

Prediction Using Supervised ML

Task - Predict the percentage of a student on the number of study hours

Submitted by - Kaustav Pal

The Libraries Used

In [1]:

```
library(ggplot2)
library(caTools)
```

Reading the Data

In [2]:

```
df<-read.csv(file.choose())
head(df)
```

A data.frame: 6 × 2

	Hours	Scores
	<dbl>	<int>
1	2.5	21
2	5.1	47
3	3.2	27
4	8.5	75
5	3.5	30
6	1.5	20

In [3]:

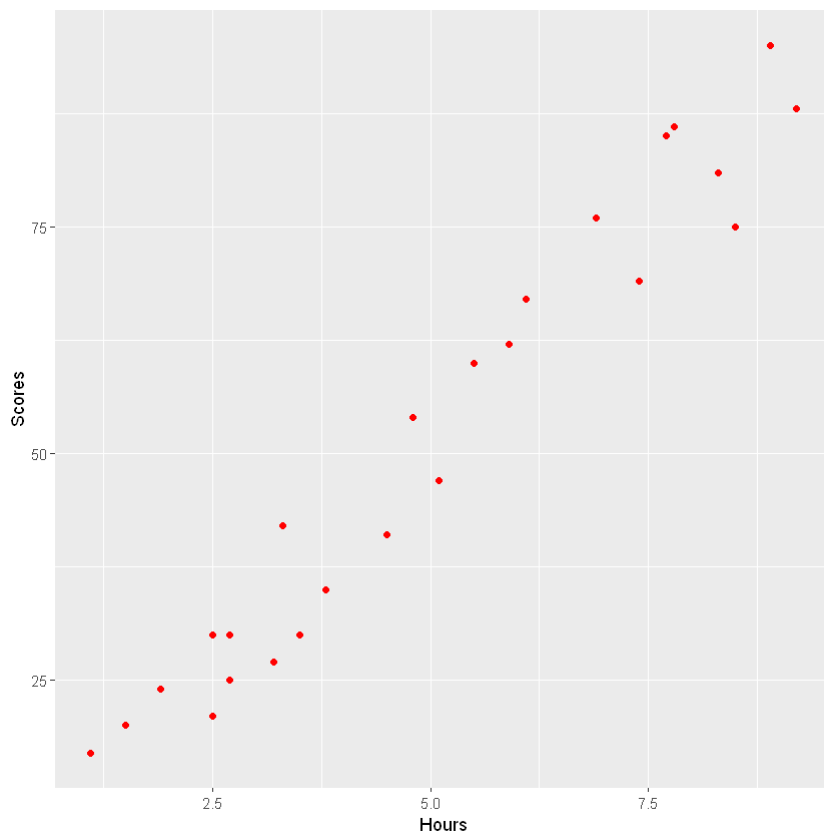
```
is.null(df)
```

FALSE

Plotting of data

In [4]:

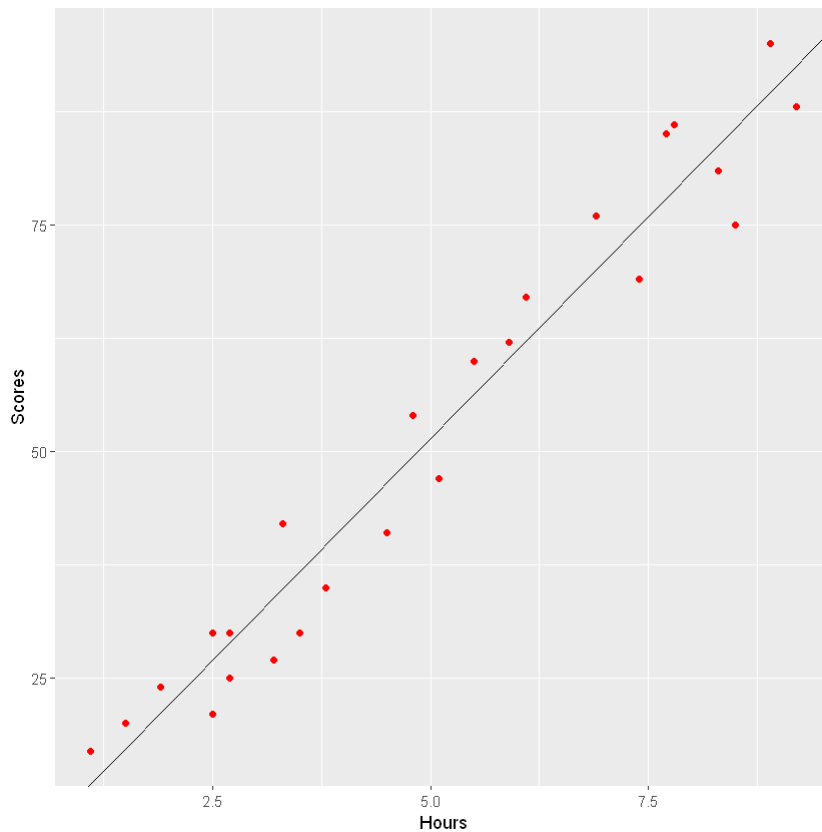
```
ggplot(df,aes(Hours,Scores))+geom_point(col='red',fill='red')
```



