16 \*\*\*

---

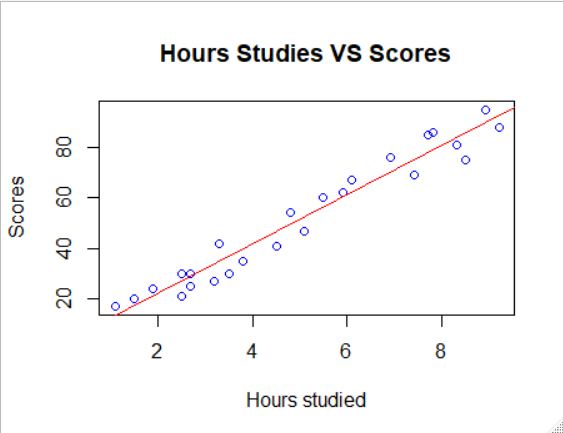
Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.603 on 23 degrees of freedom

Multiple R-squared: 0.9529, Adjusted R-squared: 0.9509

F-statistic: 465.8 on 1 and 23 DF, p-value: < 2.2e-16

> abline(student.reg,col="red")



#Predict the percentage of a student by entering study hours

> new\_hour<-data.frame(Hours=9.25)

> predict(student.reg,new\_hour)

1

92.90985

**THE PREDICTED SCORE IS 92.909.**

> #Check the error

> p<-predict(student.reg)

> AP<-data.frame(student.data$Scores,p)

> AP

student.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

> aberror<-mean(abs(student.data$Scores-p))

> aberror

[1] 4.972805