Project Report – by Amber Rastogi (UBID - 50097978)

Objective: Implement regression on web search ranking data set.

Training data Source: Query-url pair datasets from Microsoft LETOR 4.0.

Models used:

1. **Gaussian model for Linear Regression**:

Linear Basis Function	Gaussian Basis Function
$y(\boldsymbol{x},\boldsymbol{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\boldsymbol{x}) = \boldsymbol{\phi}(\boldsymbol{x})\boldsymbol{w}$	$\phi_j(\boldsymbol{x}) = \exp\left\{-\frac{(x-\mu_j)^2}{2s^2}\right\}$

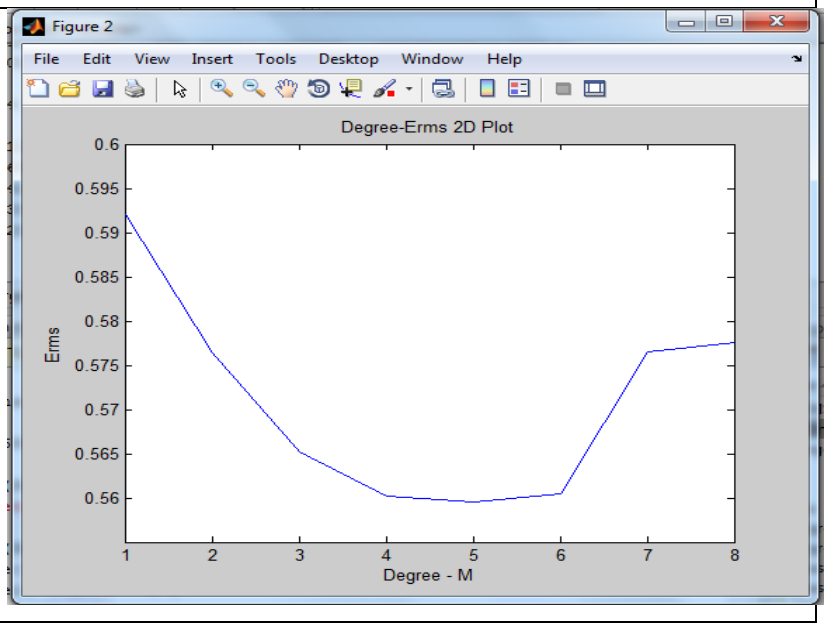
i. **Intuitive choice of methods:** The method has been used since it takes into account the correlations among the dependent variables and allow us to select the best combination.

ii. **Intuitive choice of parameters:** Since all values for the different Dimensions range from 0 to 1 it becomes intuitive to select parameters as following:

Parameter	Start Value	Step Size	End Value
mu	.1	.1	.5
S	Mean of column deviations	NA	NA
lambda	.0001	. 0001	. 0005
M (Degree)	1	NA	5

