CSE574 MACHINE LEARNING PROJECT-1
LINEAR REGRESSION WITH BASIS FUNCTIONS

Karthik Kiran
UB#: 50133145

Project Report

Introduction:

The aim of the Project is to train a Model to fulfill the task of Linear Regression with Basis Functions. Model is trained to map an Input vector 'x' into a target values 't'.
The task of Linear Regression is accomplished by generating Gaussian Basis functions for the given Input Data.

Two methods of solution are implemented in this project:

i) Closed-Form Maximum Likelihood Solution
ii) Stochastic Gradient Descent

Approach:

The given Microsoft LETOR dataset was read into a matrix and only the data required for Linear Regression were extracted from the file. This extracted data is stored in an Input Matrix which contains 47 columns. The first column represents the Target Values or Relevance Labels for the query document pair and the following 46 Columns represent the 46 Features representing each Query document pair.
The extracted input data is split into three subsequent subsets of data. The first subset contains 80% of the entire data representing the Data that is used for Training the model. The second subset of data contains 10% of the entire Dataset used for validating the model. The remaining 10% of data is used for testing the model. First column of each of these Matrices is stored in three different matrices representing the target vectors of training, validation and testing processes.

Closed-Form Maximum Likelihood Solution

Training

i) The Model is trained for different values of M (Model complexity), S (isotropic special scale) and L (Lambda or regularization coefficient). For each M value, a number of S and L values were used.

ii) Basis functions were generated for different values of M. Different means were selected randomly to generate each Basis function. The number of Basis functions generated is equal to M.

iii) Design matrix **'Phi'** is generated using the Basis Functions generated. Each column of the design matrix corresponds to one Basis function. The first column of the design matrix is always a column vector of 1's.

iv) Weight vector is calculated for the design matrix generated using Basis functions in the previous step. This weight vector is calculated using design matrix and the Target vectors for the Training Data.

v) Each weight vector generated for a particular 'M' value is fed to the Validation program.

Validation

i) 'M' value is varied from 1 to 10.

ii) The model is checked for accuracy of training. Basis functions are generated for the validation data. The number of Basis functions generated depends on the Input value 'M'. Design matrix is formed for the validation set with its first column being a vector of 1's.

iii) Root mean square error (ERMS) is calculated for the validation data using the Weights generated in the training phase. This ERMS is calculated for different values of Lambda also called the regularization term. Lambda is used to avoid over fitting.

iv) The minimum value of ERMS is stored along with the corresponding 'M' and Lambda value.

v) This process is repeated for different values of 'M', 'S' and Lambda until a minimum ERMS value is achieved.

vi) ERMS was found to be 0.5523 for M = 9 and Lambda = 2;

Testing

The Testing data contains 10% of the entire data set. It has 6963 data points.

i) The Mean vector and Lambda values for the minimum ERMS value generated during Validation are passed to the testing program.

ii) The Weight vector of the training phase which was sent to the Validation phase is also passed on to the testing phase.

iii) Basis functions for testing data is generated using Mean vector obtained from the Validation phase and the design matrix for testing data is generated using these Basis functions.

iv) ERMS of testing phase is calculated using the values of the above three steps and is noted.

Testing Result Closed form solution: ERMS = 0.6404 for M = 9 and Lambda = 1;

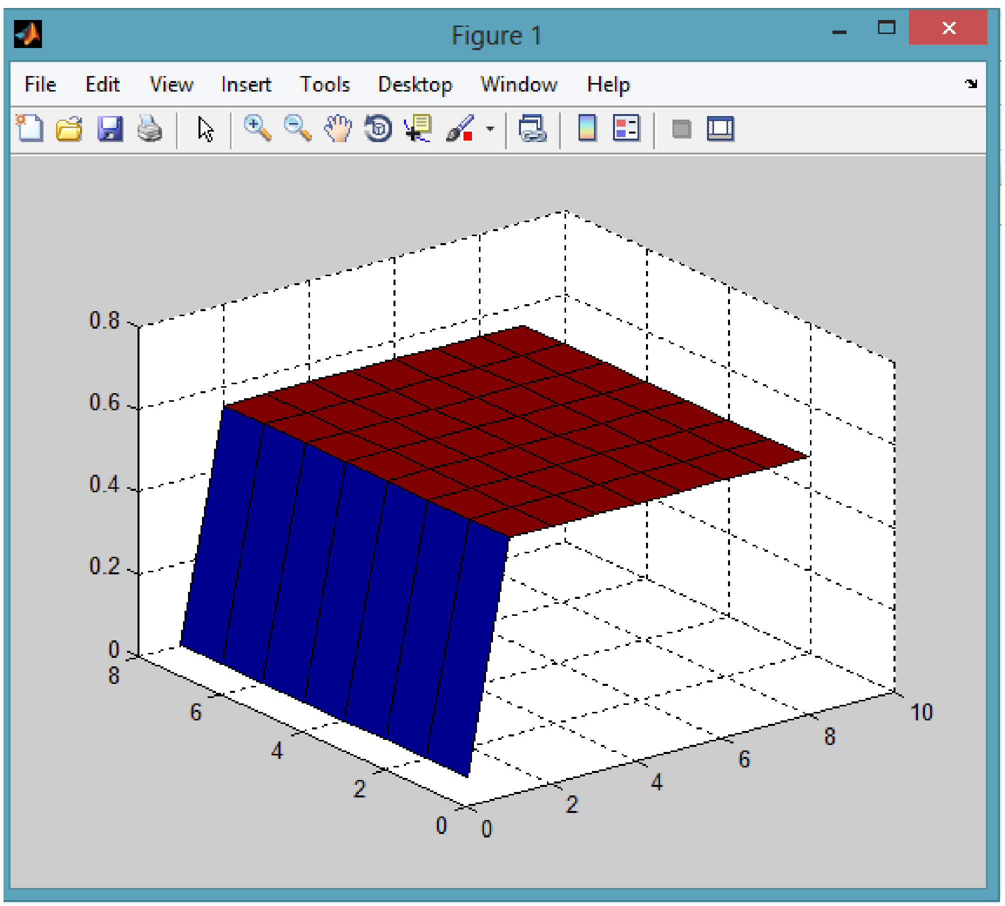

Figure1: The above plot shows the variation of Erms with respect to lambda and model Complexity (M) for Gaussian basis function

