G H Raisoni College of Engineering
SY AI Semester-IV AY 2023-24 Division-A
UCAIP210: Machine Learning Algorithms Practicals
Lab Manual
Practical Teacher : Dr Monika Y. Dangore

Experiment No: 1

Aim:

Generate a proper dataset of N points. Perform Linear Regression Analysis with Least Squares Method and plot the graph.

Introduction to Least Square Regression:

Least Square Regression is a statistical method commonly used in machine learning for analyzing and modelling data. It involves finding the line of best fit that minimizes the sum of the squared residuals (the difference between the actual values and the predicted values) between the independent variable(s) and the dependent variable.

We can use Least Square Regression for both simple linear regression, where there is only one independent variable. Also, for multiple linear regression, where there are several independent variables. We widely use this method in a variety of fields, such as economics, engineering, and finance, to model and predict relationships between variables

Finding the Line of Best Fit Using Least Square Regression

Given any collection of a pair of numbers and the corresponding scatter graph, the line of best fit is the straight line that you can draw through the scatter points to represent the relationship between them best.

$$Y = mX + c$$

Where,

Y: Dependent Variable

m: Slope

X: Independent Variable

c: y-intercept

Our aim here is to calculate the values of slope y-intercept and substitute them in the equation along with the values of independent variable X to determine the values of dependent variable Y. Let's assume that we have 'n' data points, then we can calculate slope using the scary looking formula below:

$$m = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}$$

Then, the y-intercept is calculated using the formula:

$$c = \bar{y} - m * \bar{x}$$

Then we substitute these values in the final equation Y = mX + c.

Least Squares Regression Example

Let us take a simple dataset to demonstrate the least squares regression method.

X	Y
Years of Expeience (Years)	Salary (in 1000$)
18	90
12	64
2	15
8	47
16	75
11	61
1	8
9	49
5	25

Step 1: The first step is to calculate the slope 'm' using the formula

$$m = \frac{\sum(x - \bar{x})(y - \bar{y})}{\sum(x - \bar{x})^2}$$

After substituting the respective values in the formula, m = 4.70 approximately.

Step 2: Next, calculate the y-intercept 'c' using the formula (y_{mean} — m * x_{mean}). By doing that, the value of c approximately is c = 5.3

$$c = \bar{y} - m * \bar{x}$$

Step 3: Now we have all the information needed for the equation, and by substituting the respective values in Y = mX + c, we get the following table. Using this information, you can now plot the graph.

X	Y (Actual)	Y (Predicted)	Error (Residual)
Years of Expeience (Years)	Salary (in 1000$)		
18	90	91.27	-1.27
12	64	63.07	0.93
2	15	16.07	-1.07
8	47	44.27	2.73
16	75	81.87	-6.87
11	61	58.37	2.63
1	8	11.37	-3.37
9	49	48.97	0.03
5	25	30.17	-5.17

This way, the least squares regression method provides the closest relationship between the dependent and independent variables by minimizing the distance between the residuals (or error) and the trend line (or line of best fit). Therefore, the sum of squares of residuals (or error) is minimal under this approach.

Program:

Attach the printouts of the program.

Result:

The concept of Least Square Regression is studied and program is executed successfully.