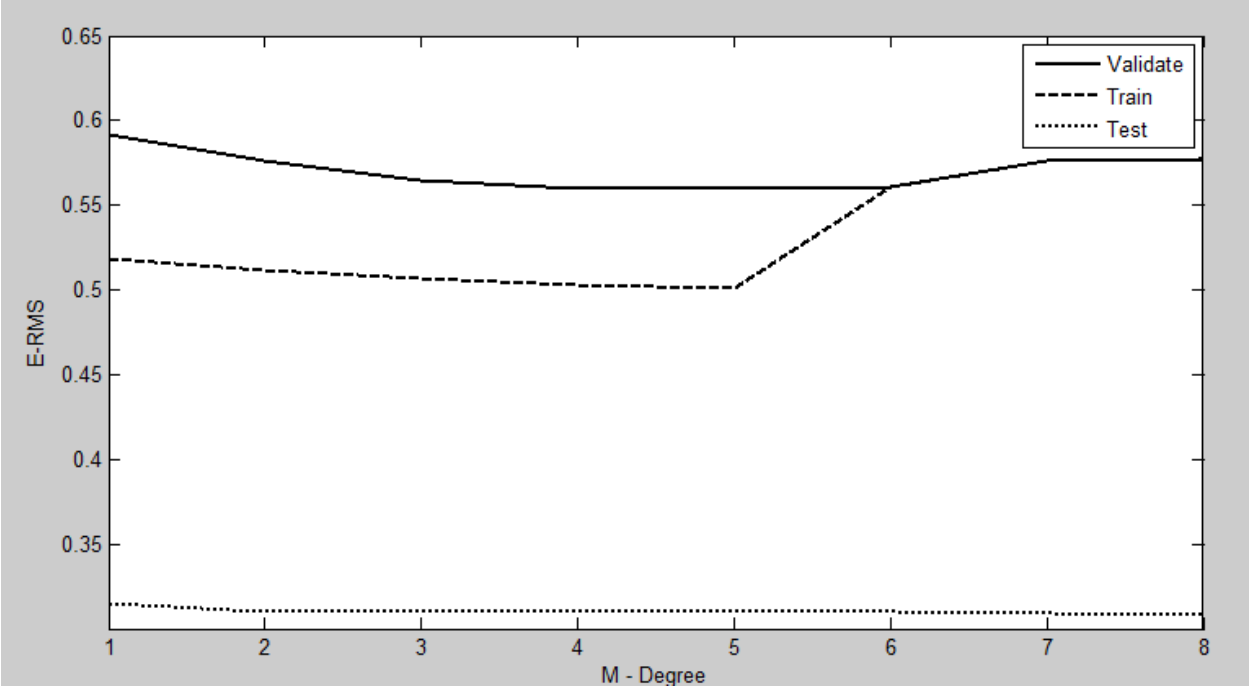
iii. **Avoiding over fitting:** Over fitting will result into an increase in error after we use certain degree of M. The model was chosen after testing with various degrees of M.

For various degree the value of Error-RMS was found to be:

M	ERMS	M – ERMS 2D Plot
1	0.592020	
2	0.576395	
3	0.565133	
4	0.560242	
5	0.559553	
6	0.560423	
7	0.576567	
8	.5775456	

The error first decreased with increasing degree of M and then it started increasing. We assumed that we have achieved a global minimum and used the optimal degree of M with lowest Error -RMS value.

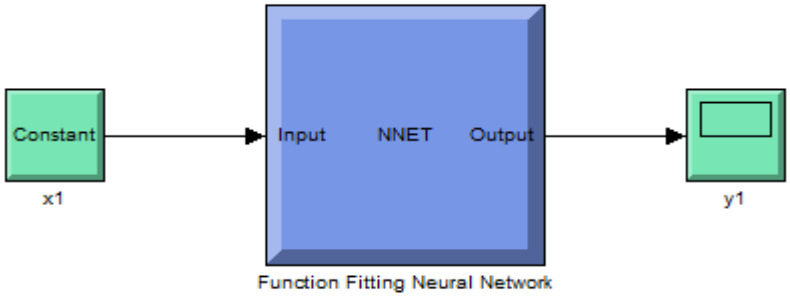
iv. E-RMS Plots:



