

COMP421X HW01: Multivariate Regression Report

Onur Yilmaz

March 2019

First of all, I divided the data set into two parts, first 160 as training and the rest as test, respectively. Then, I got the below figure.

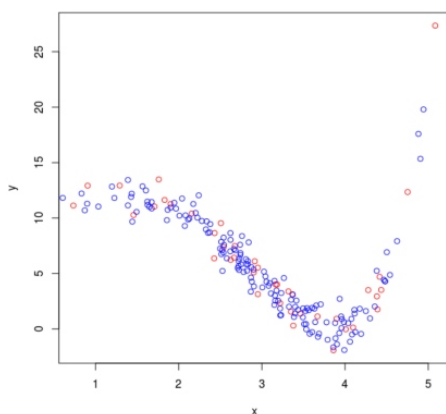


Figure 1: Train and Test Data Points

In this homework, the model that is requested to implement shown below:

$$y = w_0 + w_1x + w_2x^3 + w_3\exp(x) + w_4\sin(x)$$

To learn the parameters of the above equation, I tried to apply the recommended method by book which is multivariate regression. In multivariate linear regression, the numeric output y is assumed to be written as a linear function, that is, a weighted sum, of several input variables, x_1, x_2, \dots, x_d , and noise.

Taking the derivative with respect to the parameters w_j , $j = 0, 1, 2, 3, 4$, I got normal equations and I use the below normal equation to find parameters

$$X = \begin{pmatrix} 1 & x & x^3 & \exp(x) & \sin(x) \\ 1 & x & x^3 & \exp(x) & \sin(x) \\ 1 & x & x^3 & \exp(x) & \sin(x) \\ 1 & x & x^3 & \exp(x) & \sin(x) \\ 1 & x & x^3 & \exp(x) & \sin(x) \end{pmatrix}$$

$$X^T X w = X^T y$$

$$w = (X^T X)^{-1} X^T y$$

My parameter estimation is as below:

$$w_0 = 9.3342591, w_1 = 0.6412598, w_2 = -0.8257511, w_3 = 0.7650880, w_4 = 0.9005130$$

Looking the parameter values we can easily say that w_3 has negative effect on the output. On the other hand, the parameters other than w_3 have positive impacts on the output.

After all, I tested my algorithm with the learnt parameters on the validation data set. The root mean square error for the validation data points that tells us the difference between the predicted values and the actual values.

$$RMSE = 1.037429$$

Finally, I run the algorithm for the data interval between $[-1, 6]$ and my drawing can be seen below:

