Regression Analysis

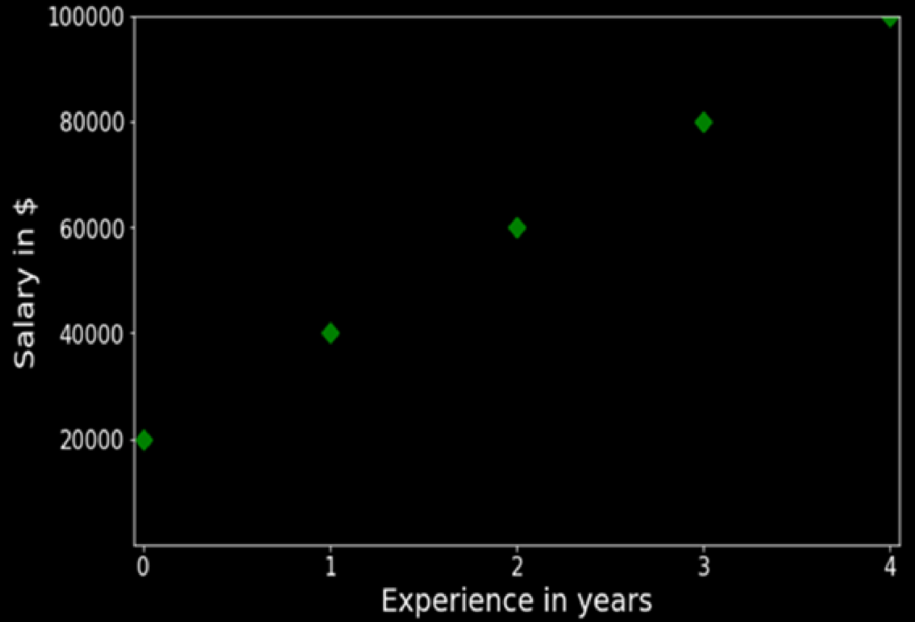
Linear Regression

Linear Regression is a process where there exists a linear relationship between independent variable $\boxed{X}$ and dependent variable $\boxed{y}$

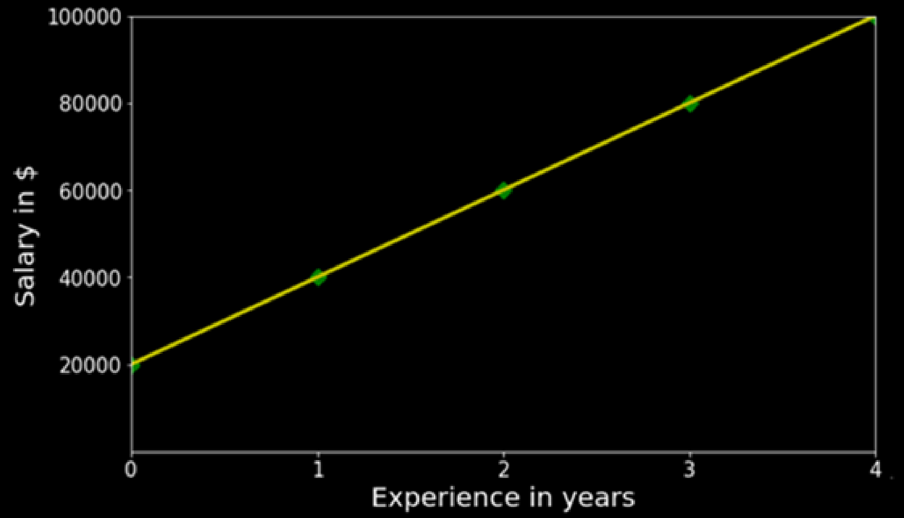
$$\boxed{y = f(X)}$$

Where

X = Independent Variable or features
y = Dependent variable or target.



