# Supervised Linear Regression

Aquino, Patrica Rose
Basallote, Lawrence Andrew
Carag, Stephanie
Jacinto, Dan Emanuel
Lunasco, Jan Osbert

***Abstract*—The goal of linear regression is to see whether the probability of getting a particular value of the nominal variable is associated with the measurement variable whereas for logistic regression is to predict the probability of getting a particular value of the nominal variable, given the measurement variable.**

## I. Introduction

There are three data sets, specifically training set, validation set, and test set. Experiment 5 deals with the difference between training set and testing set. Training set is a set of data used for learning/training the classifier while test set is used to assess the performance of a trained classifier. Training set are data where labels are given while test set are data where labels are known but not given [1].

## II. Procedure

### A. Linear Regression

For this exercise you will implement the objective function and gradient calculations for linear regression in OCTAVE.
In the ex1/ directory of the starter code package you will find the file ex1_linreg.m which contains the makings of a simple linear regression experiment. This file performs most of the boiler-plate steps for you:

1) The data is loaded from housing.data. An extra 1 feature is added to the dataset so that $\theta 1$ will act as an intercept term 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 $\theta$ for predicting the house prices and then check its performance on the testing set.
3) The code calls the minFunc optimization package. minFunc will attempt to find the best choice of $\theta$ by 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.
4) 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.

The ex1_linreg.m file calls the linear_regression.m file that must be filled in with your code. The linear_regression.m file receives the training data X, the training target values (house prices) y, and the current parameters $\theta$. Complete the following steps for this exercise:

1) Fill in the linear_regression.m file to compute $J(\theta)$for the linea r regression problem as defined earlier. Store the computed value in the variable f.

You may complete both of these steps by looping over the examples in the training set (the columns of the data matrix X) and, for each one, adding its contribution to f and g. We will create a faster version in the next exercise.
Once you complete the exercise successfully, the resulting plot should look something like the one below:

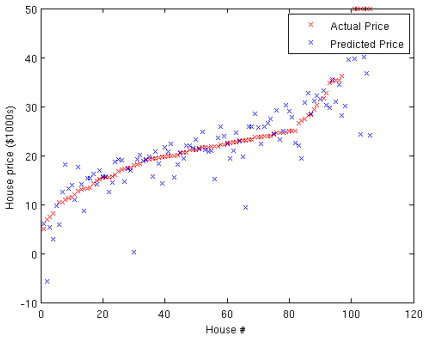


Fig. 1.

## III. Data and Results

### A. Exercise 1A: Linear Regression

In the following results shown in Fig. 2 the typical errors are between 4.5 to 5 but it may vary from different random choice of training set. From the actual price vs the predicted price, it was shown that predicted price shows very large errors if we compare it to the actual.

## IV. Conclusion

In this experiment, we used the codes from the first experiment (Linear Regression) to show the difference between training and testing. The process in this experiment is almost similar with the procedures in the first experiment, except this time, the data was split into training and testing sets. Each set has

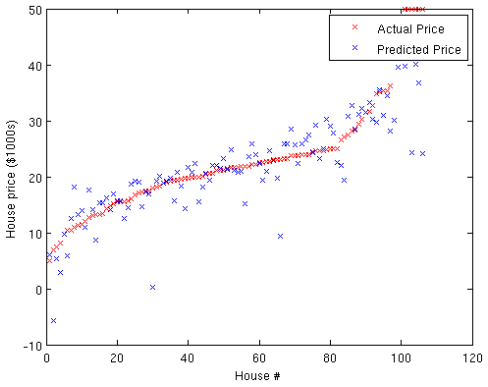


Fig. 2. House number vs House price Actual and Predicted Price

different input values from the other, and also different target values. We used the training set to find the best value of theta to predict the housing prices (with respect to the training set), and used the same theta to observe how it would fare in the testing set. Based on the resulting plot acquired, it can be observed that actual price and the predicted price are very close to one another, albeit it does not overfit nor underfit. Therefore, the hypothesis is applicable for future values.

## REFERENCES

[1] O. E. Tom Mitchell and P. Domingos, "Machine learning." [Online]. Available: https://courses.cs.washington.edu/courses/csep573/11wi/lectures/19-mlintro.pdf