Project 1

Ression on Page Relevancy
CSE4/574 Machine Learning

Ankit Jain

UBIT: 50097432

Project 1: Regression on Page Relevancy

This project is to implement and evaluate several supervised machine learning approaches to the task of linear regression. The objective is to learn how to map an input vector x into a target value t using the model

$$y(\mathrm{x}.\mathrm{w}) = \mathrm{w}^T \phi(\mathrm{x})$$

where w is a weight vector to be learnt from training samples and phi is a vector of M basis functions. Each basis function phi converts input vector x into a scalar value.

1. Models used in this Project:

1. Closed form solution (Linear Basis Function Model)

$$y(\boldsymbol{x}, \boldsymbol{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\boldsymbol{x}) = \phi(\boldsymbol{x})\boldsymbol{w}$$

I have used Gaussian basis function

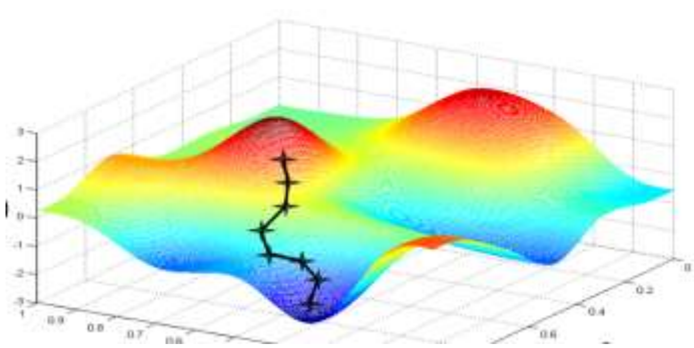
$$\phi_j(\mathrm{x}) = \exp\left(-\frac{(\mathrm{x}-\mu_j)^2}{2s^2}\right)$$

where phi is a vector in feature space and s is an isotropic spatial scale.

2. Stochastic Gradient Descent

$$\mathrm{w}^{(\tau+1)} = \mathrm{w}^{(\tau)} - \eta \nabla E_n \qquad \mathrm{w}^{(\tau+1)} = \mathrm{w}^{(\tau)} + \eta(t_n - \mathrm{w}^{(\tau)\mathrm{T}}\phi_n)\phi_n$$

Start with some w and keep changing w and w previous until we end at some minimum. Graphically, this can be represented as,



Comparison of above approaches:

In closed form solution, we loop over all data elements of training set in one iteration.

For **Stochastic Gradient Descent**, the vector gets updated as, at each iteration the algorithm goes over only one among j^{th} training set

Files included in this project

project1.m – Dividing the dataset into training, validation and testing, calling other functions and printing output.
train_cfs.m - Function to train the closed form regression model.
test_cfs.m – Function to test the generated model on test dataset.
train_gd.m - Function to train the gradient descent regression model.
test_gd.m – Function to test the generated model on test dataset.
W_cfs.mat – w data for closed form solution.
W_gd.mat – w data for gradient descent solution.
project1_data.mat – Extracted features and target variables from raw data.

Choice of these models:

This models have been chosen as they give us a good relation between dependent variables and allow us to choose the best combination of values.

Choice of parameters:

