R Notebook

Simple Linear Regression

Importing the dataset

```
dataset = read.csv('Salary_Data.csv')
```

Splitting the dataset into the Training set and Test set

install.packages('caTools')

```
library(caTools)
set.seed(123)
split = sample.split(dataset$Salary, SplitRatio = 2/3)
training_set = subset(dataset, split == TRUE)
test_set = subset(dataset, split == FALSE)
```

Fitting Simple Linear Regression to the Training set

```
regressor = lm(formula = Salary ~ YearsExperience, data = training_set)
summary(regressor)
```

```
##
## Call:
## lm(formula = Salary ~ YearsExperience, data = training_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7325.1 -3814.4   427.7  3559.7  8884.6
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)        25592       2646   9.672 1.49e-08 ***
## YearsExperience     9365        421  22.245 1.52e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 5391 on 18 degrees of freedom
## Multiple R-squared:  0.9649, Adjusted R-squared:  0.963
## F-statistic: 494.8 on 1 and 18 DF,  p-value: 1.524e-14
```

Predicting the Test set Results

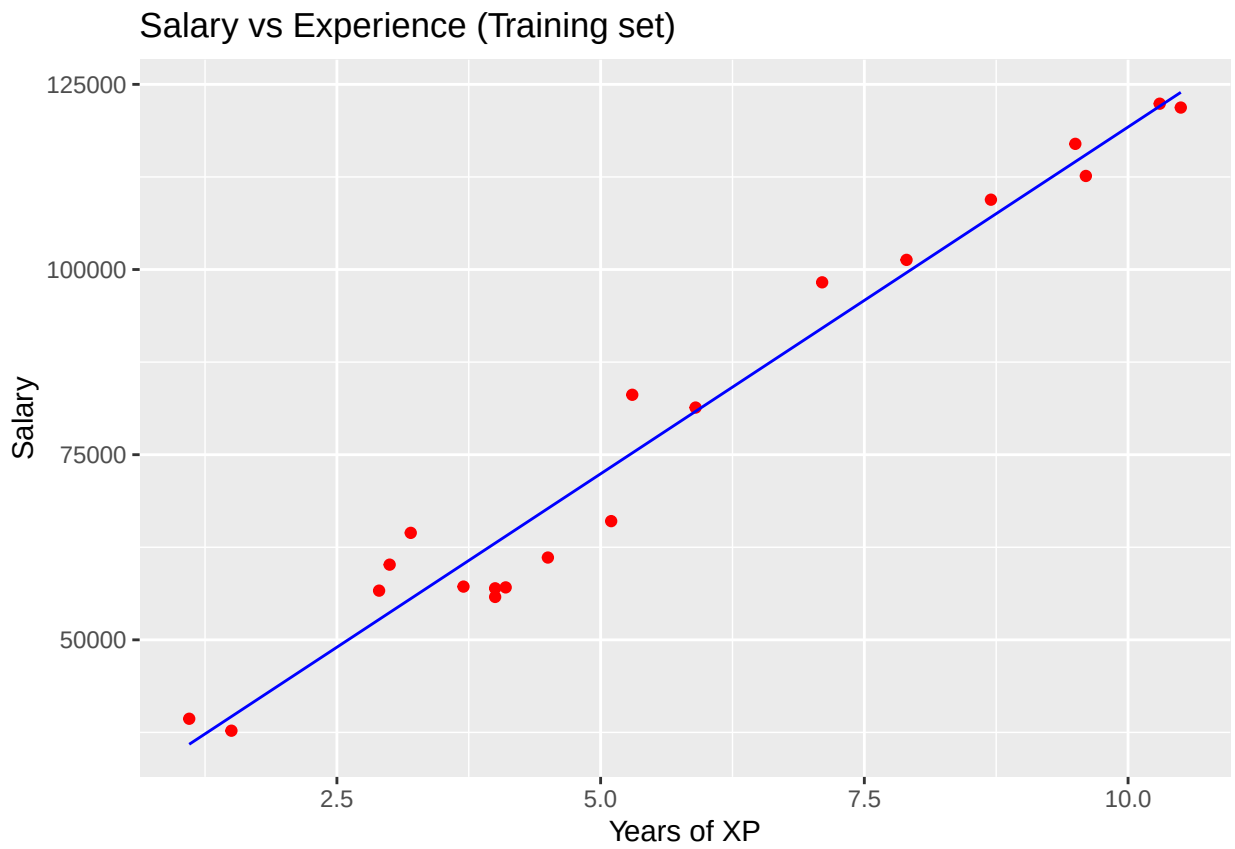
```
y_pred = predict(regressor, newdata = test_set)
y_pred #predicted values in test set
```

```
##         2         4         5         8        11        16        20        21
##  37766.77  44322.33  46195.35  55560.43  62115.99  71481.07  81782.66  89274.72
##        24        26
## 102385.84 109877.90
```