Salary in $	Experience in Years
20000	0
40000	1
60000	2
80000	3
100000	4



For the Salary dataset

$$\boxed{Salary = f(Experience)}$$

$$y = mX + b$$

$$y = wX + w_0$$

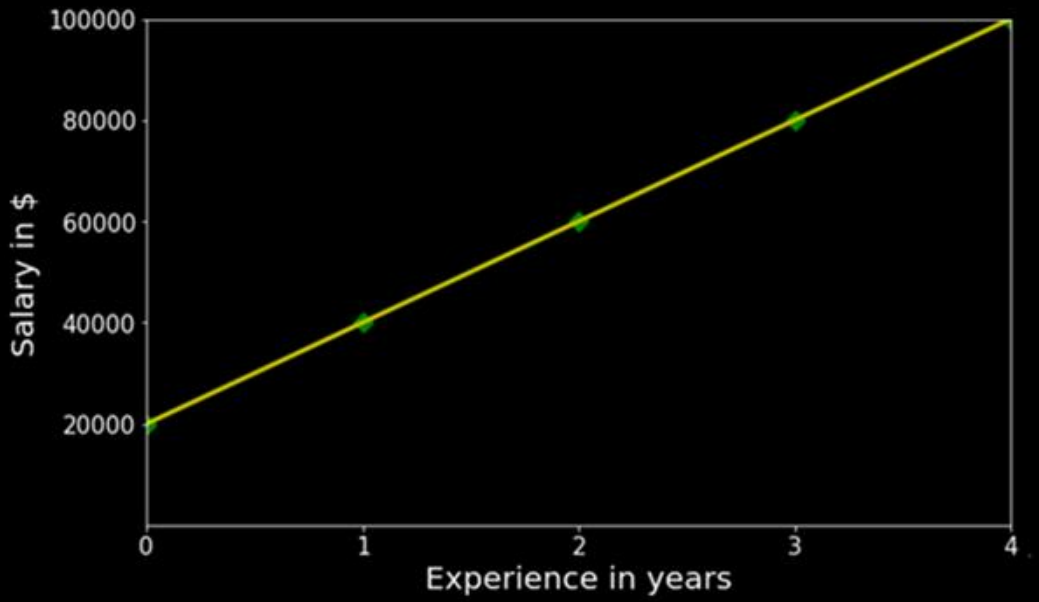
$$\boxed{y = w_0 + wX}$$ → Standard Regression Equation

Where

$$\boxed{\begin{aligned} w_0 &= \text{Bias.} \\ w &= \text{weight associated with } X \end{aligned}}$$

Simple Linear Regression

Simple Linear Regression



Salary in $	Experience in Years
20000	0
40000	1
60000	2
80000	3
100000	4

Salary Dataset

$$\boxed{y = w_0 + wX}$$

Where

w_0 = Bias.
w = weight associated with X

Multiple Linear Regression

Multiple Linear Regression

Salary in $	Experience in Years X_1	Years of Education X_2
20000	0	16
40000	1	16
60000	2	16
100000	3	18
140000	4	18

$$\boxed{y = w_0 + w_1 X_1 + w_2 X_2}$$

Where

w_0 = Bias.
w_1 = weight associated with X_1
w_2 = weight associated with X_2

Evaluation Metrics For Regression

Evaluation Metrics

- Root Mean Square Error
- Mean Absolute Error
- R^2

Root Mean Square Error

$$RMSE = \boxed{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y - y_{pred})^2}}$$

RMSE tells us how far apart the predicted values are from the observed values in a dataset, on average. The lower the RMSE, the better a model fits a dataset

Mean Absolute Error

$$\boxed{MAE = \frac{1}{n}\sum_{i=1}^{n} |y - y_{pred}|}$$

R^2

$$\boxed{R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i=1}^{n}(y - y_{pred})^2}{\sum_{i=1}^{n}(y - y_{mean})^2}}$$

This value ranges from 0 to 1.

The higher the R2 value, the better a model fits a dataset.

R^2 will be negative if the model prediction is bad i.e the sum of square error is large as compare to sum of total error.

Thank You!

Thank You!