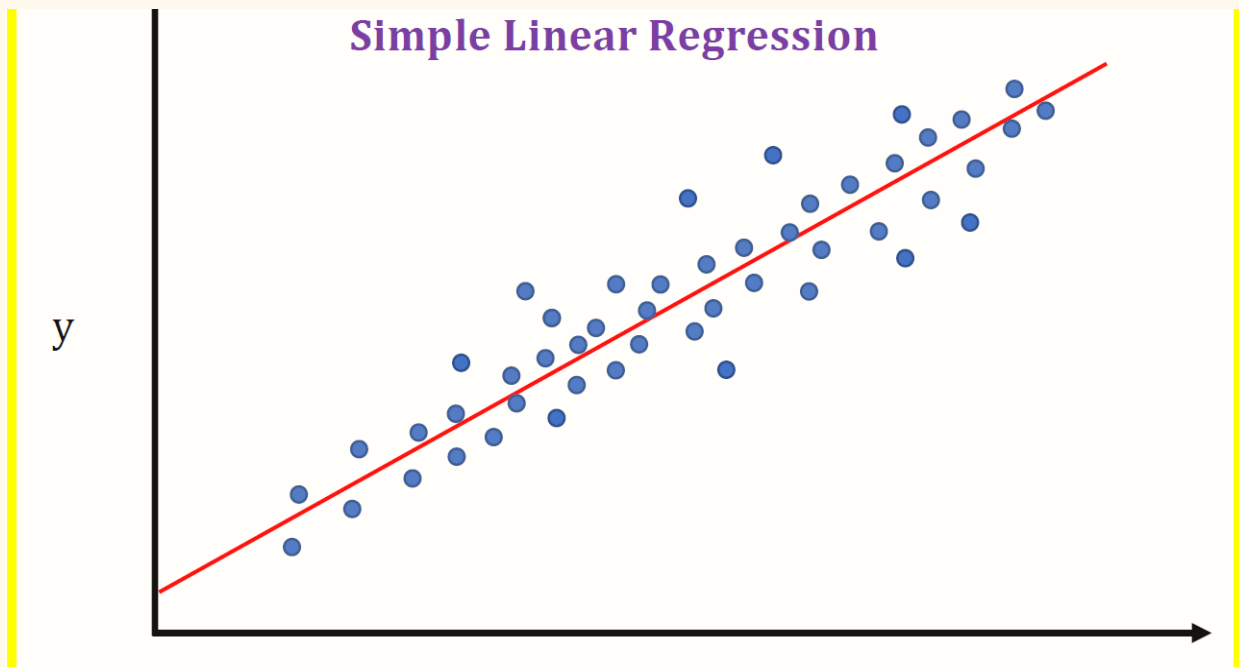# Linear Regression algorithm

(Using Salary as a case study)

—

By Ogunlaiye Sope

INTRODUCTION

Linear regression is a method of predictive analysis used to find out the value of the dependent variable with the help of an independent variable. The goal of linear regression is to find the best-fit line that describes the relationship between the dependent and the independent variable.

Linear regression algorithm

Today, we will learn how linear regression analysis works with the help of a very simple example. We will predict the salary of a worker based on the individual's years of experience. To

predict this we are going to use a salary dataset that has a column titled YearsExperience and another titled Salary. Then, you will use a Linear Regression algorithm to find the best-fit line that describes the relationship between the number of Years Experience and Salary. Once you have the best-fit line, you can use it to predict how much Salary will be paid.

Implementing linear regression using python

We will be using Jupyter notebook for this project. First of all we will import all necessary libraries for this project. These are the libraries I use for most of my data science projects.

```python
In [1]:    1  import pandas as pd
           2  import numpy as np
           3  import datetime
           4  from datetime import date, timedelta
           5  import plotly.graph_objects as go
           6  import plotly.express as px
           7  import plotly.io as pio
           8  from plotly.subplots import make_subplots
           9  pio.renderers.default='notebook'
          10  pio.templates.default = "plotly_white"
          11  import ipywidgets as widgets
          12  from sklearn.linear_model import LinearRegression
```

Next we import the data i.e. the salary csv file and also define the dependent and independent variable

```python
1  #Importing the data
2  dataset = pd.read_csv('Salary_Data (1).csv')
3  x = dataset.iloc[:, :-1].values
4  y = dataset.iloc[:, -1].values
```

The dataset variable imports the Salary_Data CSV into jupyter notebooks via pandas. The x variable identifies the independent variable 'YearsExperience' while they variable identifies the dependent variable 'Salary'.

