# To determine the weight of Brain using Multivariate Linear Regression

*Indrojyoti Mondal*[1], *Abhishek Mamidi*[2], Advisor: *Dr.Anish Chand Turlapty*[3]
Indian Institute of Information Technology, Chittoor, Sri City, A.P, India
*indrojyoti.m15@iiits.in*[1], *abhishek.m15@iiits.in*[2], *anish.turlapaty@iiits.in*[3]

**Abstract**

Regression analysis is a statistical technique for estimating the relationship among variables which have reason and result relation. Regression models with one dependent variable and more than one independent variables are called multivariate Linear regression. In this paper, data for multivariate Linear Regression Analysis is taken from R.J. Gladstone (1905). "A Study of the Relations of the Brain to to the Size of the Head", Biometrika, Vol. 4, pp105-123. A multivariate Linear regression model was fitted to this dataset. Average training and testing errors as a function of the model order M in the range of 1 to 6 is plotted and a brief report on the above analysis is presented.

**Index Terms**

Multivariate Linear Regression, Brainhead, Regression

## I. Introduction

ONE of the most used machine learning algorithm these days is Multivariate Linear Regression.In statistics, linear regression is a linear approach for modelling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. In Simple Linear Regression, there is only one independent variable and one dependent variable. We fit the model using Linear Regression. But, having only one independent variable is rarely possible. In real world problems, we have many independent variables which contribute to predict the dependent variable(y). So, we use Multivariate Linear regression.

Multivariate Linear Regression is used when we have more than one independent variable and one dependent variable. It is quite similar to the simple linear regression model, but with multiple independent variables contributing to the dependent variable and hence multiple coefficients to determine and complex computation due to the added variables. The equation of multivariate linear regression is given by:

$$Y_i = \alpha + \beta_1 x_i^{(1)} + \beta_2 x_i^{(2)} + ... + \beta_n x_i^{(n)} \tag{1}$$

$Y_i$ is the estimate of component of dependent variable y, where we have **n** independent variables and denotes the component of the independent variable. Similarly cost function is as follows,

$$E(\alpha, \beta_1, \beta_2, ..., \beta n) = \frac{1}{2m} \sum_{i=1}^{m} (y_i - Y_i) \tag{2}$$

where we have **m** data points in training data and y is the observed data of dependent variable.

## II. Description of Dataset

A selected Dataset is Brainhead dataset. The source of the dataset is R.J. Gladstone (1905). "A Study of the Relations of the Brain to to the Size of the Head", Biometrika, Vol. 4, pp105-123. In the dataset, Brain weight (grams) and head size (cubic cm) for 237 adults is classified by gender and age group. The features are Gender, Age Range, Head size, Brain weight. For Gender feature, value 1 is assigned for male and value 2 is assigned for Female. Similarly for Age feature, value 1 is assigned for a person having an age in the range of 20-46 and value 2 is assigned for a person whose age happens to be greater than 46. Head size is in $cm^3$. Brain weight is in $grams$.

| Gender | Age Range | Head size($cm^3$) | Brain weight($grams$) |
|---|---|---|---|
| 1 | 1 | 4512 | 1530 |
| 1 | 1 | 3738 | 1297 |
| 1 | 2 | 3524 | 1295 |
| 1 | 2 | 3571 | 1295 |
| 2 | 1 | 2857 | 1027 |

TABLE I: Sample Dataset

## III. Application of MVLR to the dataset

We considered the Brain weight as the output variable, and other variables such as Gender, Age range and head size as independent variables. Here, we predict the Brain weight of the person using the gender, age and head size. We chose MVLR because every input variable is independent of each other and we try to predict the Brain weight of that person, which is a dependent variable. The data is seperated into two sets(80:20), train and test sets. The train set was used to fit the model. The trained model was used to predict the brain weight on test set. We trained the model as a function of the model order M with different values ranging from 1 to 6. And the results were plotted.

## IV. Analysis of the results

We found the training and testing errors with M varying from 1 to 6.
The average training error is in the range of 56.5 to 58.5 for all the values of M in the range of 1 to 6.
The average testing error is in the range of 48.5 to 50.5 for all the values of M in the range of 1 to 6.
The training error is greater than the testing error and the testing error is least when M = 4. The figure given below gives the information about the training and testing errors w.r.t M.

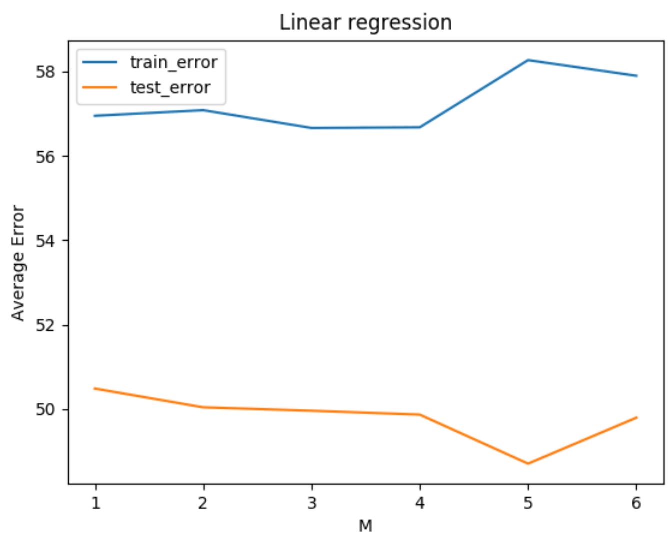


Fig. 1: Average error vs M