

Multivariate Linear Regression

Dataset

The dataset used for this activity is a multivariate dataset from UCI Machine Learning Repositories entitled "Automobile". The dataset will be used to identify the car prices according to its engine size using Multivariate Linear Regression. It consists of 205 instances with 25 features including fuel type, door number, engine location, car length, car width, engine type, engine size, horsepower, etc. Only few selected features were selected to be used in training the model namely: carwidth, carlength, carheight, enginesize, and horsepower. The values of these features were scaled from 0 to 1 and then divided into training and testing sets at an 8:2 ratio, with the training set being the 8.

Regression Result

Observing that the line which predicts the car prices adjust depending on the input feature used on the model, we can conclude that the model is working properly. It has a mean squared error value of 0.0076653067433068755 which is considered to be significantly low. This was because the dataset was scaled before fitted in the Linear Regression model. Visualizing a multivariate regression with more than 2 features seems to be complicated for a normal computer that can't show a multidimensional image. For that reason, each features were plotted separately in a subplot.