Training vs. Testing

Lab Report #5

|  |  |  |  |
| --- | --- | --- | --- |
| Sidney Carlo P. Lopez (11222522)  sidneylopez18@gmail.com | |  | Jeremy B. Reccion (11210192)  jeremybreccion@yahoo.com |
| Carl Michael O. Sy (11208929)  carl.michael.sy@gmail.com | Wellington S. Villasin (11209534)  wellvill08@gmail.com | | |

*Abstract*— a training set is actualized to develop a model, while a test (or validation) set is to validate the model fabricated. Logistic Regression is a kind of regression that forecast the probability of occurrence of an event by fitting information to a logistic function and Newton's method, additionally called the Newton-Raphson strategy, is a root-discovering calculation.

Keywords—regression; machine; logistic; Newton; method; testing; training

1. INTRODUCTION

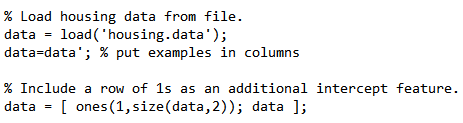
In a data set, a training set is actualized to develop a model, while a test (or validation) set is to validate the model fabricated [1]. Data points in the training set are rejected from the test (validation) set. Typically, a data set is divided into a training set, a validation set in each iteration, furthermore separated into a training set, a validation set and a test set in every cycle.

Linear Regression is a methodology for demonstrating the relationship between a scalar dependent variable y and one or more logical variables (or autonomous variables) denoted X. The instance of one informative variable is called simple linear regression. With the use of the scatterplot, it would help demonstrate the strength of two of the given variables. This was also one of the first types of regression analysis where it was greatly used in the field of practical analysis [2].

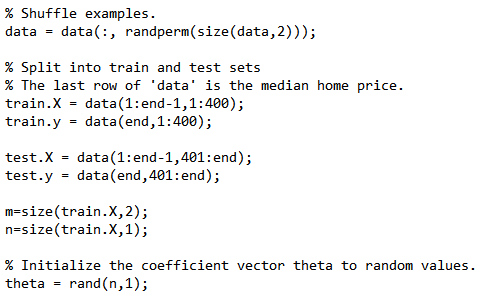
1. OBJECTIVES

* To implement the objective function and gradient calculations for linear regression in MATLAB.
* To train and test the errors that come out of linear regression.

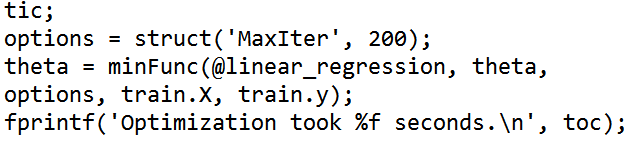
1. IMPLEMENTATION
2. The data is loaded from housing.data. An extra ‘1’ feature is added to the dataset so that θ1 will act as an intercept term in the linear function.



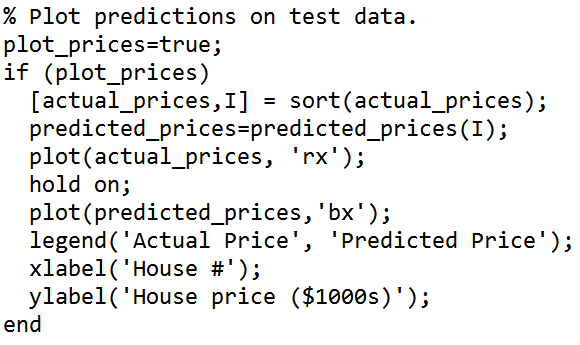
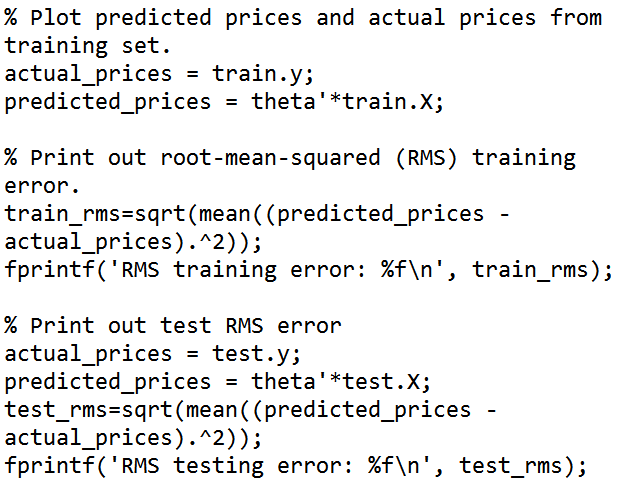
1. The examples in the dataset are randomly shuffled and the data is then split into a training and testing set. The features that are used as input to the learning algorithm are stored in the variables train.X and test.X. The target value to be predicted is the estimated house price for each example. 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.



