CSE 574 - Project 1

ANKIT SHARMA - 50134053

Oct 19, 2014

Linear Regression with Basis Functions

Division of Microsoft LETOR Dataset:

Training Set: 80% (1-50,000)

Validation Set: 10% (50,000-60,000)

Test Set: 10% (60,000-end)

Term Index:

N: Number of training samples
Nv: Number of elements in validation set
D: Dimension of each training sample
X: Input dataset (NxD)
T: Target Set (Nx1)
M: No. of basis functions
Φ: Basis matrix (NxM)
s: Isotropic spatial scale (sigma)
λ: Regularization constant (lambda)
W: Weight Matrix (Mx1)
E_D: Sum-of-square error
E_{RMS}: Root-mean-square error

Basis Functions Used:

Gaussian basis function:

$\phi_j(x) = \exp(-(x-\mu_j)^2 / 2s^2)$

Here, $x \in X$, is the input vector consisting 46 features. Therefore, D=46.

Selection of means ($\mu \in \mathbb{R}^{MxD}$):

There are M-1 *means* each of which is a vector of D elements. *Means* are calculated by dividing the training set into M-1 groups and then calculating mean for each group.

Now the basis matrix $\Phi \in \mathbb{R}^{NxM}$, can be calculated as follows:

for i=1:N{
    $\phi_1(x_i) = 1$
    for j=2:M{
        $\phi_j(x_i) = \exp(-(x_i - \mu_{j-1})^2 / 2s^2)$
    }
}

Part 1: Maximum likelihood (ML) Closed-Form Solution

We have to select three parameters: Model Complexity (M_cfs), s and λ.

We first calculate optimal values of sigma*(s)*, by taking a fixed lambda as λ=0.0001. Following are the steps involved in determining these values.

for s=0.1:0.1:1.0{

for M=2:100{

Step 1: Divide the training set into M-1 groups

Step 2: Compute M-1 means (μ) (as described above)

Step 3: Compute basis matrix Φ (as described above)

Step 4: Calculate maximum likelihood (ML) solution using the following:

$$W_{ML} = (\Phi^T * \Phi + \lambda * \mathbf{I})^{-1} * \Phi^T * T$$

Step 5: Run training and validation set to calculate sum-of-square error (E(W))

$$E_D(W) = (1/2) * \sum_{n=1}^{N} \{\mathrm{Tn} - \mathrm{W}'\Phi(\mathrm{Xn})\}\hat{}2$$

$$E(W) = E_D(W) + (1/2)\lambda\, W^T W$$

Step 6: Calculate E_{RMS} value

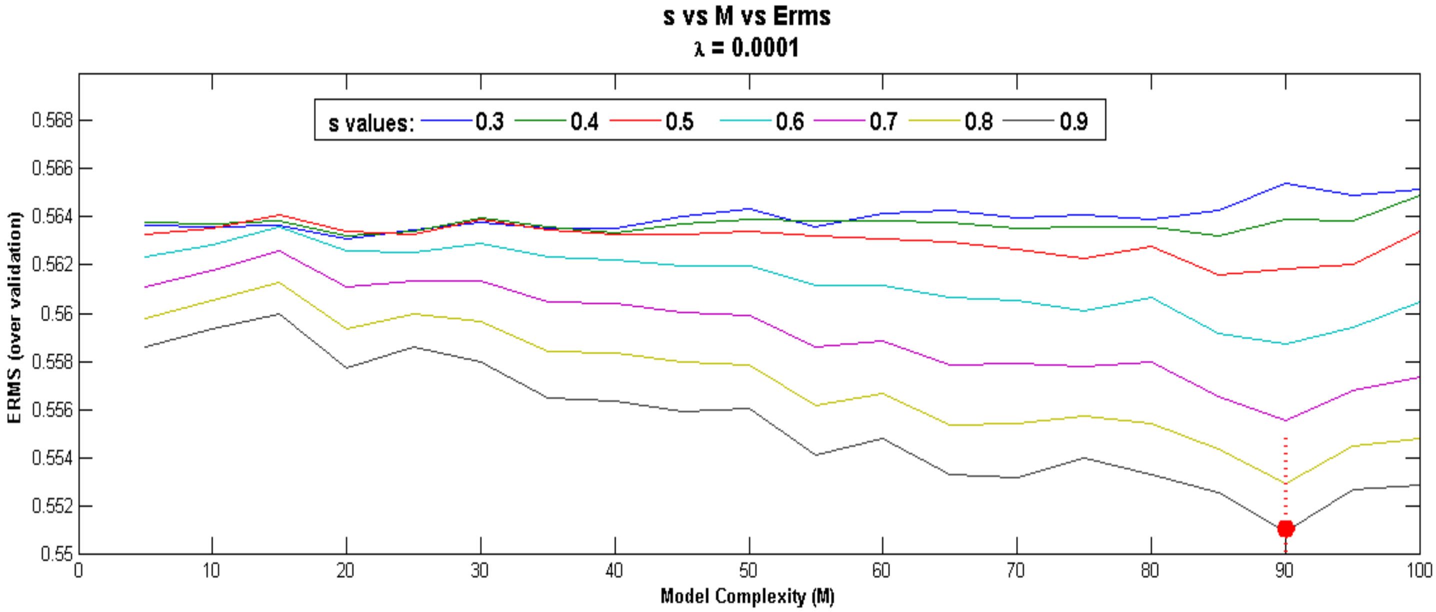
$$E_{RMS}(W) = \sqrt{2E(W*)/N\mathrm{v}}$$

