	Marwadi University Faculty of Technology Department of Information and Communication Technology	
Subject: Machine Learning (01CT0519)	**Aim: To obtain the best fit line over multiple feature scattered datapoints using Linear Regression**	
Experiment No: 03	**Date: 03-09-2024**	**Enrollment No: 92200133030**

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

IDE: Google Colab

Theory:

Multiple linear regression (MLR), also known simply as multiple regression, is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The goal of multiple linear regression is to model the linear relationship between the explanatory (independent) variables and response (dependent) variables. In essence, multiple regression is the extension of ordinary least-squares (OLS) regression because it involves more than one explanatory variable.

Formula and Calculation of Multiple Linear Regression

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \epsilon$$

where, for $i = n$ observations:

y_i = dependent variable

x_i = explanatory variables

β_0 = y-intercept (constant term)

β_p = slope coefficients for each explanatory variable

ϵ = the model's error term (also known as the residuals)

Example of How to Use Multiple Linear Regression

As an example, an analyst may want to know how the movement of the market affects the price of ExxonMobil (XOM). In this case, their linear equation will have the value of the S&P 500 index as the independent variable, or predictor, and the price of XOM as the dependent variable.

In reality, multiple factors predict the outcome of an event. The price movement of ExxonMobil, for example, depends on more than just the performance of the overall market. Other predictors such as the price of oil, interest rates, and the price movement of oil <u>futures</u> can affect the price of XOM and stock prices of other oil companies. To understand a relationship in which more than two variables are present, multiple linear regression is used.

Multiple linear regression (MLR) is used to determine a mathematical relationship among several random variables. In other terms, MLR examines how multiple independent variables are related to one dependent variable. Once each of the independent factors has been determined to predict the dependent variable, the information on the multiple variables can be used to create an accurate prediction on the level of effect they have

	Marwadi University **Faculty of Technology** **Department of Information and Communication Technology**
Subject: Machine Learning (01CT0519)	**Aim: To obtain the best fit line over multiple feature scattered datapoints using Linear Regression**
Experiment No: 03	**Date: 03-09-2024**	**Enrollment No: 92200133030**

on the outcome variable. The model creates a relationship in the form of a straight line (linear) that best approximates all the individual data points.

Program (Code):

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

	Marwadi University
![Marwadi University logo]	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**
Subject: Machine Learning (01CT0519)	Aim: To obtain the best fit line over multiple feature scattered datapoints using Linear Regression
Experiment No: 03	**Date: 03-09-2024**	**Enrollment No: 92200133030**

Results:

To be attached with

    a. Datapoints scattering (without best fit line)

Multivariate Linear Regression
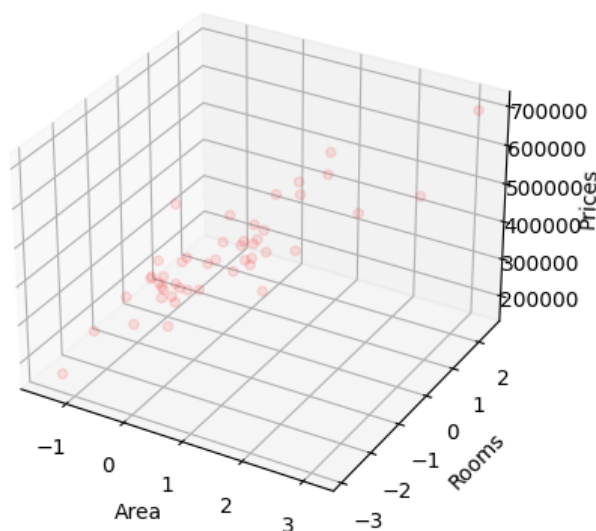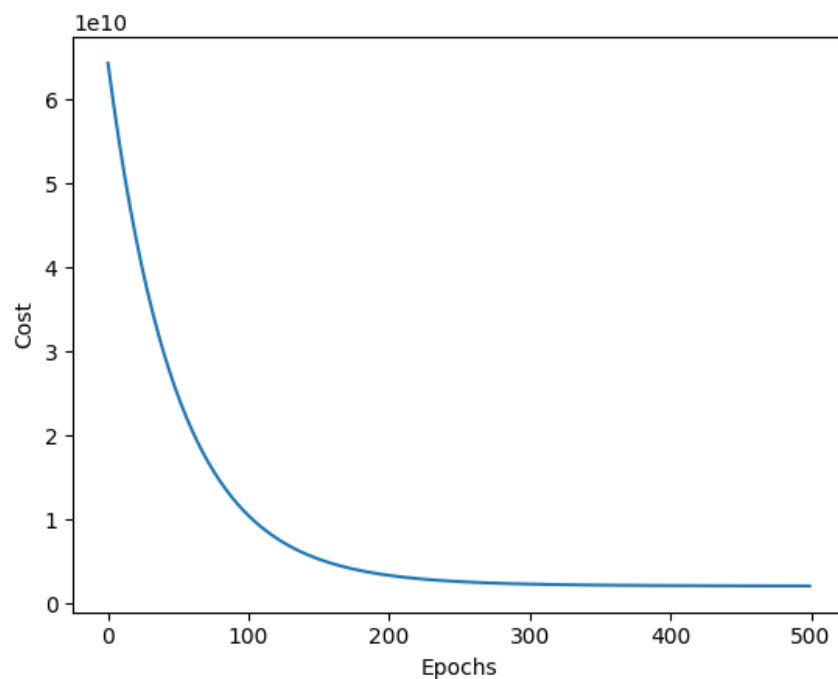
    b. Data Statistics before Normalization