1. The code calls the minFunc optimization package. minFunc will attempt to find the best choice of θby minimizing the objective function implemented in linear\_regression.m. It will be your job to implement linear\_regression.m to compute the objective function value and the gradient with respect to the parameters.



1. After minFunc completes (i.e., after training is finished), the training and testing error is printed out. Optionally, it will plot a quick visualization of the predicted and actual prices for the examples in the test set as shown in fig. 1.



1. DATA AND RESULTS
2. Plot predictions on test data. (Section III)

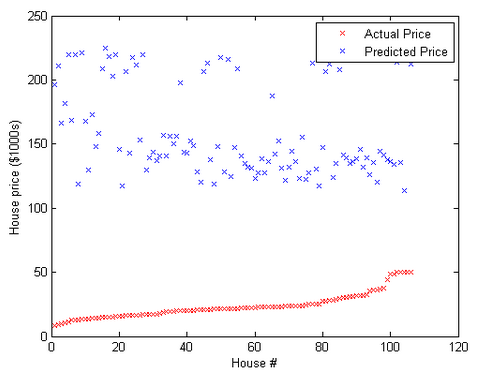


Fig. 1. Plot of house number vs price without modifications

1. Problem Plot. (Section V)

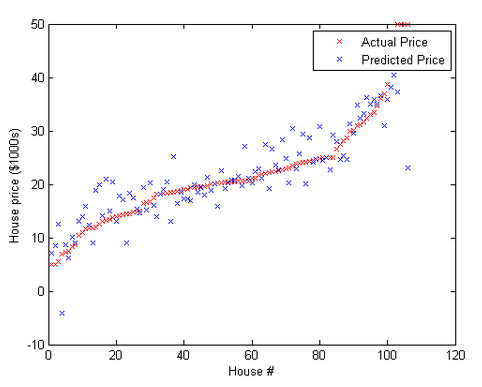
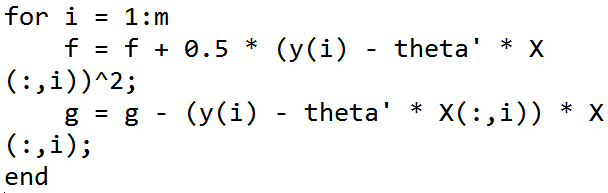


Fig. 2. Plot of house number vs price with modifications

1. PROBLEM
2. Fill in the linear\_regression.m file to compute J(θ) for the linear regression problem as defined earlier. Store the computed value in the variable f.



Plot can be found at the Data and Results section (Figure 2).

1. ANALYSIS AND CONCLUSION

For this experiment, linear regression is used again in order to simulate the training and testing for machine learning. In the previous experiments, the class had only done the training part for the simulations, but for this experiment, a testing part is also done. Basically, the experiment tested the training data hypothesis to a new, unknown test data. This allows the users to know how well the training set applies to the testing set which this might also be the measurement on how accurate the machine learning algorithm is. Overall the group successfully simulated the experiment by solving the missing algorithm in the linear\_regression.m which is mostly similar to the cost function equation but in a for loop format.

REFERENCES

1. Researchgate.net, 'Whats the difference between training set and test set?', 2015. [Online]. Available: http://www.researchgate.net/post/Whats\_the\_difference\_between\_training\_set\_and\_test\_set. [Accessed: 06- Oct- 2015].
2. Y. Personnel, "Yale Education," Yale University,[Online].Available: http://www.stat.yale.edu/Courses/1997-98/101/linreg.htm. [Accessed 8 September 2015].
3. Mathworld.wolfram.com, 'Newton's Method -- from Wolfram MathWorld', 2015. [Online]. Available: http://mathworld.wolfram.com/NewtonsMethod.html. [Accessed: 22- Sep- 2015].
4. Machine Learning of Stanford, OpenClassroom by Andrew Ng

(http://openclassroom.stanford.edu/MainFolder/DocumentPage.php?course=MachineLearning&doc=exercises%2Fex3%2Fex3.html)