Marwadi University	Marwadi University Faculty of Technology Department of Information and Communication Technology	
Subject: Artificial Intelligence (01CT0616)	Aim: To obtain the best fit line over single feature scattered datapoints using Linear Regression	
Experiment No: 02	Date:	Enrollment No: 92200133030

Aim: To obtain the best fit line over single feature scattered datapoints using Linear Regression

IDE: Google Colab

Theory:

Linear regression is a method for determining the best linear relationship between two variables X and Y. If variables X and Y are uncorrelated, it is pointless embarking upon linear regression. However, if a reasonable degree of correlation exists between X and Y then linear regression may be a useful means to describe the relationship between the two variables. The usual approach is to use the *least-squares* method, which minimizes the squared difference between the actual data points and a straight line. Let $[x_i, y_i]$, $i = 1,2,3,\dots,N$ be the N pairs of data values of the variables X and Y. The straight-line relating X and Y is $y = mx + c$, where m and c are the gradient and constant values (to be determined) defining the straight line. Thus, $y(x_i) - y_i$ is the difference between the line and data point i (see Fig. 1). Taking all the data points, we seek values of m and c that minimize the squared difference SD.

$$\sum_{1}^{N} [y(x_i) - y_i]^2$$

This is achieved by calculating the partial derivatives of SD with respect to m and c and finding the pair $[m,c]$ such that SD is at a minimum.

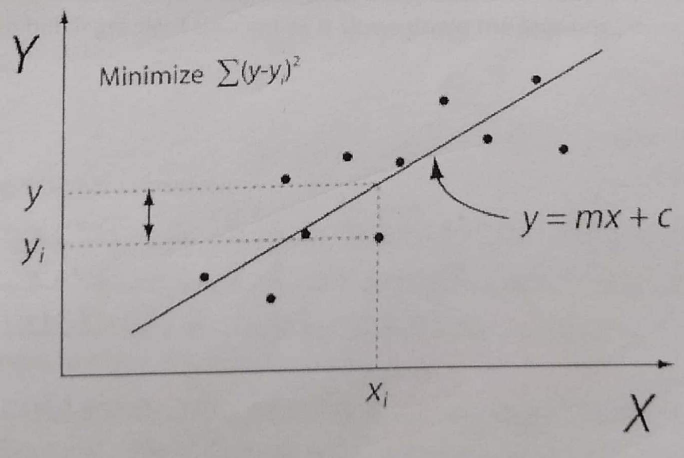

Figure 1: Illustration of Linear Regression. Linear least squares regression, the idea is to find the line y = mx + c that minimizes the mean squared difference between the line and the data points

Batch Gradient Descent:

Gradient Descent is an optimization algorithm used for minimizing the cost function in various machine learning algorithms. It is basically used for updating the parameters of the learning model. Batch gradient descent which processes all the training examples for each iteration of gradient descent. But if the number of training examples is large, then batch gradient descent is computationally very expensive.

Let m be the number of training examples. Let n be the number of features.

Algorithm for batch gradient descent :

Let $h_\theta(x)$ be the hypothesis for linear regression. Then, the cost function is given by:
Let Σ represents the sum of all training examples from i=1 to m.
$J_{train}(\theta) = (1/2m)\ \Sigma(h_\theta(x^{(i)}) - y^{(i)})^2$

Repeat {
$\theta j = \theta j - (\text{learning rate}/m) * \Sigma(h_\theta(x^{(i)}) - y^{(i)})x_j^{(i)}$
For every j =0 ...n
}

Where $x_j^{(i)}$ Represents the j^{th} feature of the i^{th} training example. So if *m* is very large(e.g. 5 million training samples), then it takes hours or even days to converge to the global minimum.That's why for large datasets, it is not recommended to use batch gradient descent as it slows down the learning.

Pre Lab Exercise:

a. Explain the meaning of linear regression

Linear Regression is a statistical method used to model the relationship between one or more indepen dent variable and dependent variable.

b. Write three applications of linear regression

The applications of linear regression extens to
1) House Price Prediction 2) Stock Market trends
3) Sales Forcasting etc.

