

2018 Fall, Database Project, HW1

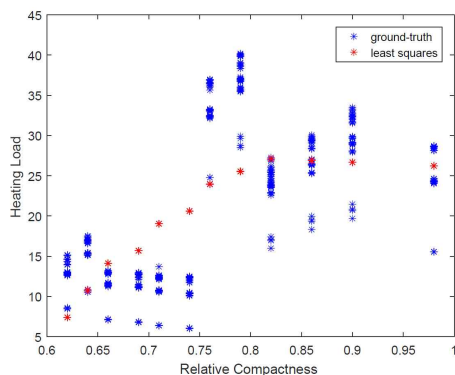
Submission: You should submit your report and the MATLAB code by Oct. 11th PM11:59 (upload them on icampus).

Late Policy: We accept late submissions until Oct. 12th AM11:59 with 50% points reduction.

Implement the Least Squares and the Ridge Regression by following the instructions below.

(the original dataset: <https://archive.ics.uci.edu/ml/datasets/Energy+efficiency>)

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.
2. Implement the Least Squares Regression, and report the \mathbf{w} by rounding the numbers to 4 decimal places.
3. Compute the RMSE on the test dataset. Round the value to 8 decimal places.
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.



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

