Grip Task 1

Nishant Kumar (19138-BMS 2C)

12/02/2021

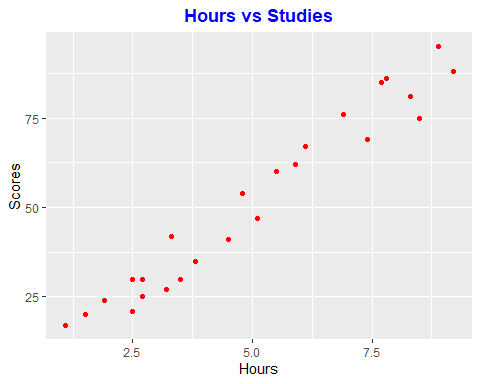
file<-'http://bit.ly/w-data'  
data<-read.csv(file,header = T)  
attach(data)  
head(data,5)

## Hours Scores  
## 1 2.5 21  
## 2 5.1 47  
## 3 3.2 27  
## 4 8.5 75  
## 5 3.5 30

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

#Ploting the dataset  
par(mar=c(1,1,1,1))  
library(ggplot2)  
g1<-ggplot(data,aes(x=Hours,y=Scores))+geom\_point(col='red')+labs(title = "Hours vs Studies")  
g1<-g1+theme(plot.title=element\_text(hjust=0.5,face="bold", color="blue"))  
g1



#Splitting into training and testing data:  
dt = sort(sample(nrow(data), nrow(data)\*.8))  
train<-data[dt,]  
test<-data[-dt,]  
print(train)

## Hours Scores  
## 1 2.5 21  
## 2 5.1 47  
## 3 3.2 27  
## 6 1.5 20  
## 7 9.2 88  
## 8 5.5 60  
## 9 8.3 81  
## 10 2.7 25  
## 11 7.7 85  
## 13 4.5 41  
## 15 1.1 17  
## 16 8.9 95  
## 17 2.5 30  
## 18 1.9 24  
## 19 6.1 67  
## 20 7.4 69  
## 21 2.7 30  
## 22 4.8 54  
## 23 3.8 35  
## 24 6.9 76

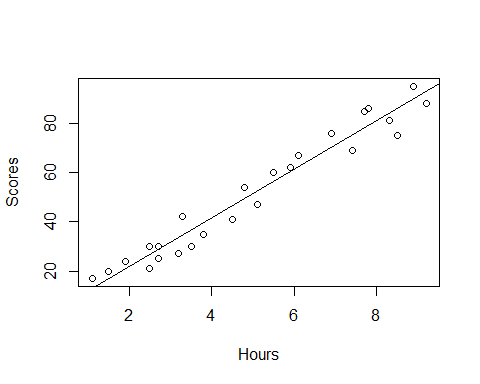
print(test)

## Hours Scores  
## 4 8.5 75  
## 5 3.5 30  
## 12 5.9 62  
## 14 3.3 42  
## 25 7.8 86

#Regression Model:  
model<-lm(Scores~Hours,train)  
model

##   
## Call:  
## lm(formula = Scores ~ Hours, data = train)  
##   
## Coefficients:  
## (Intercept) Hours   
## 1.863 9.914

#plotting the regression line:  
with(data,plot(Hours, Scores))  
abline(model)



#predicting hours using train data:  
pred = predict(model,test)  
pred

## 4 5 12 14 25   
## 86.13389 36.56281 60.35692 34.57996 79.19393

compare <-data.frame(test$Scores,pred)  
names(compare)[names(compare) == "test.Scores"] <- "Actual"  
names(compare)[names(compare) == "pred"] <- "Predicted"  
compare

## Actual Predicted  
## 4 75 86.13389  
## 5 30 36.56281  
## 12 62 60.35692  
## 14 42 34.57996  
## 25 86 79.19393

#calculate the error margin between prediction and actual values in test data  
library(MLmetrics)

## Warning: package 'MLmetrics' was built under R version 4.0.3

##   
## Attaching package: 'MLmetrics'

## The following object is masked from 'package:base':  
##   
## Recall

error=test$Scores-pred  
print(paste("Mean square error:",sum(error^2)/nrow(data)))

## [1] "Mean square error: 10.8445203930425"

print(paste("MAPE =",MAPE(compare$Predicted,compare$Actual)))

## [1] "MAPE = 0.129904215811921"

#Testing and predicting for the the given 9.25 hours:  
  
pred\_score=predict(model,data.frame(Hours=9.25))  
print(paste("The predicted score is: ",pred\_score))

## [1] "The predicted score is: 93.5695476444019"