Homework 2

For this homework, we were given a set of training data and testing data and we were expected to use the training data to train a linear regression model and apply the model to the test samples. The way I made my model robust is by data processing, deleting columns with high correlation, and performing importance selection.

First, I processed the data by reading the excels into dataframes and normalizing the columns of both the training and the testing data. And then I found and deleted columns with high correlation with each other by running selection on the correlation matrix. Note that columns with high correlation are not going to be helpful in making accurate predictions. Next, I implemented an importance selection process, which allows me to evaluate how useful each sample is in predicting. And finally, I made my final predictions using the w vector.