CSE 574: Introduction to Machine Learning

Fall 2018

Submitted by: Sai Kalyan Katta (UB Person Number: 50292522)

Functions Description:

Our project starts with the retrieval of data i.e. the target values and the input samples. The functions '**GetTargetVector**' and '**GenerateRawData**' are used to get the target values '**t**' and the input samples '**x**' on which the likelihood function which is defined.

After taking the target values and the input samples, we divide the data into training, validation and the testing sets where the training set contains of 80% of the data, validation set contains 10% of the data and the rest is used as the testing set. Here, the functions '**GenerateTrainingTarget**', '**GenerateTrainingDataMatrix**', '**GenerateValData**', '**GenerateValTargetVector**' divide the original data into the training sets and validation sets.

The '**GenerateBigSigma**', '**GetScalar**', '**GetRadialBasisOut'** are used to calculate the Gaussian basis function using the input values and their means. The '**GenerateBigSigma**' function calculates the ∑ part of the Gaussian radial basis functions in which the input values x and their mean μ are involved. The '**GetScalar**' function calculates the whole scalar part except the exponent of the Gaussian radial basis function Ø(x). The '**GetRadialBasisOut'** function gives the final value of the Guassian radial basis function. The '**GetPhiMatrix**' gives the matrix of the values of the basis functions.

The **'GetWeightsClosedForm'** function returns the weights by using the formula

$$W_{ML} = (\lambda I + Ø^T Ø)^{-1} Ø^T t$$

The '**GetValTest'** function gives the values of the linear regression function y(x,w) which is defined as

$$y(x,w) = w^T Ø(x)$$

The '**GetErms**' gives the Root Mean Square error between the test output and the validation data.

Changing the Hyper Parameters:

Now let us try to change the hyper parameters namely the number of clusters M, the means μ, the ∑ value, λ value and the learning rate and see the change in the performance of the model.

λ value:

Let us change the value of λ by keeping the values of M and the learning rate constant for stochastic gradient descent constant at 8 and 0.01 respectively.

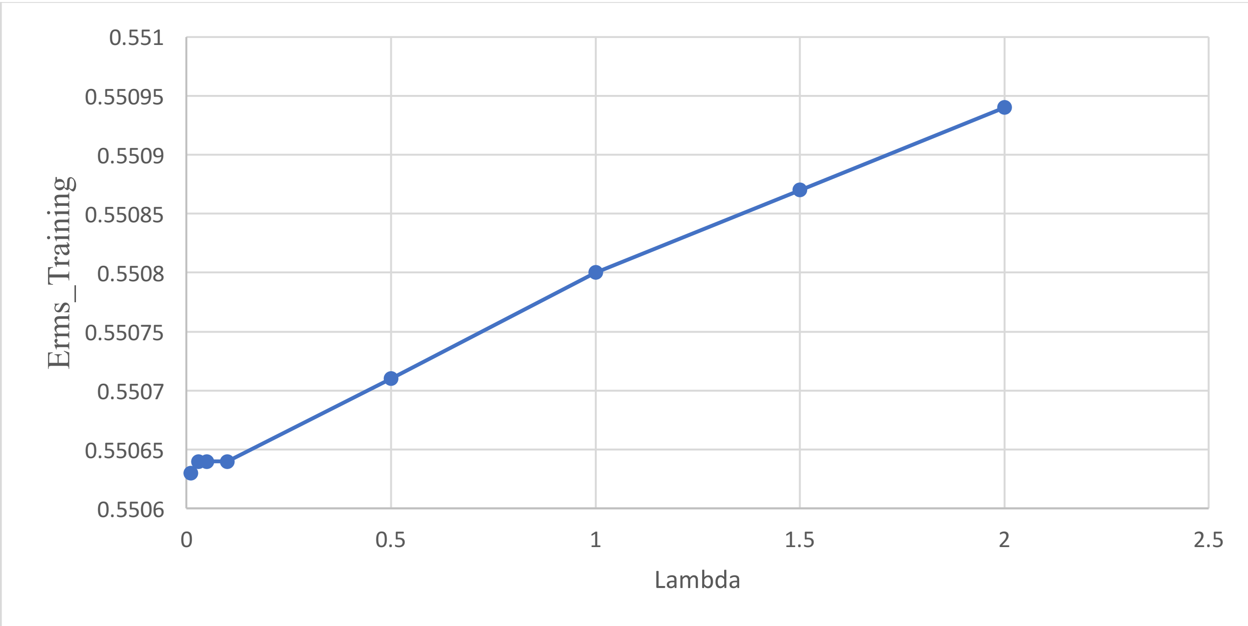
Considering the values of λ between 0.03 and 1, E_rms values of training, validation and the testing sets for the closed form solution and gradient descent solution are

