

## Multiple Linear Regression

In the previous topic, we have learned about Simple Linear Regression, where a single Independent/Predictor(X) variable is used to model the response variable (Y). But there may be various cases in which the response variable is affected by more than one predictor variable; for such cases, the Multiple Linear Regression algorithm is used.

Moreover, Multiple Linear Regression is an extension of Simple Linear regression as it takes more than one predictor variable to predict the response variable. We can define it as:

Multiple Linear Regression is one of the important regression algorithms which *models the linear relationship between a single dependent continuous variable and more than one independent variable.*

Multiple regression is like linear regression, but with more than one independent value, meaning that we try to predict a value based on **two or more** variables.

### Example:

Prediction of CO<sub>2</sub> emission based on engine size and number of cylinders in a car.

### Some key points about MLR:

- For MLR, the dependent or target variable(Y) must be the continuous/real, but the predictor or independent variable may be of continuous or categorical form.
- Each feature variable must model the linear relationship with the dependent variable.

### MLR equation:

In Multiple Linear Regression, the target variable(Y) is a linear combination of multiple predictor variables  $x_1, x_2, x_3, \dots, x_n$ . Since it is an enhancement of Simple Linear Regression, so the same is applied for the multiple linear regression equation, the equation becomes:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Where,

**Y= Output/Response variable**

**$b_0, b_1, b_2, b_3, b_n, \dots$  = Coefficients of the model.**

**$x_1, x_2, x_3, x_4, \dots$  = Various Independent/feature variable**

### Implementation of Multiple Linear Regression model using Python:

To implement MLR using Python, we have below problem:

#### **Problem Description:**

We have a dataset of **50 start-up companies**. This dataset contains five main information: **R&D Spend, Administration Spend, Marketing Spend, State, and Profit for a financial year**. Our goal is to create a model that can easily determine which company has a maximum profit, and which is the most affecting factor for the profit of a company.

Since we need to find the Profit, so it is the dependent variable, and the other four variables are independent variables. Below are the main steps of deploying the MLR model:

1. **Data Pre-processing Steps**
2. **Fitting the MLR model to the training set**

3. **Predicting the result of the test set**