

Simple Linear Regression

Importing the Libraries

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

Importing the dataset

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

	YearsExperience	Salary
0	1.1	39343.0
1	1.3	46205.0
2	1.5	37731.0
3	2.0	43525.0
4	2.2	39891.0

Splitting the dataset into Train 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()
```

Predicting the Test set results

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

Visualising the Training set results

- We will use the matplotlib library. To be exact pyplot for plotting our training set results.

```
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('EXPERIENCE (Years)')
plt.ylabel('SALARY (USD)')
```

Text(0, 0.5, 'SALARY (USD)')



Visualising the Test set results

- We will use the matplotlib library. To be exact pyplot for plotting our training set results.

```
plt.scatter(X_test, y_test, color='green')
plt.plot(X_train, regressor.predict(X_train), color='blue')
plt.title('SALARY vs EXPERIENCE (Test Set)')
plt.xlabel('EXPERIENCE (Years)')
plt.ylabel('SALARY (USD)')
```

Text(0, 0.5, 'SALARY (USD)')



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 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.

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

```
[9312.57512673]  
26780.09915062818
```

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

Salary=9345.94xYearsOfExperience+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.