Machine Learning

Project 1 -Project Report

Submission – 10/19/2014 Sunday

Submitted by

Bala Seshadri Sura

50097470

balasesh.buffalo.edu

Introduction:

The objective of this assignments is to train a regression and a gradient descent model. The given consists of 69623 rows and 48 columns out of which 69623*46 is taken into matrix and divided for training validation and testing. The rows also known as data points and the columns as features. This data is split into 3 parts of 80% into training data, 10% into validation data and 10% into testing data.

We need to find a linear model that produces the given data and the model that will be used are the Gaussian and gradient descent model. In order to get the best fit we must obtain the Model complexity of the design matrix “M” and use the regularization parameter. The best fit is obtained by minimizing the error function that calculates the misfit between the function and the target value. The metric used to calculate the error is Root Mean Square Error, which is specified in the detailed project description.

Training:

Here we calculate the design matrix which the Gaussian basis function. And we use this to calculate the E(w), which in turn is used to calculate the Erms value. This is the value that is minimized over iterations. We also use the gradient descent model to calculate the error which reduces the data up to minimum value and is saved to the system and printed.

Validation:

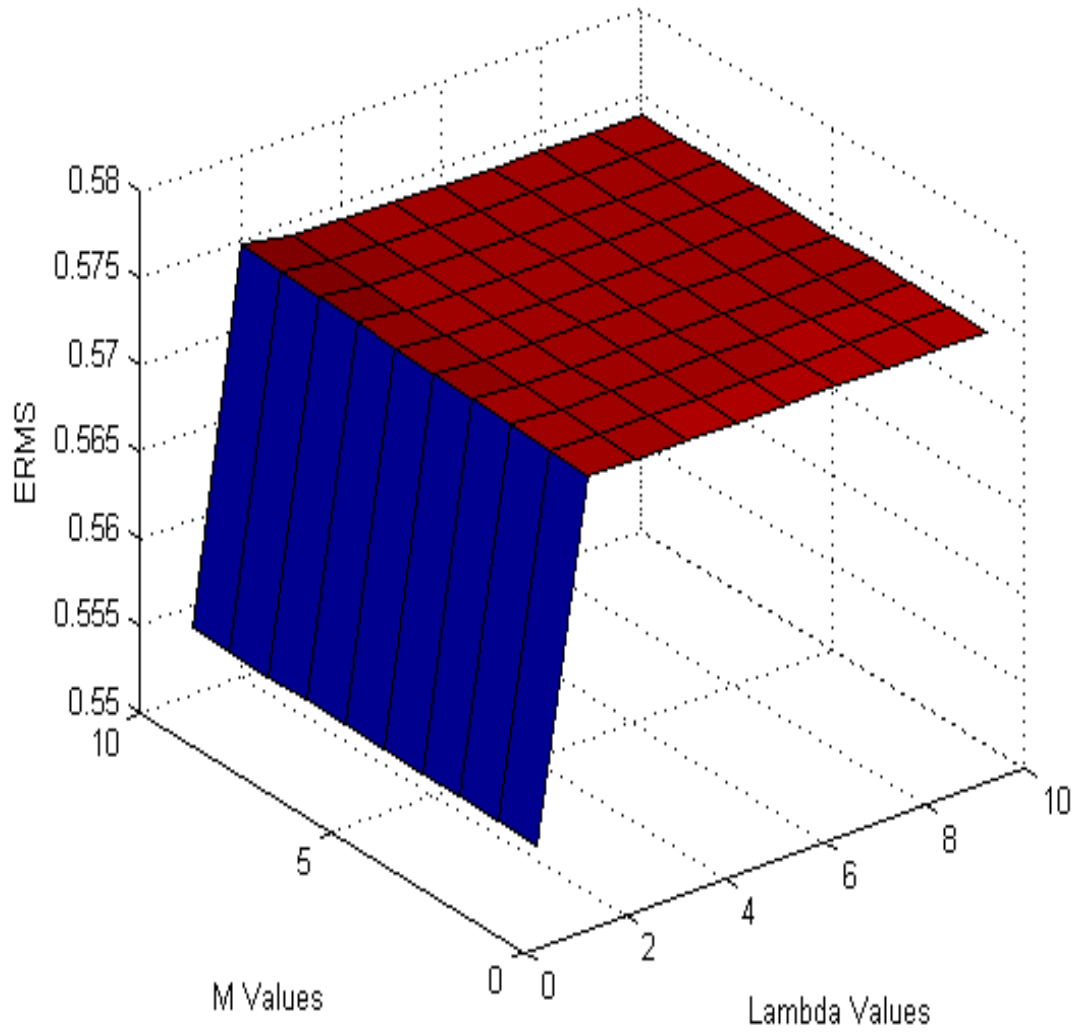
In this section we obtain the data from the training model and validate it and make sure we are getting a minimal error value. We then send this data to testing phase so that we can make sure the data is tested.

Testing:

In the test phase we check if the calculated weight vector ‘w’, the model complexity ‘M’ and the regularization parameter ‘lambda’ is providing the best fit for the test data. This was calculated, to be Erms = 0.6693 for model complexity M=7 and lambda=5 for the Linear Regression Model with a Gaussian basis function. As for the Gradient descent model it is: 0. 5889.

The graph shows us that the data gradually decreases and stops at one point. At this point the error rate is minimum and this was recorded to be M = 7 and Lambda = 5. Please find the below graph produced for the data collected from the training model.

The X axis represent – Lambda values.
The Y axis represents – The model complexity values, and
The Z axis represents – The Erms values.



References:

http://www.mathworks.com/help/matlab/ref/surf.html

http://www.mathworks.com/help/matlab/

http://en.wikipedia.org/wiki/Gaussian_function

http://en.wikipedia.org/wiki/Gradient_descent