1. **Linear Closed form model**

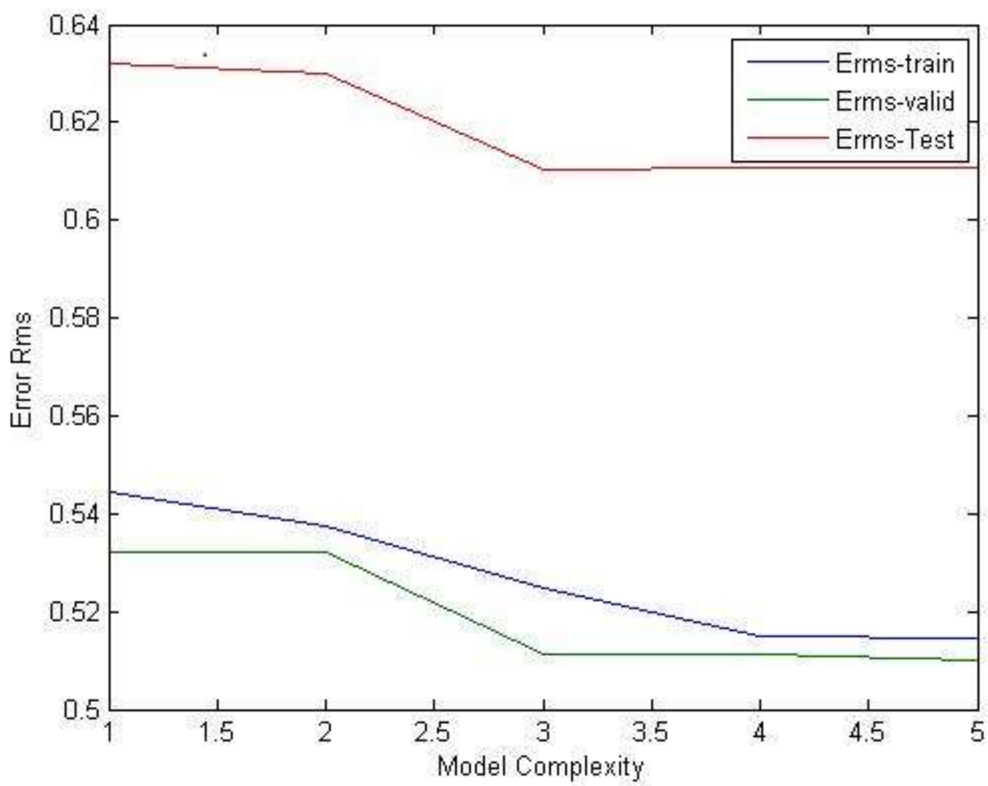
Parameter	Start Value	Step Size	End Value
Mu	0.1	0.1	0.5
S	Mean of Variance of columns values	NA	NA
Lambda	0.1	0.1	0.5
Model Complexity (modelC)	1	1	7

2. Gradient Descent Model

Parameter	Start Value	Step Size	End Value
Mu	0.1	0.1	0.5
S	Mean of Variance of columns values	NA	NA
Lambda	0.1	0.1	0.5
Model Complexity (modelC)	1	1	7

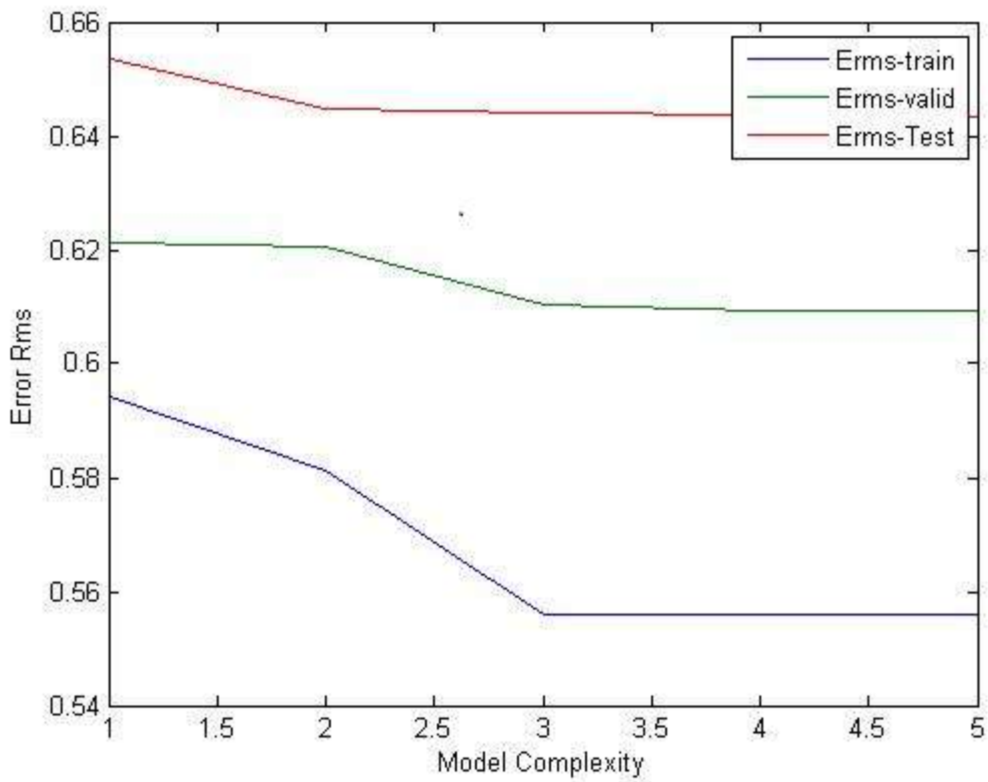
Evaluating models:

