Programming assignment #1: Linear Regression

4/9/2015

In a group of 2-3
Due: Friday, 04/17/15, 11:59PM.

This assignment derives from Stanford UFLDL tutorial. You can check the tutorial for more information about problem formulation.

For this assignment, you will implement solution for a linear regression problem.

Use the github repository as a starter code. To download it, click"download zip" in the right bottom corner. The code is in matlab, so you are encouraged to use matlab for this assignment. Using other language is okay but it'll be more complicated since you have to write all the code.

Use the ex1/ex1a_linreg.m as the starter coder. This file performs most of trivial steps for you:

1. The data is loaded from housing.data. We assume that the house price $y^{(i)} \approx \theta_0 + \sum_j \theta_j x_j^{(i)}$. An extra '1' feature is added to the dataset so that it will act as the intercept term θ_0 in the linear function.
2. The examples in the dataset are randomly shuffled and the data is then split into a training and testing set. The features that are used as input to the learning algorithm are stored in the variables *train.X* and *test.X*. The target value to be predicted is the estimated house price for each example. The prices are stored in *train.y* and *test.y*, respectively, for the training and testing examples. You will use the training set to find the best choice of θ for predicting the house prices and then check its performance on the testing set.
3. It's your job to implement three versions of solutions:
    a. Use matlab optimization package minFunc to decide parameters of our model θ. minFunc uses line-search to solve the optimization problem given *linear_regression.m* . You need to compute the objective function value and the gradient with respect to the parameters in *linear_regression.m*. The *linear_regression.m* file receives training data X, training target values y, and current parameters.
    b. Implement your gradient descent method to decide parameters. Set proper step size and criterion to stop.
    c. Implement the closed-form solution for the same problem.

4. Finally, evaluate the three methods by calculating the Root mean square error of test data, and plot the results for test data on one figure. The ex1a_linreg.m file provides an example of the calculation and plotting.