	Area	Rooms	Prices
count	47.000000	47.000000	47.000000
mean	2000.680851	3.170213	340412.659574
std	794.702354	0.760982	125039.899586
min	852.000000	1.000000	169900.000000
25%	1432.000000	3.000000	249900.000000
50%	1888.000000	3.000000	299900.000000
75%	2269.000000	4.000000	384450.000000
max	4478.000000	5.000000	699900.000000

	Marwadi University
![Marwadi University logo]	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**
Subject: Machine Learning (01CT0519)	Aim: To obtain the best fit line over multiple feature scattered datapoints using Linear Regression
Experiment No: 03	**Date: 03-09-2024**	**Enrollment No: 92200133030**

c. Data Statistics after Normalization

	Area	Rooms	Price
count	4.700000e+01	4.700000e+01	47.000000
mean	9.448707e-18	2.710598e-16	340412.659574
std	1.010811e+00	1.010811e+00	125039.899586
min	-1.461049e+00	-2.882690e+00	169900.000000
25%	-7.233261e-01	-2.260934e-01	249900.000000
50%	-1.433229e-01	-2.260934e-01	299900.000000
75%	3.412850e-01	1.102205e+00	384450.000000
max	3.150993e+00	2.430504e+00	699900.000000

d. Learning Curve (Cost function vs iterations)

![Marwadi University logo]	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To obtain the best fit line over multiple feature scattered datapoints using Linear Regression**	
Experiment No: 03	**Date: 03-09-2024**	**Enrollment No: 92200133030**

Observation and Result Analysis:

a. Nature of the dataset

_____

_____

_____

_____

b. During Training Process

_____

_____

_____

_____

c. After the training Process

_____

_____

_____

_____

d. Observation over the Learning Curve

_____

_____

_____

![Marwadi University logo] Marwadi University	**Marwadi University** **Faculty of Technology** **Department of Information and Communication Technology**	
Subject: Machine Learning (01CT0519)	**Aim: To obtain the best fit line over multiple feature scattered datapoints using Linear Regression**	
Experiment No: 03	**Date: 03-09-2024**	**Enrollment No: 92200133030**

Post Lab Exercise:

a. What is the difference between single and multiple variable linear regression

b. What does it mean for a multiple linear regression to be linear?

c. What is the use of Normalization?

d. Is there any change in the behavior of data before and after normalization? Prove using a toy example.

	Marwadi University
	Faculty of Technology
	Department of Information and Communication Technology
Subject: Machine Learning (01CT0519)	**Aim: To obtain the best fit line over multiple feature scattered datapoints using Linear Regression**
Experiment No: 03	**Date: 03-09-2024**	**Enrollment No: 92200133030**

Post Lab Activity:

Consider any dataset from **https://archive.ics.uci.edu/ml/datasets** and perform the multiple variable linear regression analysis over the dataset and obtain the best fit line. Make sure that the dataset is not matching with your classmates. You can also select the dataset from other ML repositories with prior permission from your concerned subject faculty.