For different model Complexity, the value of Error-RMS for closed form solution was found to be:



Analysis:
For test data, the error first decreased with increasing model complexity and then it started increasing slightly. We assumed that we have achieved a global minimum and used the optimal model complexity with lowest Error -RMS value.

For different model Complexity, the value of Error-RMS for gradient descent solution was found to be:

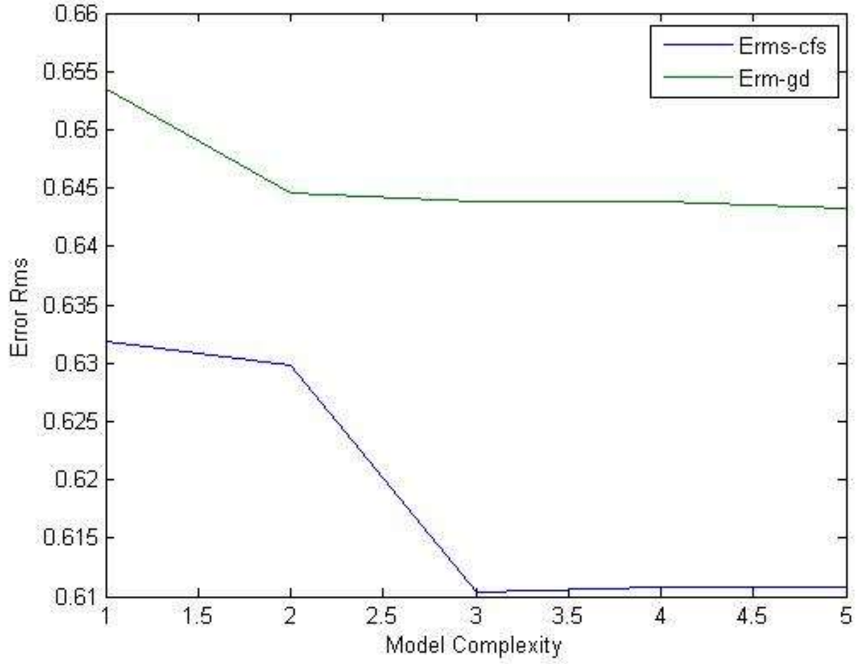


Analysis:
For test data, the error first decreased with increasing model complexity and then it started increasing slightly. We assumed that we have achieved a global minimum and used the optimal model complexity with lowest Error -RMS value.

Parameter	**Linear Closed form**	**Gradient Descent**
Training Time	Less (80 sec)	More (120 sec)
Erms-test	Less (0.63)	More (0.64)
Lambda	0.5	0.5
Model Complexity	7	7

When the training set is large, Stochastic Gradient Descent can be useful (as we need not go over the full data to get the first set of the parameter w)

Comparison of Erms-test for linear closed form solution and gradient descent solution



Comparison of Erms-test of closed form solution and gradient descent solution through histogram