TASK1.R

2022-02-04

## THE SPARKS FOUNDATION- DATA SCIENCE AND BUSINESS ANALYTICS   
### AUTHOR: DEVANSH SAINI  
#### GRIP- FEB22  
#### TASK1: Prediction using Supervised Machine Learning  
##### Objective: Predict the percentage of a student based on the no. of study hours

#### To import data set   
library(readxl)  
STUDENT\_SCORES <- read\_excel("D:/DESKTOP/STUDENT SCORES.xlsx")  
#### To view the data in R  
View(STUDENT\_SCORES)  
#### To attach the data   
 attach(STUDENT\_SCORES)  
#### To Summarize the data  
 summary(STUDENT\_SCORES)

## 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

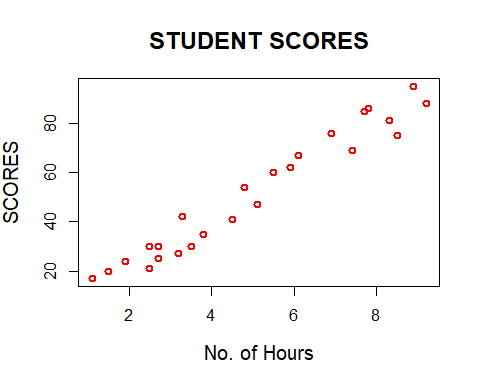
#### To Call out variable names  
 names(STUDENT\_SCORES)

## [1] "Hours" "Scores"

#### To make simple regression model  
 mod <- lm(Scores~Hours)  
  
#### To know the summarry of model  
 summary(mod)

##   
## Call:  
## lm(formula = Scores ~ Hours)  
##   
## 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

### To plot the data   
plot(Hours,Scores,main="STUDENT SCORES",  
 xlab="No. of Hours",ylab="SCORES",cex.main=1.5,cex.lab=1.25,  
 col="red",lwd=2)



#### To call the coeficient   
 coef(mod)

## (Intercept) Hours   
## 2.483673 9.775803

#### Prediction of Marks if study 9.25 hours  
 2.483673 + (9.775803)\*9.25

## [1] 92.90985

#or using code  
 predict(mod, newdata = data.frame(Hours=9.25))

## 1   
## 92.90985