SHRI VILEPARLE KELAVANI MANDAL'S
DWARKADAS J. SANGHVI COLLEGE OF ENGINEERING
(Autonomous College Affiliated to the University of Mumbai)
NAAC ACCREDITED with "A" GRADE (CGPA : 3.18)

DEPARTMENT OF INFORMATION TECHNOLOGY

COURSE CODE: DJ19TEL7014   **DATE:**

COURSE NAME: Machine Learning   **CLASS:** Final Year B.Tech

EXPERIMENT NO. 2

CO Measured:

CO1 Solve real-world problems using suitable machine learning techniques.

TITLE: Model-building using regression

AIM / OBJECTIVE:

To perform linear regression and find the error associated with the model.

DESCRIPTION OF EXPERIMENT:

Linear regression is one of the easiest and most popular Supervised Machine Learning algorithms. It is a statistical method that is used for predictive analysis. Linear regression makes predictions for continuous/real or numeric variables such as sales, salary, age, product price, etc. Linear regression algorithm shows a linear relationship between a dependent (y) and one or more independent (x) variables, hence called as linear regression. Since linear regression shows the linear relationship, which means it finds how the value of the dependent variable is changing according to the value of the independent variable.The linear regression model provides a sloped straight line representing the relationship between the variables.Cleaning Data in PythonWe will now separate the numeric columns from the categorical columns.

Mathematically, we can represent a linear regression as: $y = b_0 + b_1 x + \varepsilon$
Here,
    y= Dependent Variable (Target Variable)
    x= Independent Variable (predictor Variable)
    b_0= intercept of the line (Gives an additional degree of freedom)
    b_1 = Linear regression coefficient (scale factor to each input value).
    ε = random error
The values for x and y variables are training datasets for Linear Regression model representation

The different values for weights or coefficient of lines (b_0, b_1) gives the different line of regression, and the cost function is used to estimate the values of the coefficient for the best fit line. Cost function optimizes the regression coefficients or weights. It measures how a linear regression model is performing. We can use the cost function to find the accuracy of the **mapping function**, which maps the input variable to the output variable. This mapping function is also known as **Hypothesis function**. For Linear Regression, we use the **Mean Squared Error (MSE)** cost function, which is the average of squared error occurred between the predicted values and actual values. It can be written as:

$$MSE = \frac{1}{N} \sum_{i=1}^{n} (y_i - (b_1 x_i + b_0))^2$$

where,
N=Total number of observation
y_i = Actual value
($b_1 x_i + b_0$)= Predicted value.

Linear regression using Least Square Method

We have linear regression equation as **y= b₀+b₁x**

Using least square method,

$$b_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

PROCEDURE:

1. Describe the procedure that is used to perform Linear regression using Least Square Method carry out the experiment step-by-step for simple linear regression for following dataset without using scikit library. Describe every line of code with the proper interpretation of the output.

X	2	3	4	5	6	7	8	9	10
Y	1	3	6	9	11	13	15	17	20

2. Perform Regression with respect to one dataset of your choice and discuss results of all the steps.

OBSERVATIONS / DISCUSSION OF RESULT:

1. Find predicted value of y using Linear Regression for one epoch and RMSE for x =4.

X	2	3	4	5	6	7	8	9	10
Y	1	3	6	9	10	13	14	17	21

CONCLUSION:

Base all conclusions on your actual results; describe the meaning of the experiment and the implications of your results.

REFERENCES:

(List the references as per format given below and citations to be included the document)

[1] Ponniah P., "Data Warehousing: Fundamentals for IT Professionals", 2nd Edition, Wiley India, 2013.

[2] Ageed, Z. S., Zeebaree, S. R., Sadeeq, M. M., Kak, S. F., Yahia, H. S., Mahmood, M. R., & Ibrahim, I. M. (2021), "Comprehensive survey of big data mining approaches in cloud systems", Qubahan Academic Journal, 1(2), 29-38.

Website References:

Author's Last Name, First Initial. Middle Initial. (Date of Publication or Update). Title of work. Site name. Retrieved Month Day, Year, from URL from Homepage

[3] U.S. Census Bureau. U.S. and world population clock. U.S. Department of Commerce. Retrieved July 3, 2019, from https://www.census.gov/popclock.