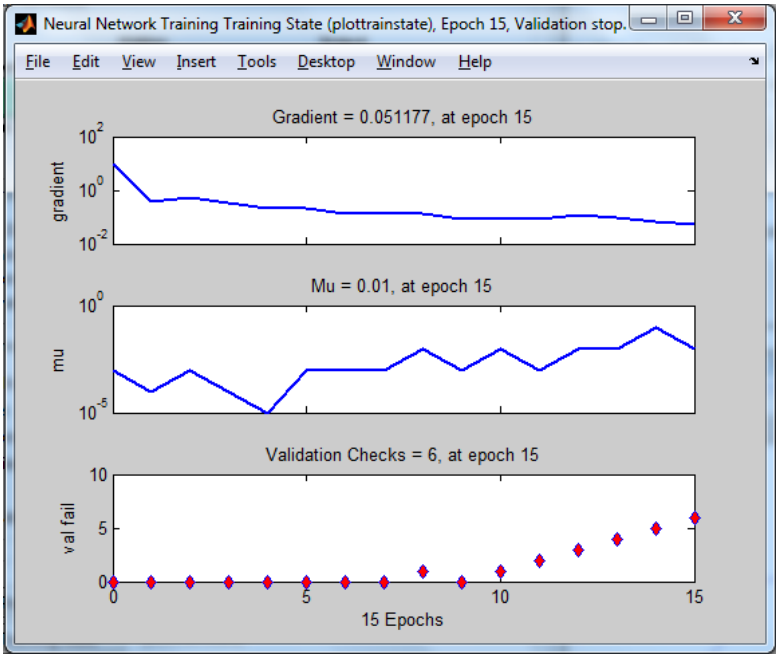
---

2. Neural network Model:

Linear Basis Function	Neural Basis Function
$y(\boldsymbol{x},\boldsymbol{w}) = \sum_{j=0}^{M-1} w_j \phi_j(\boldsymbol{x}) = \boldsymbol{\phi}(\boldsymbol{x})\boldsymbol{w}$	$y(\mathrm{x},\mathrm{w}) = f\left(\sum_{j=1}^{M} w_j \phi_j(\mathrm{x})\right)$



i. **Intuitive choice of methods:** Neural networks have advantages when dealing with data that does not adhere to the generally chosen low order polynomial forms, or data for which there is little a priori knowledge of the appropriate CER to select for regression modeling.
Source: http://www.eng.auburn.edu/~smithae/publications/journal/tony.pdf

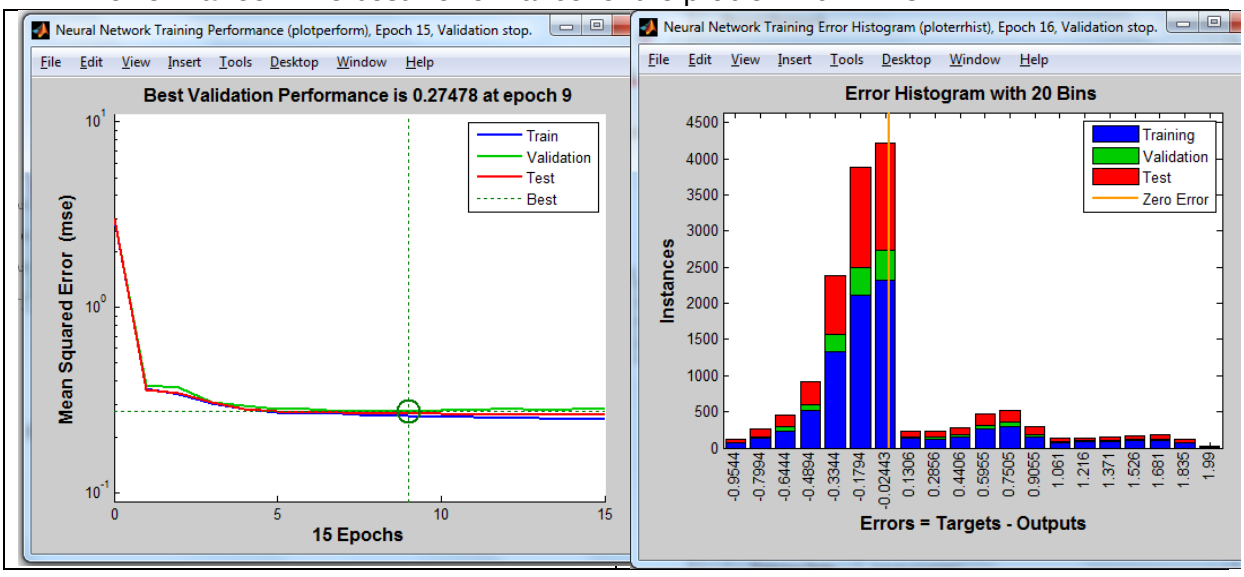


ii. **Intuitive choice of parameters:** We used following parameters for training our Neural network:

Parameter	Start Value	Step Size	End Value
Epoch	1	NA	15
Neurons	10	NA	NA
Mu	.00001	100 Multiplicative factor	NA
Lambda	.0001	. 0001	10^10
Performance	.248	NA	NA

iii. **Avoiding over fitting:** Over fitting will result into an increase in error after we increase EPOCH after a certain value. We stopped further training the neural network after we got the best Validation Performance at EPOCH 9. The value of E-RMS was stabilized or seemed to be increasing after the value.

iv. **Performance:** The best Performance for the problem : 0.27478



Comparison of the Models

Parameter	Neural Regression	Gaussian Regression
Training Time	Less : 20 Sec	Higher : >1 min for high M
E-RMS	Less (Performance: .24)	High (E-RMS: 0.559551271)