Visualising the Training set results

```
 # install.packages('ggplot2')

library(ggplot2)
ggplot() +
  geom_point(aes(x = training_set$YearsExperience, y = training_set$Salary),
             colour = 'red') +
  geom_line(aes(x = training_set$YearsExperience, y = predict(regressor, newdata = training_set)),
            colour = 'blue') +
  ggtitle('Salary vs Experience (Training set)') +
  xlab('Years of XP') +
  ylab('Salary')
```

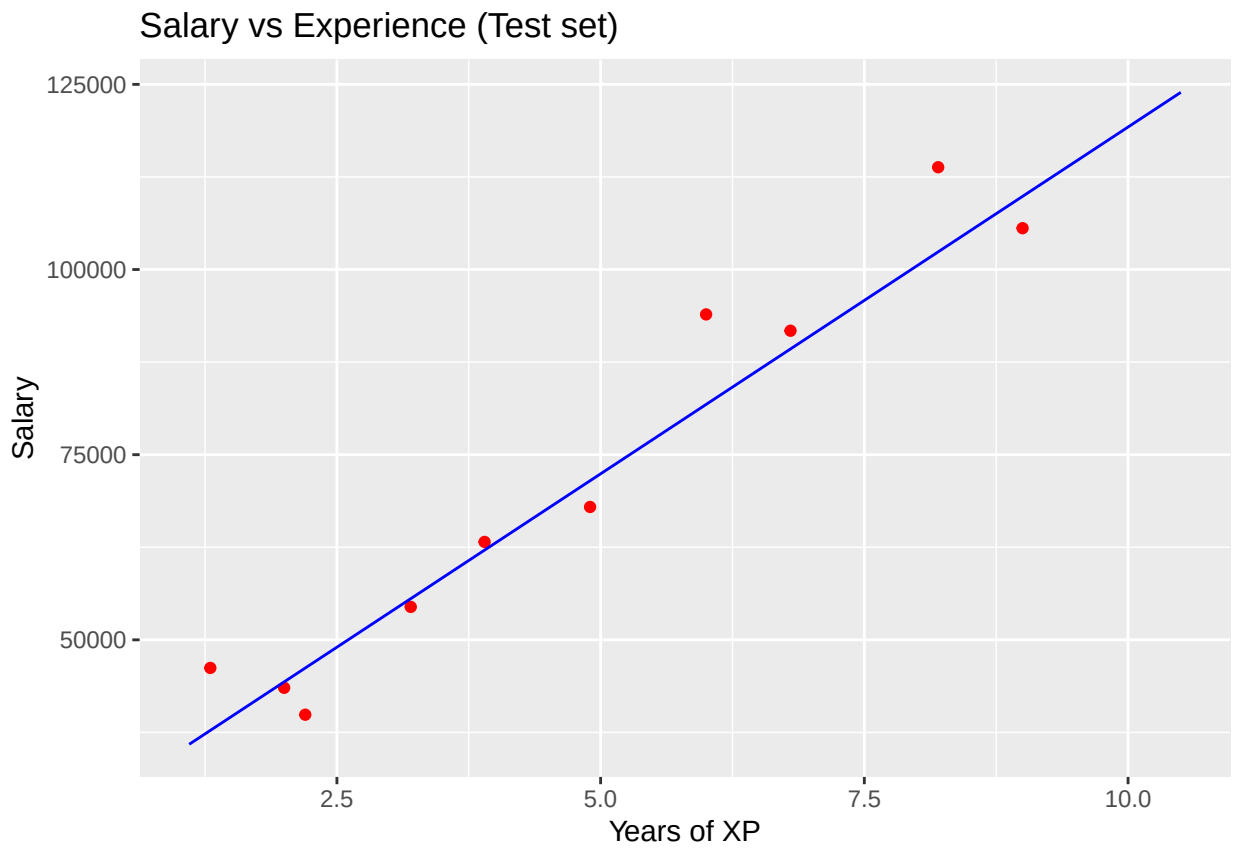


Visualising the Test set results. The same line above in the Test set.

```
# install.packages('ggplot2')

library(ggplot2)

ggplot() +
  geom_point(aes(x = test_set$YearsExperience, y = test_set$Salary),
             colour = 'red') +
  geom_line(aes(x = training_set$YearsExperience, y = predict(regressor, newdata = training_set)),
            colour = 'blue') +
  ggtitle('Salary vs Experience (Test set)') +
  xlab('Years of XP') +
  ylab('Salary')
```


Salary vs Experience (Test set)
125000
100000
75000
50000
Salary
2.5
5.0
7.5
10.0
Years of XP