}

}

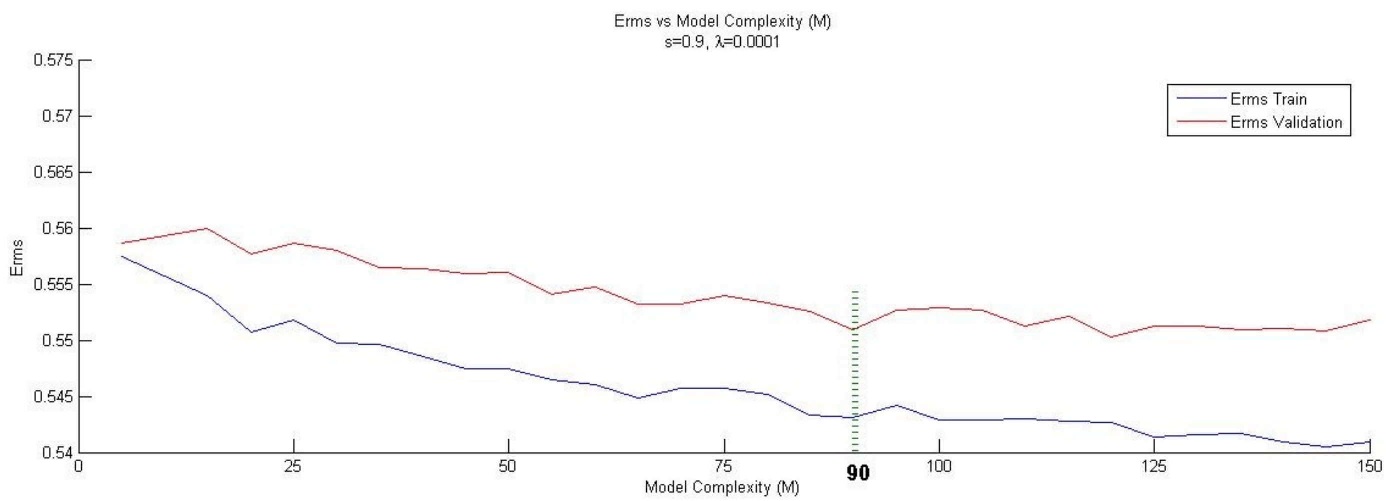
Validating s and M

Following the graph of E_{RMS} of validation set with respect to the values of M and s.



We can see that least possible values of E_{RMS} can be obtained when we select **s_cfs=0.9**

