CS 4491/CS 7990- Homework 2

In HW2, we aim to write a code for the implementation of Linear regression.

Download the training (housing_training.csv) and test (housing_test.csv) data from the course web page, where there are 300 samples in the training data and 206 samples in the test data. The first 13 columns will be used as independent variables (**X**) and the last column will be used as a dependent variable (**y**).

Train the model with "housing_training.csv", and list the optimal coefficient of the model. Then, given the data set of "housing_test.csv", create two plots: (1) prediction vs. ground truth and (2) the difference between prediction and the ground truth, and their residual sum of squares. Note that the ground truth of the new data is included in the last column of "housing_test.csv". You have to write code for linear regression. Don't use any library related to linear regression e.g. lm() in R.

The expected outputs:

1. Table of the coefficients in the linear regression model (see the examples below)

Intercept	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_10	c_11	c_12	c_13
-12.399	1.201	0.015	0.024	0.602	-8.828	9.131	-0.047	-1.013	0.168	-0.015	-0.642	0.017	-0.110

2. Plot of prediction vs. the ground truth (see Fig. 1)
3. Plot of difference between prediction and the ground truth (see Fig. 2)
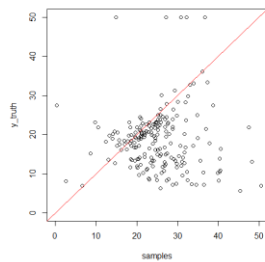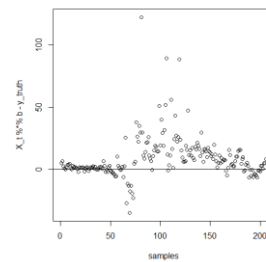

Fig. 1 Pred. vs Ground Truth

Fig. 2 Difference between pred. and ground truth

4. Residual sum of squares. Let b^* be the optimal coefficients of the model. The residual sum of squares can be computed by:

$$\sum_{i=1}^{n}(y_i - x_i b_i^*)^2,$$

where $\mathbf{X} = \{x_1, \ldots, x_p\}$, and $\mathbf{X}$ is a test data. y_i is the last column data of the test data.

Submission:

You have to submit the followings to D2L:

1. MS word file
   - Describe what you did for the homework assignment.
   - Include the four outputs (three figures and residual sum of squares)

2. Source code file(s)
   - Any languages, but recommend R or Python
   - Must be well organized (comments, indentation, …)

<u>Deadline:</u>

You have to submit HW2 by **<u>Friday, Feb 17, 2017</u>**. Late assignments will be accepted up to 24 hours after the due date for 50% credit. Assignments submitted more than 24 hours late will not be accepted for credit.