Marwadi University	Marwadi University Faculty of Technology Department of Information and Communication Technology	
Subject: Artificial Intelligence (01CT0616)	Aim: To obtain the best fit line over single feature scattered datapoints using Linear Regression	
Experiment No: 02	Date:	Enrollment No: 92200133030

c. Write three advantages of linear regression

The advantages of linear regressions are 1) Simple and interpretability 2) Efficiency 3) Low computational cost etc.

d. Write three limitations of linear regression

The Limitations of Linear Regressions are 1) Assume Linearity 2) sensitive to outliers and 3) Colinearity Issues etc.

Methodology:

1. Load the basic libraries and packages
2. Load the dataset
3. Analyse the dataset
4. Pre-process the data
5. Visualize the Data
6. Separate the feature and prediction value columns
7. Write the Hypothesis Function
8. Write the Cost Function
9. Write the Gradient Descent optimization algorithm
10. Apply the training over the dataset to minimize the loss
11. Find the best fit line to the given dataset
12. Observe the cost function vs iterations learning curve

Program (Code):

To be attached with

10

Marwadi University	Marwadi University Faculty of Technology Department of Information and Communication Technology	
Subject: Artificial Intelligence (01CT0616)	Aim: To obtain the best fit line over single feature scattered datapoints using Linear Regression	
Experiment No: 02	Date:	Enrollment No: 92200133030

Observation and Result Analysis:

a. Nature of the dataset

The dataset is having 2 colums x and y having numerical values. the columns having max values 3530 and 108 and min values 0 and -3.83 accordingly.

b. During Training Process

The model uses training data to learn the relationship between features and the target variables. The Mean Squared Error is minimized by adjusting the model parameters.

c. After the training Process

The model is tested on unseen data to check its accuracy. uses metrices like R^2 score, MSE and RMSE to assess the model. The model's ability to perform on unseen dataset is analyzed.

Post Lab Exercise:

a. What are the major assumptions considered in linear regression

The Major Assumptions considered in linear regressions are 1) Linearity 2) Homoscedasticity 3) Normality of Residuals 4) No Multicollinearity

b. Why MSE is used instead of MAE for calculating the loss function

In the loss function calculations, MSE is preferred in the linear regression as it is differentiable making it easier to optimize.

c. How can the behaviour of outliers be understood while dealing with the unseen dataset

we can identify the outliers by plotting scatter plot or by calculating the IQR ranges or plotting box plot.

Marwadi University	Marwadi University Faculty of Technology Department of Information and Communication Technology	
Subject: Artificial Intelligence (01CT0616)	Aim: To obtain the best fit line over single feature scattered datapoints using Linear Regression	
Experiment No: 02	Date:	Enrollment No: 92200133030

d. Derive the Normal Equation for the Linear Regression.

→ In the Linear Regression :-

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{n} (y_i - (\theta_0 + \theta_1 x_i))^2$$

and the Hypothesis function :-

$$h(\theta) = \theta_0 x_0 + \theta_1 x_1 + \dots + \theta_n x_n$$

so, $h(\theta) = \theta^T x$

and cost = [Predicted value - Actual value]^2

$$J(\theta) = \begin{bmatrix} h\theta(x^0) \\ h\theta(x^1) \\ \vdots \\ h\theta(x^2) \end{bmatrix} - \begin{bmatrix} y^0 \\ y^1 \\ \vdots \\ y^m \end{bmatrix}$$

$$= \begin{bmatrix} \theta^T(x_0) \\ \theta^T(x_1) \\ \vdots \\ \theta^T(x_m) \end{bmatrix} - y \quad . \; X\theta - y$$

cost function $= (X\theta - y)^T (X\theta - y) = J(\theta)$

now, $\frac{\partial J(\theta)}{\partial \theta} = \frac{\partial}{\partial \theta} [(X\theta - y)^T (X\theta - y)]$

$$2X^T(X\theta - y)$$

now, $2X^T(X\theta - y) = 0$

$$2X^T X\theta - 2X^T y = 0$$

$$X^T X \theta = X^T y$$

$$\theta = (X^T X)^{-1}(X^T y)$$