

The Spark Foundation Task 1

RISHAB JAIN

06/01/2021

PROBLEM STATEMENT :

What will be the predicted score if studies for 9.25hr/day

Solution : We will be using linear regression model for this prediction.. At first we will the load the library

```
library(tidyverse)
```

```
## -- Attaching packages ----- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2      v purrr   0.3.4
## v tibble  3.0.3      v dplyr  1.0.2
## v tidyr   1.1.1      v stringr 1.4.0
## v readr   1.3.1      v forcats 0.5.0
```

```
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
```

```
library(carat)
```

Now, we will load the dataset from the url given

```
library(readr)
scores <- read_csv("https://raw.githubusercontent.com/AdiPersonalWorks/Random/master/student_scores%20-%20.csv")
scores
```

```
##      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
## 7      9.2      88
## 8      5.5      60
## 9      8.3      81
## 10     2.7      25
## 11     7.7      85
```

```
## 12  5.9    62
## 13  4.5    41
## 14  3.3    42
## 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
## 25  7.8    86
```

#Now we will train the model

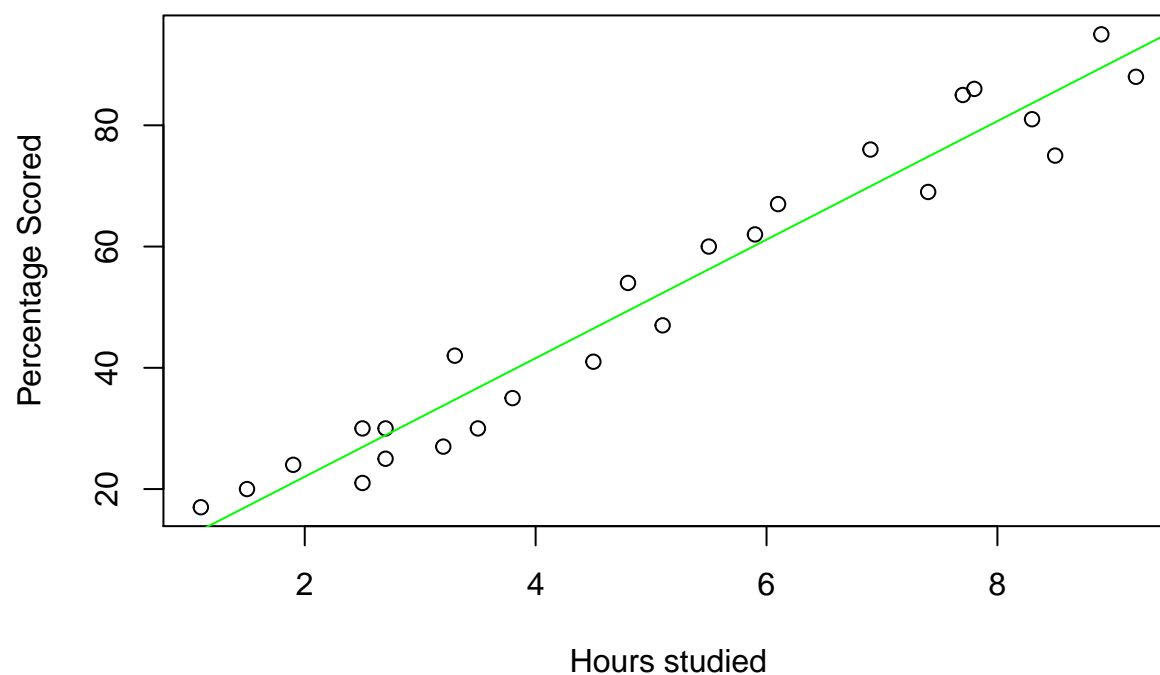
```
hour <- scores$Hours
scr <- scores$Scores
model <- lm(scr~hour)
model
```

```
##
## Call:
## lm(formula = scr ~ hour)
##
## Coefficients:
## (Intercept)      hour
##      2.484      9.776
```

#Plotting the regression line

```
attach(scores)
plot(Hours,Scores,type = "p",xlab = "Hours studied",ylab = "Percentage Scored",main = "Hours vs Percent
abline(model,col = "green")
```

Hours vs Percentage



#Comparing the actual and predicted scores

```
actual = scr
hour = data.frame(hour)
predicted <- predict(model, hour)
compare = data.frame(actual, predicted)
compare
```

```
##      actual predicted
## 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
```

Now after training our model we can solve the question given in our task that what will be the score if it studies for 9.25/days

```
test = data.frame(hour = 9.25)
test2 = predict(model, test)
test2
```

```
##          1
## 92.90985
```

From this we come into conclusion that the prediction score for studying 9.25hrs/day is around 92.9