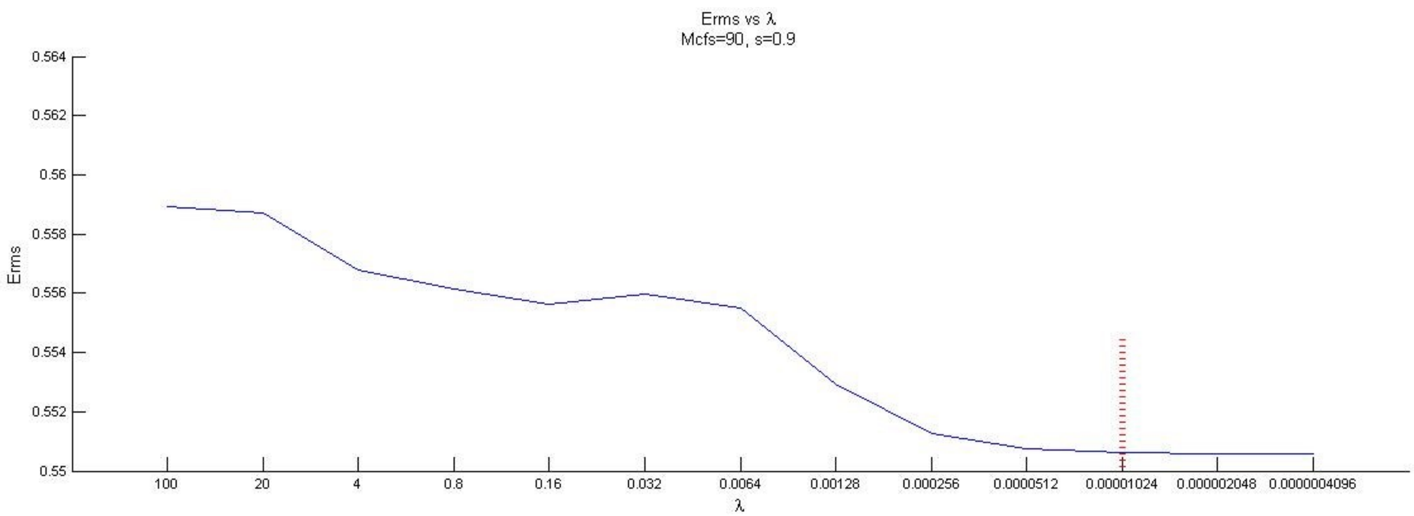
Now we calculate the Model Complexity (M_cfs) using the values of above graph when s=0.9. Following is the graph of E_{RMS} of both validation and training sets with respect to the values of M at s=0.9.



From the above graph it can be observed that Erms for Training and Validation substantially decreases as M increases. After M=90 the Erms for training keep decreasing but Erms for Validation does not decrease. Hence, we selected **M=90** for this model as M_cfs.

Validating L (lambda)

Now, we calculate the most optimal value for lambda (λ_cfs). We already know the optimal values of s and M. Therefore, we set these values and calculate Erms for validations set for different values of λ. Following the graph we obtained:



From the above graph it can be observed that Erms for Validation substantially decreases as λ decreases. After λ=0.00001 the Erms for Validation remains constant. Hence, we selected **λ_cfs=0.00001** for this model.

Therefore, after validation, we obtain the following most optimal parameters for our system:

Final Outcomes of Closed-Form Solution

M_cfs = 90

S_cfs = 0.9

λ_cfs= 0.00001

Erms Train: **0.5431**

Erms Validation: **0.5506**

Erms Test: **0.5971**

Part 2: Stochastic Gradient Descent

We have to select three parameters: Model Complexity (M_gd), s and α.

We first calculate optimal values of sigma*(s)*, by taking a fixed alpha as α=0.001.

Calculating Weights for Stochastic Gradient Descent:

```
Repeat{
    for i=1:N{
        for j:=1:M{
                W_j := W_j +  α*(T_i – W^T*ϕ(x_i))*ϕ_j(x_i);
        }
    }
}
```

Procedure:

Following are the steps involved in determining the parameters with respect to ERMS values:

for s=0.1:0.1:1.0{

for M=2:100{

Step 1: Divide the training set into M-1 groups

Step 2: Compute M-1 means (μ) (as described above)

Step 3: Compute basis matrix Φ (as described above)

Step 4: Calculate the weights using Stochastic Gradient Descent (as described above)

Step 5: Run training and validation set to calculate sum-of-square error (E(W))

$$E_D(W) = (1/2) * \sum_{n=1}^{N}\{\mathrm{Tn} - \mathrm{W}'\Phi(\mathrm{Xn})\}\text{^}2$$

