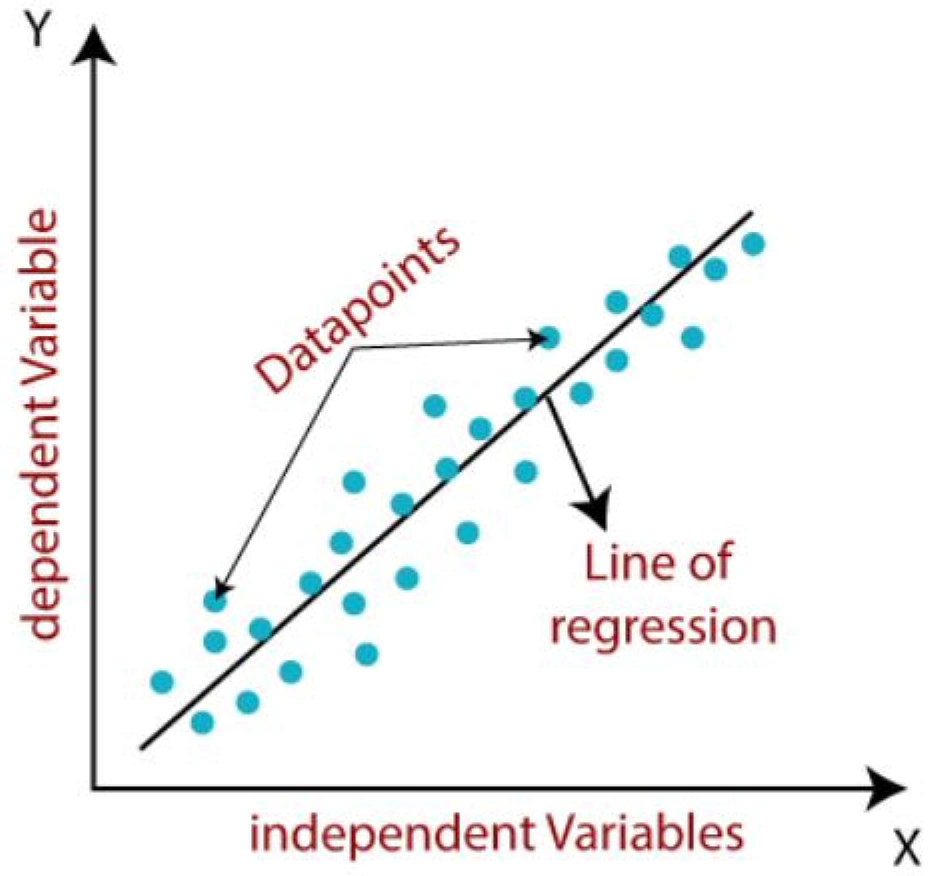
Chapter 2: Simple Linear Regression

Linear Regression:

- It is a statistical method for predictive analysis.
- It can make prediction for continuous/real or numeric variables such as price, salary, etc.
- Linear regression is used to find out how the value of dependent variable is changing according to the value of independent variable.



Types of relationship:

- Regression line can show two types of relationship:
    - **Positive Linear Relationship:**
        - In this the value of y increases as the value of x increases.
    - **Negative Linear Relationship:**
        - In this the value of y decreases as the value of x increases.

Simple Linear Regression:

- Simple Linear Regression models the relationship between a dependent variable and a single independent variable.
- SLR has two objectives:
    - Model the relationship between the two variables:
        - House area and house price
        - Experience and salary
    - Forecasting new observations:
        - Predict price as per area
        - Predict salary as per experience

P1

Area	Price
2.6	5.5
3	5.65
3.2	6.1
3.6	6.8
4	7.2

$\sum x = 2.6 + 3 + 3.2 + 3.6 + 4 =$

$\sum y = 5.5 + 5.65 + 6.1 + 6.8 + 7.2 =$

$\sum x2 = 2.6^2 + 3^2 + 3.2^2 + 3.6^2 + 4^2 =$

$\sum y^2 = 5.5^2 + 5.65^2 + 6.1^2 + 6.8^2 + 7.2^2 =$

$\sum xy = (2.6x5.5) + (3x5.65) + (3.2x6.1) + (3.6x6.8) + (4x7.2) =$

$\beta_1 = (n (\sum xy) - (\sum x)(\sum y)) \ / \ (n \sum x2 - (\sum x)^2)$

$\beta_1 =$

$\beta_1 =$

$\beta_1 =$

$\beta_1 =$

$x' = 16.4 / 5 =$

$y' = 31.25 / 5 =$

$\beta_0 = y' - \beta_1 x'$

$\beta_0 =$

$\beta_0 =$

$y = \beta_0 + \beta_1 x$

$y =$

Model Evaluation:

- It is necessary to obtain the accuracy on training data
- But it is also important to get a genuine and approximate result on unseen data otherwise Model is of no use.
- So to build and deploy a generalized model we require to Evaluate the model on different metrics which helps us to better optimize the performance.

Model Score:

- **score() / r2_score()** provides an indication of goodness of fit and therefore a measure of how well unseen samples are likely to be predicted.
- It is also called **R^2 score**, the coefficient of determination.
- Best possible score is 1.0 and worst would be some –ve value.

Score Calculation:

Score =	**1 – u / v**
u =	**residual sum of squares** **((y_true - y_pred) ** 2).sum()**
v =	**total sum of squares** **((y_true - y_true.mean()) ** 2).sum()**

P1

Score Calculation:

$$y = 2.02 + 1.28 * x$$

Area X	Price y_true
3.6	6.8
3.2	6.1

Price y_pred

(y_true - y_pred)^2
u =

(y_true - y_true.mean())^2
v =

Score:

P2

Experience	Salary
1	10
3	21
4	45
5	41
7	50
8	72
10	90
11	95
12	120
15	150

$\sum x$ = **1 + 3 + 4+ 5 + 7 + 8 + 10 + 11 + 12 + 15 =**

$\sum y$ = **10 + 21 + 45 + 41 + 50 + 72 + 90 + 95 + 120 + 150 =**

$\sum x2$ = $\mathbf{1^2 + 3^2 + 4^2 + 5^2 + 7^2 + 8^2 + 10^2 + 11^2 + 12^2 + 15^2 =}$

$\sum y2$ = $\mathbf{10^2 + 21^2 + 45^2 + 41^2 + 50^2 + 72^2 + 90^2 + 95^2 + 120^2 + 150^2 =}$

$\sum xy$ = **(1x10) + (3x21) + (4x45)+ (5x41) + (7x50) + (8x72) + (10x90) + (11x95) + (12x120) + (15x150) =**

$\beta_1 = (n (\sum xy) - (\sum x)(\sum y)) \; / \; (n \sum x2 - (\sum x)^2)$

$\beta_1 =$

$\beta_1 =$

$\beta_1 =$

x' =

y' =

$\boldsymbol{\beta_0 = y' - \beta_1 x'}$

$\boldsymbol{\beta_0 =}$

$\boldsymbol{\beta_0 =}$

$\boldsymbol{y = \beta_0 + \beta_1 x}$

y =

P2

Score Calculation:

$$y = -1.22 + 9.19 * x$$

Exp X	Sal y_true
15	150
8	72
3	21

Pred. Sal y_pred

$(y\_true - y\_pred)^2$
u =

$(y\_true - y\_true.mean())^2$
v =

Score: