

# Simple Linear Regression

## Importing the libraries

In [0]:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
```

## Importing the dataset

In [0]:

```
dataset = pd.read_csv('Salary_Data.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
```

## Splitting the dataset into the Training set and Test set

In [0]:

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 1/3, random_state = 0)
```

## Training the Simple Linear Regression model on the Training set

In [0]:

```
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
```

Out[0]:

```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

## Predicting the Test set results

In [0]:

```
y_pred = regressor.predict(X_test)
```

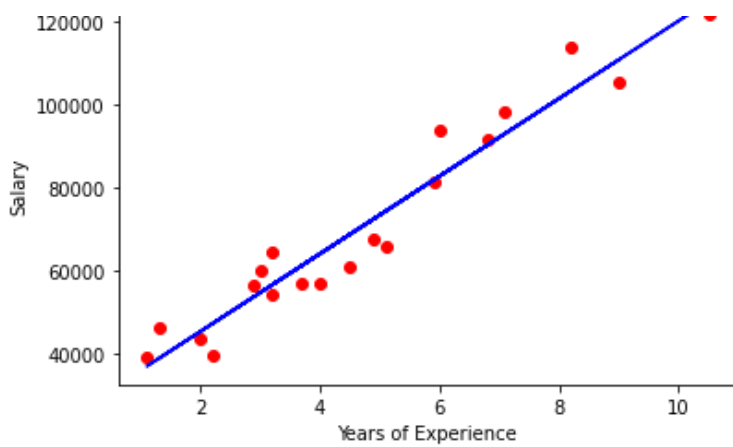
## Visualising the Training set results

In [0]:

```
plt.scatter(X_train, y_train, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title('Salary vs Experience (Training set)')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.show()
```

Salary vs Experience (Training set)





## Visualising the Test set results

In [0]:

```
plt.scatter(X_test, y_test, color = 'red')
plt.plot(X_train, regressor.predict(X_train), color = 'blue')
plt.title('Salary vs Experience (Test set)')
plt.xlabel('Years of Experience')
plt.ylabel('Salary')
plt.show()
```



## Making a single prediction (for example the salary of an employee with 12 years of experience)

In [0]:

```
print(regressor.predict([[12]]))
```

[138967.5015615]

Therefore, our model predicts that the salary of an employee with 12 years of experience is \$ 138967,5.

**Important note:** Notice that the value of the feature (12 years) was input in a double pair of square brackets. That's because the "predict" method always expects a 2D array as the format of its inputs. And putting 12 into a double pair of square brackets makes the input exactly a 2D array. Simply put:

12  
→ scalar

[12]  
→ 1D array

[[12]]  
→ 2D array

## Getting the final linear regression equation with the values of the coefficients

In [0]:

```
print(regressor.coef_)  
print(regressor.intercept_)
```

```
[9345.94244312]  
26816.192244031183
```

Therefore, the equation of our simple linear regression model is:

$$\begin{aligned} \text{Salary} \\ = 9345.94 \\ \times \text{YearsExperience} + 26816.19 \end{aligned}$$

**Important Note:** To get these coefficients we called the "coef " and "intercept" attributes from our regressor object. Attributes in Python are different than methods and usually return a simple value or an array of values.