

Homework 1

Due Date: March 18, 2016

1 Linear Regression

Note: this is a coding homework. You are required to submit your codes and a report **before** due date. You can use any programming language to finish this homework. There are two parts in this assignments: Required and Optional. You must finish the Required part. If you finish the Optional part, you will get extra credits. For more details about homework submission, see the SubmissionGuide.pdf on the course website. All the data for this homework can be downloaded from the course website.

1.1 Required

You are required to use linear regression to find the relation between **price** and **sqft_living** in **train.csv**, and predict the house price based on **sqft_living** in **test.csv**. Please implement the following three different methods to find the best fitting function:

1. Gradient Descent
2. Newton's method
3. Normal Equation

After fitting a line to the training data,

1. You are required to make a scatterplot of **price** vs **sqft_living**, and plot your fitting line on this scatterplot.
2. You are required to use your fitting line to predict the house price based on **sqft_living** in **test.csv**, and compute the RMSE(root-mean-square error) of your prediction.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_i^{predict} - y_i^{true})^2}{N}} \quad (1)$$

, where N is the number of test samples, $y_i^{predict}$ is the prediction price for the i^{th} sample, y_i^{true} is the true price for the i^{th} sample.

In your report, you are required to compare the three methods with the help of three **plots**, three **fitting functions** and three **RMSEs**.

1.2 Optional

Actually, you can use more than one features to train your linear regression model, in order to obtain a better prediction on the test data. In this part, you are encouraged to find a better model to predict the house price.

As the dimension of your input features may be more than three, which makes it different to plot, you do not have to make plots this time. And you can pick one of the three methods to fit the linear regression line. In your report, you should include the **fitting function** (please also specify the features you choose) and the **RMSE**.