National University of Computer and Emerging Sciences



Lab Exercise12

CL461-Artificial Intelligence Lab

|  |  |
| --- | --- |
| Course Instructor | Dr. Mubashir Baig |
| Lab Instructor (s) | Samia Akhter  Nimra Abbas |
| Section | H |
| Semester | Spring 2022 |

Department of Computer Science

FAST-NU, Lahore, Pakistan

Table of Contents

[1 Objectives 3](#_Toc73535090)

[2 Task Distribution 3](#_Toc73535091)

[3 Exercise (50 Marks) 3](#_Toc73535092)

[3.1 Steps to perform 3](#_Toc73535093)

[4 Submission Instructions 4](#_Toc73535094)

# Objectives

After performing this lab, students shall be able to understand the following:

* Linear Regression

# Task Distribution

| **Total Time** | **170 Minutes** |
| --- | --- |
| Exercise | 160 Minutes |
| Online Submission | 10 Minutes |

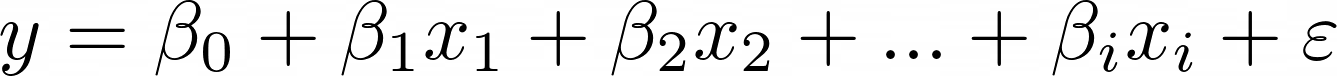
# Exercise (50 Marks)

Implement multiple linear regression using Ordinary Least Square method from scratch in Python. You are required to at least write your own the fit(), predict(), mean square method.

Use the following dataset: **Advertising.csv**

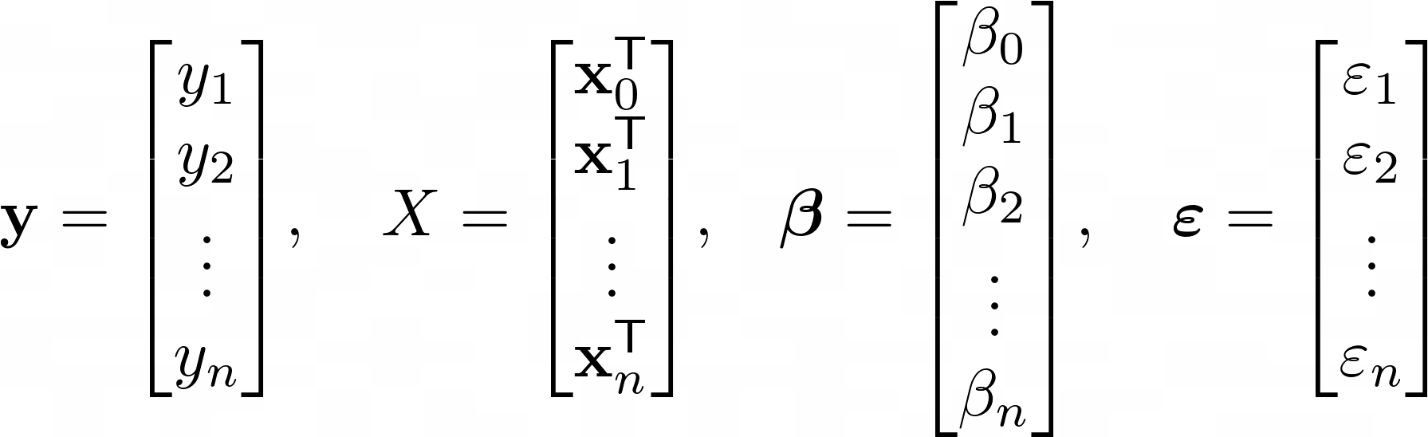
## Explanation

The general formula for multiple linear regression outcomes is as follows:

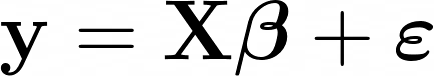


* β0 is known as the intercept.
* β0 to βi are known as coefficients.
* x1 to xi are the features of the data set.
* ε are the residual terms.

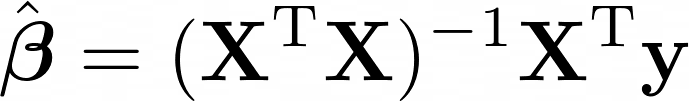
We can also represent the formula in vector notation.



The output equation would look like:



To find the coefficient values, we use the ordinary least square method which minimizes the sum of the square of the residuals. The formula for resulting coefficients in vector notation is:

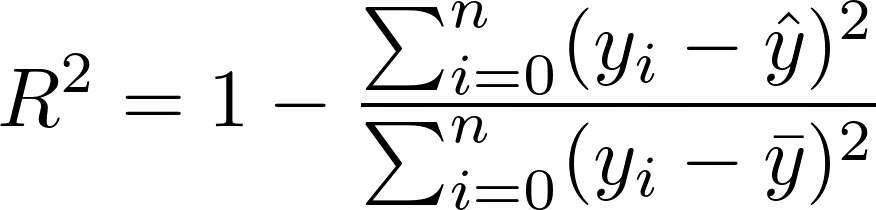


The derivation of this formula is out of the scope of this lab.If you want to learn, read more about it here: <https://web.stanford.edu/~mrosenfe/soc_meth_proj3/matrix_OLS_NYU_notes.pdf>

The outcome of the algorithm, beta β with a hat (^) on top, is a vector containing all of the coefficients that can be used to make predictions by using the formula presented in the beginning for multiple linear regression.

Assuming that you have a number of predictions for some observed data, how can you measure how well the model predicted the ground truth?

You can measure the accuracy of how well the multiple linear regression model performs by using a metric called r squared.



* yi is one value of y at index i.
* ŷ is pronounced as y hat and is the predicted values of y.
* y̅ is pronounced as y bar and is the average of y.

The metric measures the relationship between the residual sum of squares (RSS) and the total sum of squares (TSS). The RSS is computed as the ground truth minus the predicted ground truth, while the TSS is computed as the ground truth minus the average of the ground truth.

**Good Luck!☺**

# Submission Instructions

Always read the submission instructions carefully.

* Rename your Jupyter notebook to your roll number and download the notebook as **.ipynb** extension.
* To download the required file, go to **File->Download .ipynb**
* Only submit the **.ipynb** file. DO NOT **zip** or **rar** your submission file.
* Submit this file on Google Classroom under the relevant assignment.
* Late submissions will not be accepted.