Simple Linear Regression

Importing the Libraries

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

Importing the DataSet

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

Splitting the dataset into training set and Test set

```
from sklearn.model_selection import train_test_split
x_train , x_test , y_train , y_test = train_test_split(x , y , test_size= 0.2 , random_state = 0)
```

Training the Simple linear regression model on the training set

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

```
LinearRegression
LinearRegression()
```

Predicting the test set Result

```
ypredict = regressor.predict(x_test)
```

Visualizing the Training Set Result

```
plt.scatter(x_train , y_train , color = "red")
plt.plot(x_train , regressor.predict(x_train))
plt.title("Salary vs Experience (Training Set)")
plt.xlabel("Years of Experience ")
plt.ylabel("Salary")
plt.show()
```


```
from google.colab import drive
drive.mount('/content/drive')
```

Visualizing the Test Set

```
plt.scatter(x_test , y_test , color = "red")
plt.plot(x_train , regressor.predict(x_train))
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)

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

```
[138531.00067138]
```

Therefore, our model predicts that the salary of an employee

1. List item
2. List item

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:

$12 \rightarrow \text{scalar}$

$[12] \rightarrow \text{1D array}$

$[[12]] \rightarrow \text{2D array}$

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

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

```
[9312.57512673]
26780.09915062818
```

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

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

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.