Regression line Plot

In [5]:

```
ggplot(df,aes(Hours,Scores))+geom_point(col='red',fill='red')+geom_abline(intercept = 2.484
```



Preparation of data

In [6]:

```
sample<-sample.split(df$Scores,SplitRatio = 0.7)  
train<-subset(df,sample==T)  
test<-subset(df,sample=F)
```

Training of Model

In [7]:

```
model<-lm(Scores~Hours,train)
```

Testing of Model

In [8]:

```
predictions<-predict(model,test)
results<-cbind(predictions,test$Scores)
colnames(results)<-c('predicted','actual')
results<-as.data.frame(results)
head(results)
```

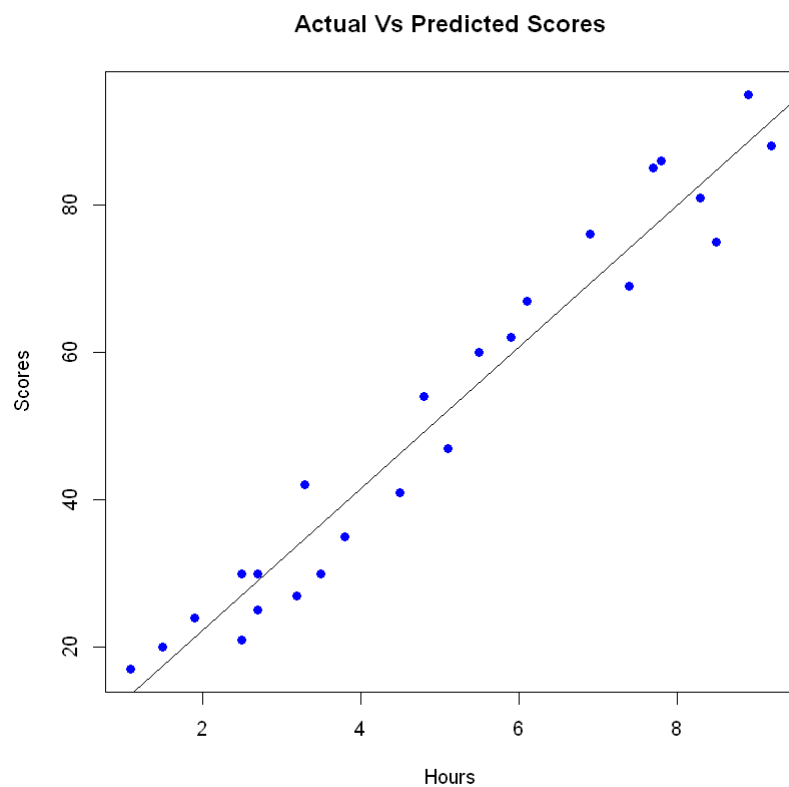
A data.frame: 6 × 2

	predicted	actual
	<dbl>	<dbl>
1	27.10340	21
2	52.06354	47
3	33.82344	27
4	84.70371	75
5	36.70345	30
6	17.50335	20

Evaluation of model

In [9]:

```
plot(test$Hours,test$Scores,pch=16,col='blue',xlab='Hours',ylab='Scores',main='Actual Vs Pr  
abline(model)
```



In [10]:

```
min(results)
```

13.6633270568807

In [11]:

```
mse<-mean((results$actual-results$predicted)^2)
mse
```

29.1400613142702

In [12]:

```
rmse<-mse^0.5
rmse
```

5.3981535096985

In [13]:

```
SSE<-sum((results$predicted-results$actual)^2)
SST<-sum((mean(df$Scores)-results$actual)^2)
R2<-1-SSE/SST
R2
```

0.952528988673659

Prediction

Hours studies per day = 9.25

Predicted marks =

In [14]:

```
Hours<-9.25
Hours<-as.data.frame(Hours)
ans<-predict(model,newdata = Hours)
ans
```

1: 91.9037537885272

Thank You! Have a good day.