Experiment No: 3

Aim: Implement simple linear regression using analytical and machine learning without using SkLearn.

Theory:

- A simple linear regression is a type of regression algorithms that models the relationship between a dependent variable and one independent variable.
- The relationship shown by simple Linear Regression model is Linear or a sloped straight line, hence it is called Simple Linear Regression.
- The key point in simple linear regression is that dependent variable must be continuous / real value.
- However, independent variable can be measured on continuous or categorical values.
- The regression line is represented using the following equation,

$$y' = aX + b + e$$

- In the above equation y' represents the predicted value, a represents the slope of the line, b shows the Y intercept and e is the random error.

- Here, we assume the mean value of random error is 0, so the equation it becomes,

$$Y' = aX + b$$

- The value of a and b can be calculated as -

$$a = \frac{n \cdot \Sigma XY - \Sigma X \cdot \Sigma Y}{n\Sigma X^2 - (\Sigma X)^2}$$

$$b = \frac{1}{n}(\Sigma Y - a\Sigma X)$$

- For e.g., consider the set of data as $\{(-1,-1), (2,2), (3,2)\}$

	X	Y	XY	X^2
	-1	-1	+1	1
	2	2	4	4
	3	2	6	9
Total	6	3	11	14

- The equation for the regression line is,

$$Y' = aX + b$$

$$\therefore a = \frac{n.\Sigma XY - \Sigma X.\Sigma Y}{n.\Sigma X^2 - (\Sigma X)^2}$$

$$= \frac{3 \times 11 - 4 \times 3}{3 \times 14 - 16}$$

$$\therefore a = 0.807$$

$$\therefore b = \frac{1}{n}(\Sigma Y - a \times \Sigma X)$$

$$= \frac{1}{3}(3 - 0.807 \times 4)$$

$$= -0.076$$

- Now, the equation for the line becomes,

$$Y' = 0.807X - 0.076$$

A
27/4/22