Next we view the relationship between dependent and independent variables with a plotly visualization. The next page shows a screenshot of the syntax and resulting visualization indicating the relationship between the two variables.

scatter_plot = px.scatter(data_frame = dataset,

x = 'YearsExperience',

y = 'Salary',

trendline = 'ols')

scatter_plot.show('notebook')

```
In [5]:   1  #the relationship between salary and years of experience
          2  scatter_plot = px.scatter(data_frame = dataset,
          3                            x = 'YearsExperience',
          4                            y = 'Salary',
          5                            trendline = 'ols')
          6  scatter_plot.show('notebook')
```
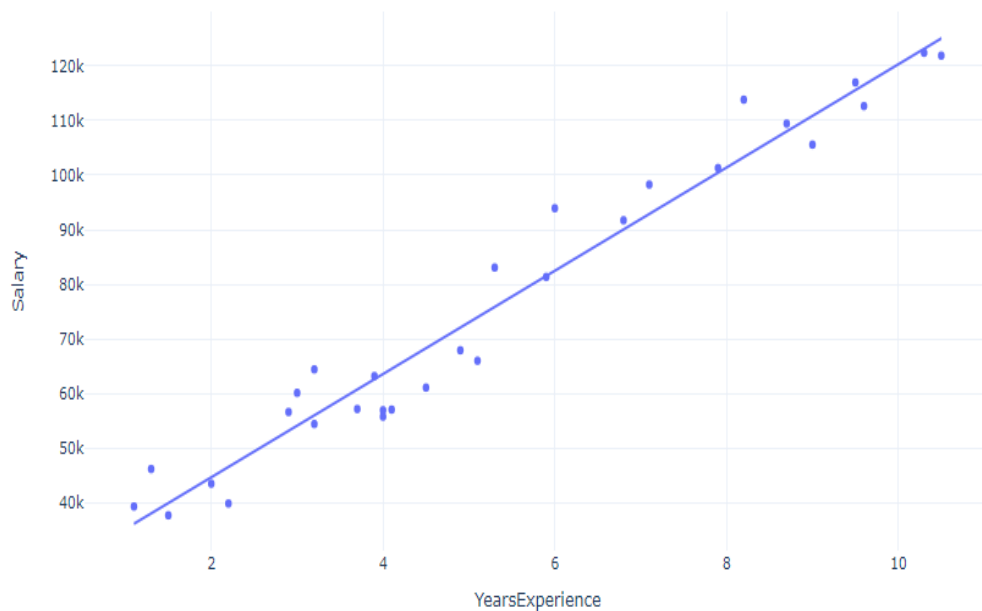

Next, we'll create a linear regression model using machine learning as follows

x = dataset.iloc[:, :-1].values

y = dataset.iloc[:, -1].values

model = LinearRegression().fit(x, y)

```
:   1 #create a linear regression model
    2 x = dataset.iloc[:, :-1].values
    3 y = dataset.iloc[:, -1].values
    4 model = LinearRegression().fit(x, y)
```

Finally, we can predict the value of a dependent variable using the value of an independent variable using our Machine Learning model:

x = [[20]]

y = model.predict(x)

print("Predicted salary:", y[0])

```
1 #predict the salary based on the number of years
2 x = [[20]]
3 y = model.predict(x)
4 print("Predicted salary:", y[0])

Predicted salary: 214791.4466277702
```

We predicted the salary of an individual who has worked for 20yrs and the predicted salary was 214791.466277702 (₦214,791.46) based on the number of years and the history of salaries provided.

Hope you enjoyed this!