Step 6: Calculate E_{RMS} value

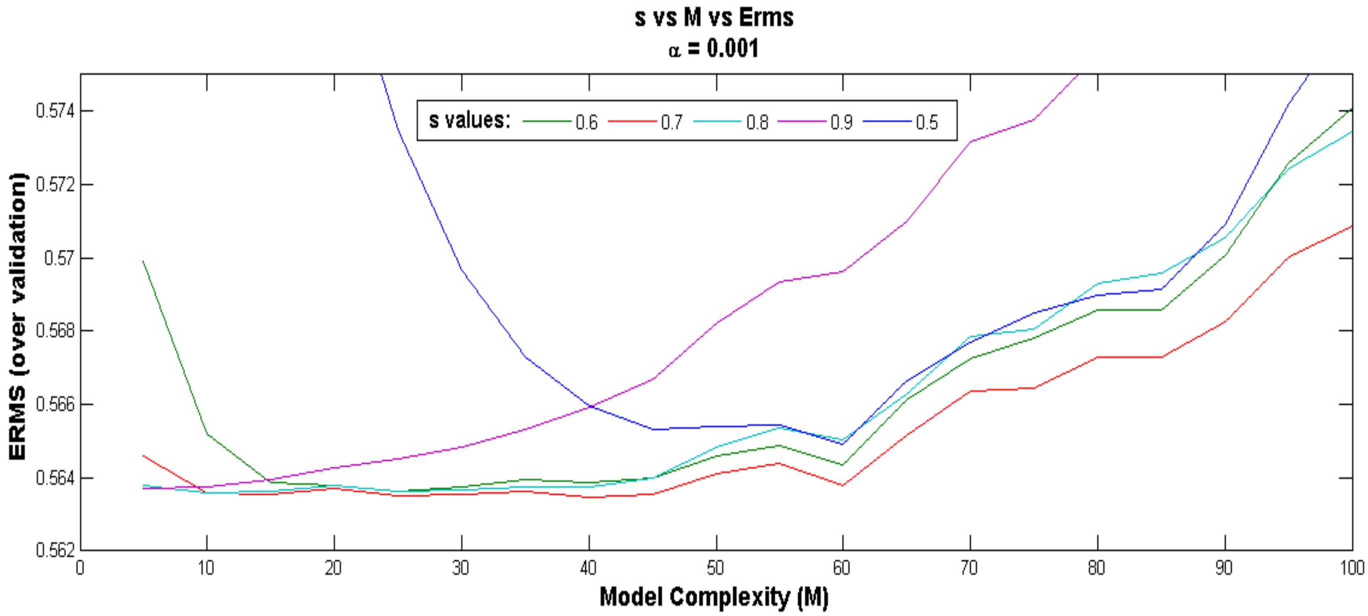
$$E_{RMS}(W) = \sqrt{2E(W*)/N\mathrm{v}}$$

}

}

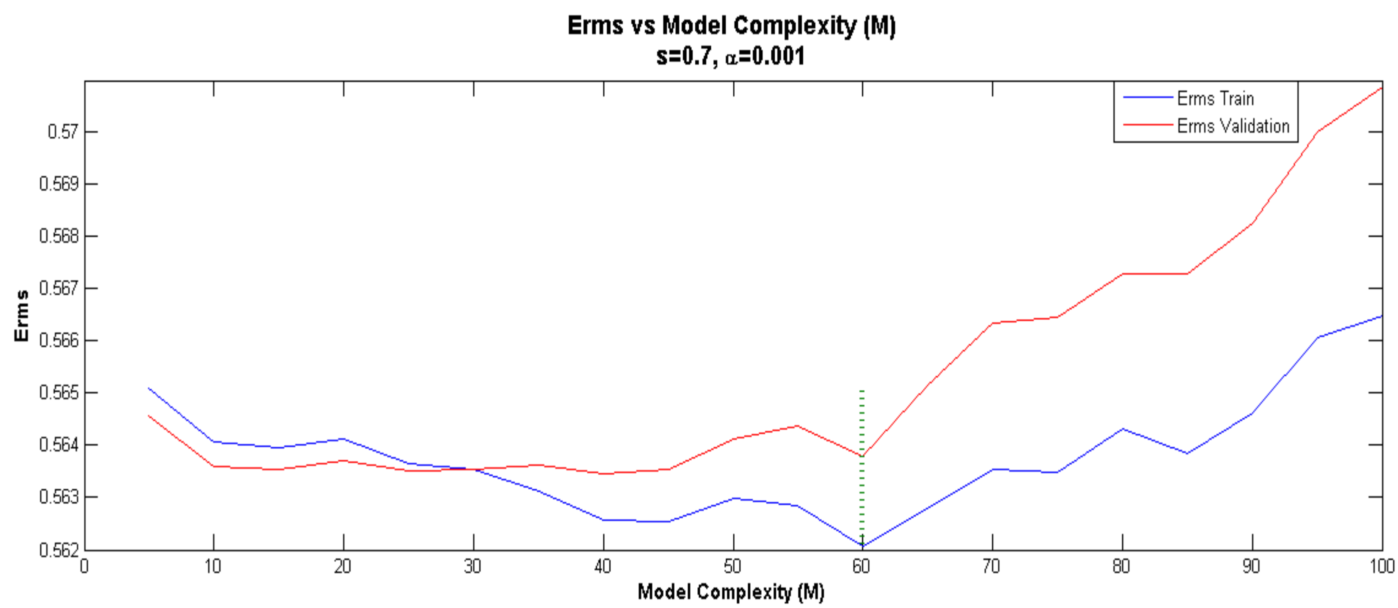
Validating s and M

Following the graph of E_{RMS} of validation set with respect to the values of M and s.



