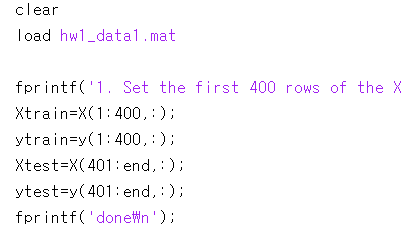
Database Project HW1

2016312107 문경진

**1. Run the given hw1\_1\_script (MATLAB script). The first column of the X matrix indicates 'Relative Compactness', and the second column indicates 'Surface Area', and the vector y indicates 'Heating Load'. Set the first 400 rows of the X matrix as the training data whereas the rest is reserved for the test data.**



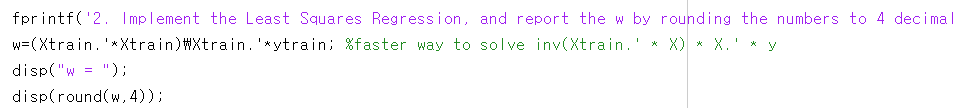
After load data1.mat, I set train and test data like this, which splits data.

**2. Implement the Least Squares Regression, and report the w by rounding the numbers to 4 decimal places.**

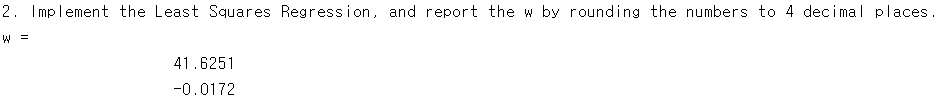
The purpose of least squares regression is to minimize objective function, and its form is…

And to minimize those error, and should be perpendicular, indicates the orthographic projection of on range(X). So,

We should use training data(Xtrain and ytrain) for X and y value.



X.’ is same as transpose(X). And, there are several ways to represent in MATLAB, such as A\B, inv(A)\*B, etc. In this case, A\B is faster than inv(A)\*B. But their answer is almost same. So in this time, I used the former one. To print w by rounding the numbers to 4 decimal places, round() is used, by using 4 for second parameter. And the answer is… [ 41.6251 ; -0.0172].



**3. Compute the RMSE on the test dataset. Round the value to 8 decimal places.**

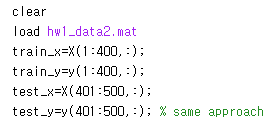
RMSE is root-mean-square-error. So to calculate RMSE, we should ‘square’, ‘mean’, and ‘sqrt’ the value. Note that predicted value should be , to compute RMSE on the test dataset.

**4. For each feature, represent the ground-truth y values and the predicted y values for the entire dataset. That is, you should have two figures: (1) Heating Load vs. Relative Compactness, and (2) Heating Load vs. Surface Area. Refer to the following example.**

(1) Heating Load (y) vs Relative Compactness ( X( : , 1) )

(2) Heating Load (y)vs Surface Area ( X( : , 2) )

**5. Run the given hw1\_2\_script (MATLAB script). The first column of the X matrix indicates 'Relative Compactness', and the second column indicates 'Surface Area', the third column indicates 'Wall Area', and the fourth column indicates 'Roof Area'. Set the first 400 rows of the X matrix as the training data whereas the rest is reserved for the test data.**



Same approach as Q1. The only difference is file name.

**6. Compute the RMSE of the Least Squares and the Ridge Regression (λ = 0.01). Round the numbers to 4 decimal places.**

As the purpose of the Ridge Regression is to minimize this objective function :

And the solution is…

**7. For each feature, represent the ground-truth y values, the predicted y through the Least Squares and the predicted y through the Ridge Regression for the entire dataset. Note that you should have four figures. The following figure shows an example of the four figures.**