Maximum Likelihood with Gradient Descent:

Training and Validation

The Model was trained and validated for the best fit M (Model Complexity), S (isotropic special scale) and L (Lambda or regularization coefficient) obtained from Closed form solution.

The following process was executed to train and validate the Model:

i) Value of eeta (learning parameter) is initialized to 1 at the start of the training and validation phase.

ii) The weight vector is taken initially as a vector of zeroes with dimension (M*1).

iii) The design matrix for Training data using Gaussian Basis functions is calculated for the value of input M.

iv) For each row in the Training data design matrix a new weight vector is obtained. The second weight vector is obtained using the initialized weight vector consisting of all zeroes, a new weight vector is obtained for the training data.

v) The weight vector obtained in the previous step is passed to the validation phase where the ERMS is calculated using the corresponding design matrix of validation set.

vi) The next weight vector from the Training set is obtained by using weights calculated in the previous iteration.

vii) ERMS is calculated for each weight vector from the training phase.

viii) In the validation phase, the current ERMS value is compared with the ERMS value of the previous iteration. If current value is lesser than previous value, then learning parameter value is not changed. If current ERMS value is greater than previous ERMS value, learning parameter value is divided by 2 for the next iteration.

ix) After the completion of all iterations of calculating weights and ERMS, the minimum value of ERMS and corresponding weight vector is passed on to the testing phase.

x) ERMS was found to be 0.6002 for M = 9 and Lambda = 1;

Testing

i) The design matrix required to calculate ERMS is generated for the testing data.

ii) Using the values of Lambda and weights obtained from validation phase, the ERMS is calculated for testing set.

Testing Result Stochastic Gradient Descent: ERMS = 0.6364 for M = 9 and Lambda = 1;

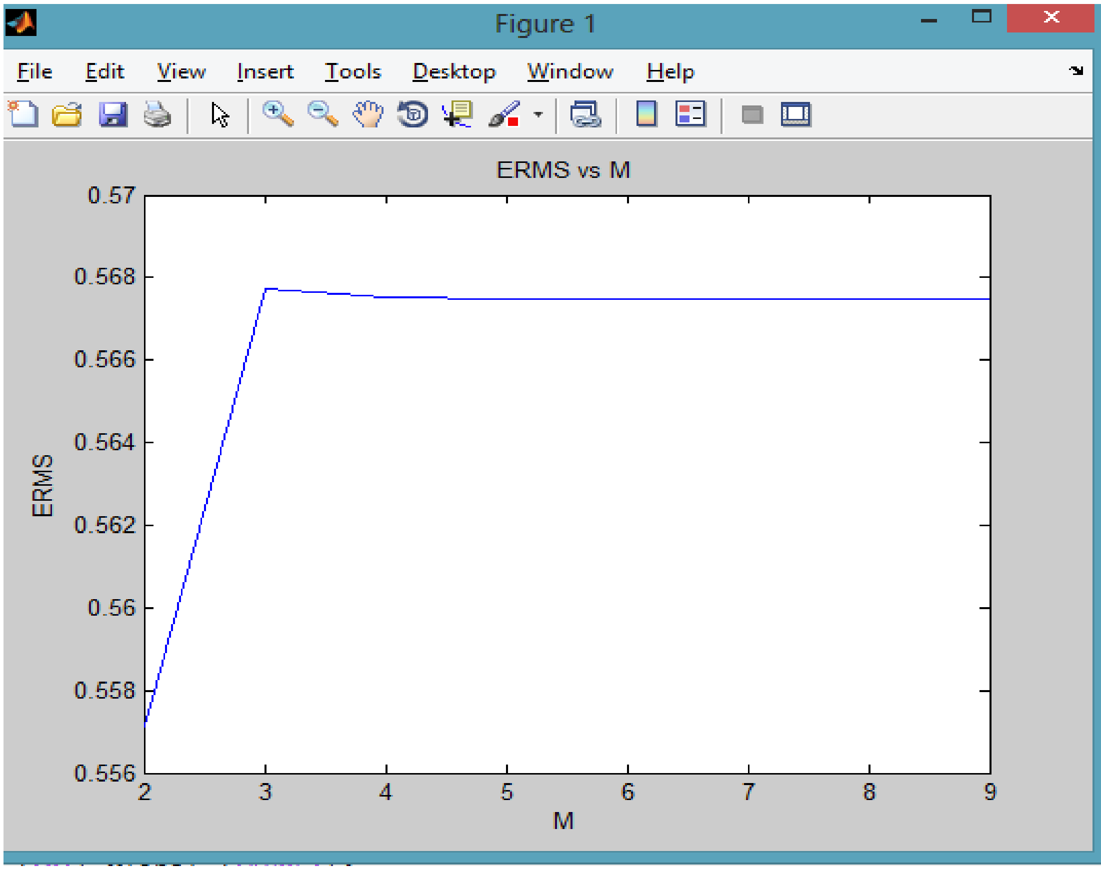

Figure2: The above plot shows the variation of Erms with respect to model Complexity (M) for Stochastic Gradient Descent Method