Experiment No.: 03

Title: Write a program to implement Multiple Linear Regression

Objectives: 1. To learn **Multiple linear regression**

Theory:

Most of time output Y can not be predicted by single independent variable but needs multiple Independent variables.
The Regression that has one output variable and more than one input/independent variables with Linear relationship between input and output is called as **multiple linear regression**.

Example:
y = h(x)
= w0 + w1x1 + w2x2 + w3x3 + ….
Prediction of house price based on size of house, age of house, distance from the center of city, etc.

Coefficient

The coefficient is a factor that describes the relationship with an unknown variable.

Example: if x is a variable, then 2x is x two times. x is the unknown variable, and the number 2 is the coefficient.

e.g.

1. House price prediction based on size of the house, number of rooms in the house, Number of floors, Age of the building, open space around the building.
   Here:
   y = House Price
   X_1 = Size of the House
   X_2 = Number of Rooms
   X_3 = Number of floors
   X_4 = Age of the building
   X_5 = open space

Sr. No.	X_1	X_2	X_3	X_4	X_5	y
0	X_1^0	X_2^0	X_3^0	X_4^0	X_5^0	y^0
1	X_1^1	X_2^1	X_3^1	X_4^1	X_5^1	y^1
2	X_1^2	X_2^2	X_3^2	X_4^2	X_5^2	y^2
3	X_1^3	X_2^3	X_3^3	X_4^3	X_5^3	y^3
4	X_1^4	X_2^4	X_3^4	X_4^4	X_5^4	y^4

Response or Target variable $\hat{y}$ is defined as

$$\hat{y} = h(x) = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_nx_n$$

Where

$X_1, X_2, X_3, \ldots, X_n$ are input/independent/predictor variables
$\hat{y}$ is the output variable.
$W_0, W_1, W_2, \ldots, W_n$ are parameters or coefficients of regression.

Since there is possibility of difference between actual output value and Predicted value, we can write actual output as

$$y = \hat{y} + e = w_0 + w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_nx_n + e$$
$$e = y - w_0 + w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_nx_n$$
$$= y - \hat{y}$$
if e is negative, e = $\hat{y}$ - y

Parameter Estimation in Multiple Linear Regression:

Gradient Descent Algorithm is used to estimate parameters in Multiple Linear Regression

The Cost function is:

$$J(W) = \frac{1}{2n} \sum_{i=1}^{n} (h(x^i) - y^i)^2$$

X^i ith input in the dataset

y^i ith output in the dataset

Basic Gradient Descent Algorithm:

Repeat Until Converge
{

$$w_{new} = w_{old} - \eta \frac{\partial J(w)}{\partial w}$$

}

Algorithm-

1] Import all necessary libraries.
2] Read data set into pandas dataframe
3] Create linear Regression object
4] Train model using fit function
5] Use build model for prediction