We can see that least possible values of E_{RMS} can be obtained when we select **s_gd=0.7**

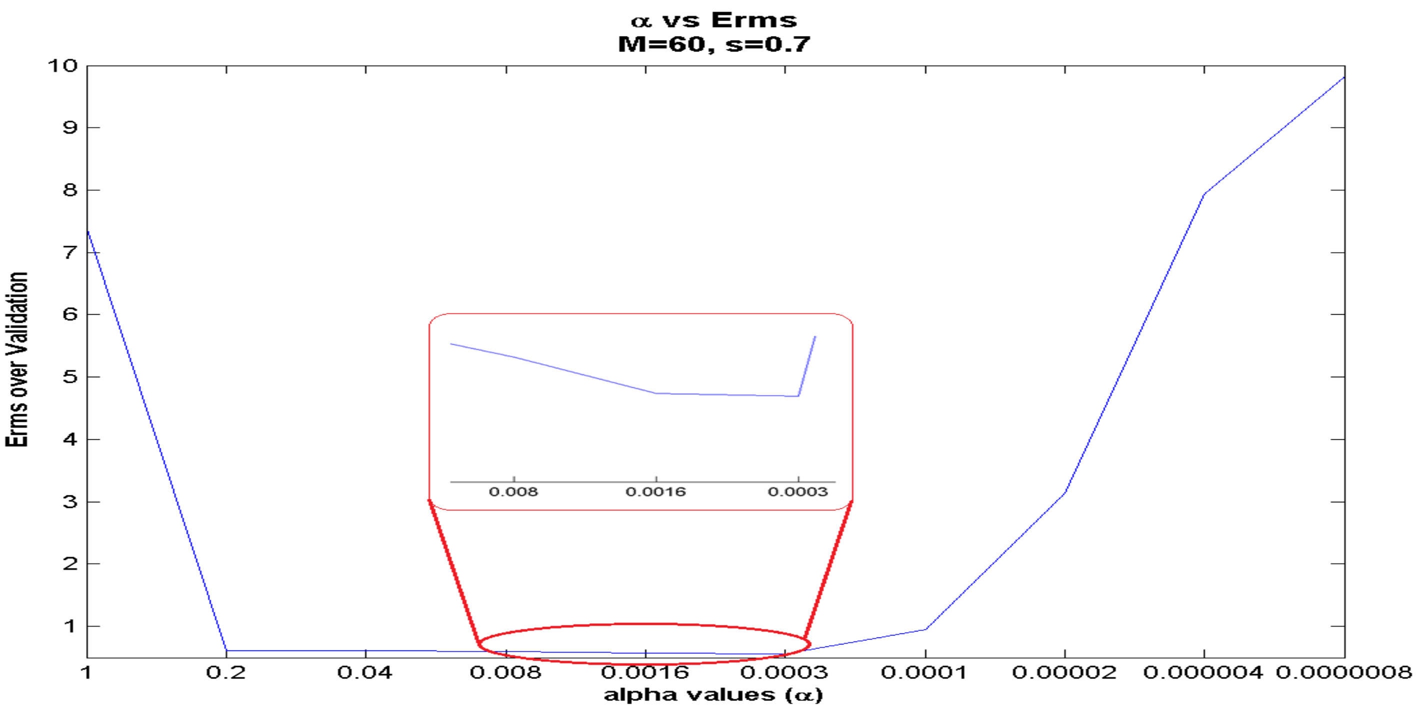
Now we calculate the Model Complexity (M_gd) using the values of above graph when s=0.7. Following is the graph of E_{RMS} of both validation and training sets with respect to the values of M at s=0.7.



From the above graph it can be observed that Erms for Training decreases as M increases and Erms of Validation remains almost constant for some time. After M=60 the Erms for both Training and Validation increases substantially. Hence, we selected **M=60** for this model as M_gd.

Validating α (alpha)

Now, we calculate the most optimal value for alpha (α_gd). We already know the optimal values of s and M. Therefore, we set these values and calculate Erms for validations set for different values of α. Following the graph we obtained:



From the above graph it can be observed that Erms for Validation substantially decreases as α increases. After α=0.001 the Erms for Validation is the least. Hence, we selected **α_gd =0.001** for this model.

Final Outcomes of Gradient Descent Solution

Therefore, after validation, we obtain the following most optimal parameters for our system:

M_gd = 70

s_gd = 0.7

α_gd = 0.001

Erms Train: **0.5621**

Erms Validation: **0.5638**

Erms Test: **0.6153**

Conclusion

The regression using the Closed-Formed Solution provides a slightly better prediction of the target values for the Testset than the Gradient Descent Solution.

For Closed-Formed Solution: Erms Test = 0.5971

For Gradient Descent Solution: Erms Test = 0.6153