**2SVKM’s NMIMS**

**Mukesh Patel School of Technology Management & Engineering**

Program: B.Tech CSDS(311)/B.Tech-CSBS

**Course: Machine Learning**

**Experiment No.03**

PART A

(PART A : TO BE REFFERED BY STUDENTS)

**A.1 Aim:** Implementation of Simple Linear Regression

**A.2 Prerequisite:**

Python Programming

**A.3 Outcome:**

**After successful completion of this experiment students will be able to:**

 To understand the concept of simple linear regression

 To apply simple linear regression on actual dataset to do prediction

**A.4 Theory:**

Types of Regression:

 Linear Regression

 Logistic Regression

 Polynomial Regression

Linear Regression It is one of the most widely known modeling technique. Linear regression is usually among the first few topics which people pick while learning predictive modeling. In this technique, the dependent variable is continuous, independent variable(s) can be continuous or discrete, and nature of regression line is linear.

Linear Regression establishes a relationship between dependent variable (Y) and one or more independent variables (X) using a best fit straight line (also known as regression line). It is represented by an equation Y=a+b\*X + e, where a is intercept, b is slope of the line and e is error term. This equation can be used to predict the value of target variable based on given predictor variable(s).

It is used to estimate real values (cost of houses, number of calls, total sales etc.) based on continuous variable(s). Here, we establish relationship between independent and dependent variables by fitting a best line.

This best fit line is known as regression line and represented by a linear equation Y= a \*X + b. The best way to understand linear regression is to relive this experience of childhood. Let us say, you ask a child in fifth grade to arrange people in his class by increasing order of weight, without asking them their weights! What do you think the child will do? He / she would likely look (visually analyze) at the height and build of people and arrange them using a combination of these visible parameters. This is linear regression in real life! The child has actually figured out that height and build would be correlated to the weight by a relationship, which looks like the equation above.

In this equation:  Y – Dependent Variable  a – Slope  X – Independent variable  b – Intercept

These coefficients a and b are derived based on minimizing the sum of squared difference of distance between data points and regression line. Look at the below example.

**Programing Steps for Simple Linear Regression:**

1. Consider the given dataset ie. HousePrice.csv file (CSV Comma Separated Values) (or any other CSV file)

2. Read CSV data into pandas dataframe object

3. Choose independent variable (X) and dependent variable (Y) from given dataset

4. Find the bo and b1 values to get Ypredicted

5. After getting Ypred, calculate the SSE (sum of squared error)

6. Calculate the RMSE (Root Mean Square Error) value

7. Calculate the coefficient of determination (r 2 ) r-square

8. Plot regression line along with the given data points

**Task1: Using python perform the following**

1. Create a .csv file, with two columns

|  |  |
| --- | --- |
| Area | Price |
| 2600 | 550000 |
| 3000 | 565000 |
| 3200 | 610000 |
| 3600 | 680000 |
| 4000 | 725000 |

2.Find out the prices of homes whose area is, 3800 and 5000.

3.Retrospect the value of slope and intercept for the simple linear regression

4. Find MSE and R-squared error

5. Draw the regression line

**Task2: Using python perform the following**

**1.Perform multilinear regression on the housing-price dataset**

**2.Identufy dependent and independent variable**

**3.Find out which attribute is contributing more on the dependent variable**

**4.Mention the training and testing criteria used in this program**

**5.Find the slope and coefficients for the housing-price dataset**

**6.Draw the regression line for the same.**

PART B

(PART B : TO BE COMPLETED BY STUDENTS)

***(Students must submit the soft copy as per following segments within two hours of the practical.)***

|  |  |
| --- | --- |
| Roll No. | Name: |
| Class : | Batch : |
| Date of Experiment: | Date of Submission |
| Grade : |  |

**B.1 Task1**

**B.2 Task2**

**B.2 Conclusion:**

**Question of Curiosity:**

1. Consider the following set of points: {(-2, -1) , (1 , 1) , (3 , 2)}  
   a) Find the least square regression line for the given data points.

*(Students must write the conclusion in their own words.)*