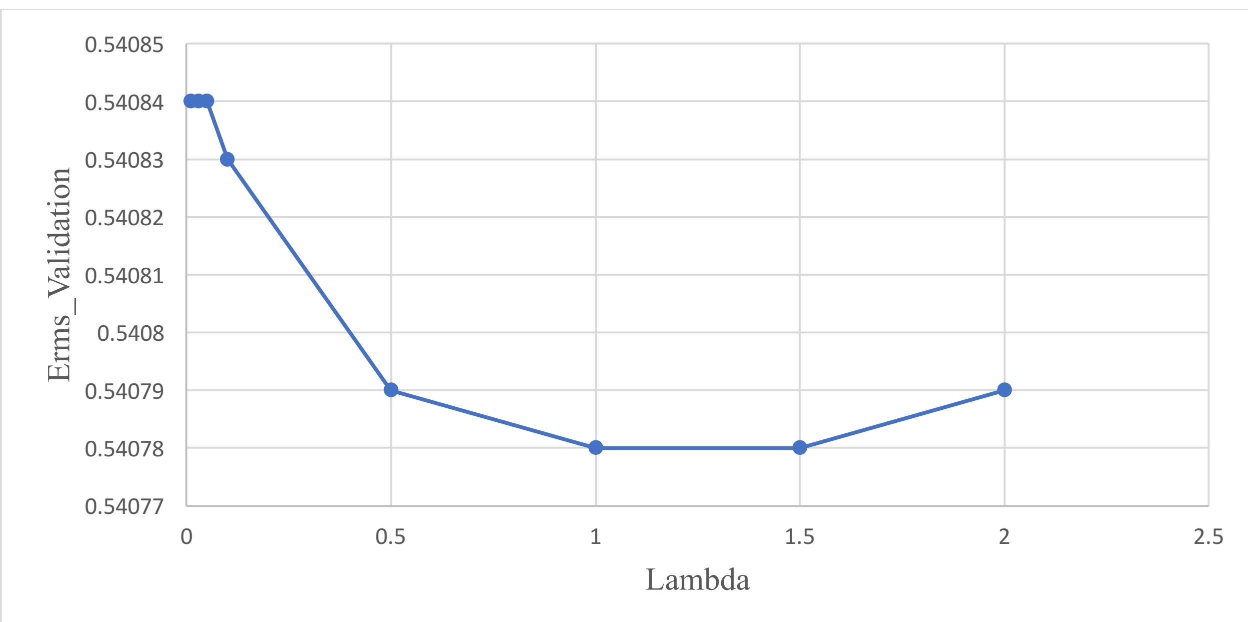
M	λ	E_rms_Training	E_rms_Validation	E_rms_Testing
8	0.01	0.55063	0.54084	0.62927
8	0.03	0.55064	0.54084	0.62924
8	0.05	0.55064	0.54084	0.62921
8	0.1	0.55064	0.54083	0.62914
8	0.5	0.55071	0.54079	0.62884
8	1	0.5508	0.54078	0.62866
8	1.5	0.55087	0.54078	0.62856
8	2	0.55094	0.54079	0.62851

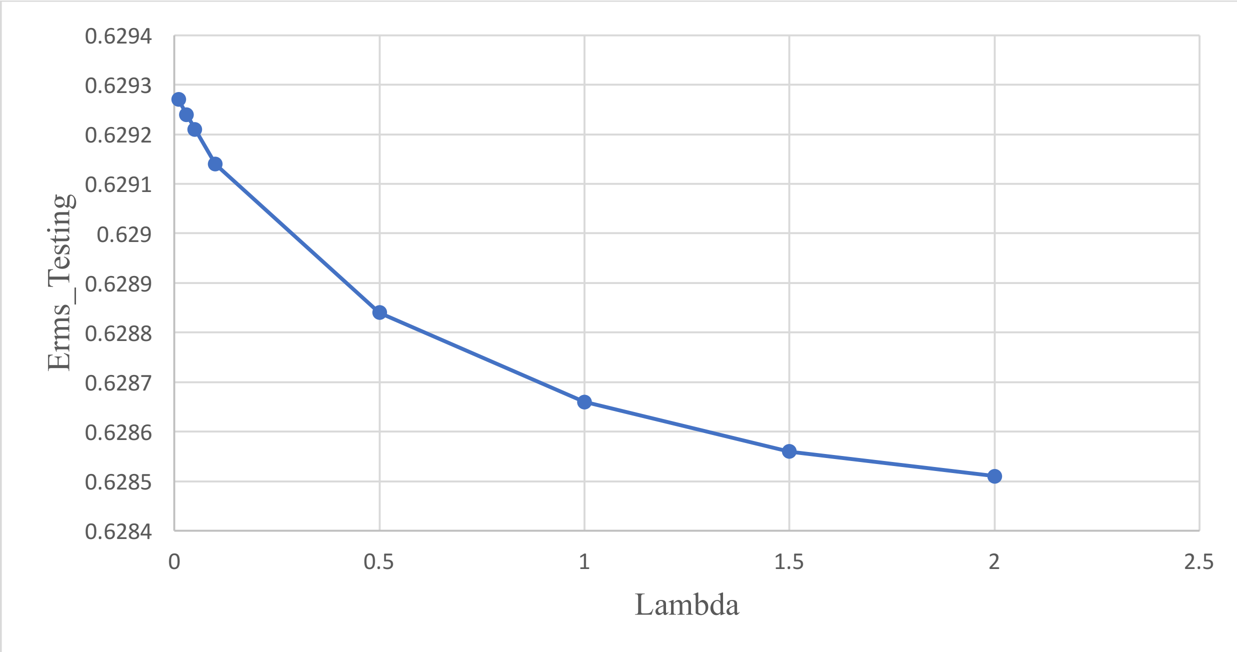
M	λ	Learning rate	E_rms_Training	E_rms_Validation	E_rms_Testing
8	0.01	0.01	24.89795	24.89783	24.80138
8	0.03	0.01	23.14134	23.05836	23.05836
8	0.05	0.01	21.43034	21.35399	21.27152
8	0.1	0.01	17.68279	17.62094	17.55341
8	0.5	0.01	3.65412	3.63158	3.61987
8	1	0.01	0.65521	0.64771	0.72917
8	1.5	0.01	0.55135	0.54124	0.6252
8	2	0.01	0.55102	0.5408	0.62446

Graphs:

For closed-form solution:

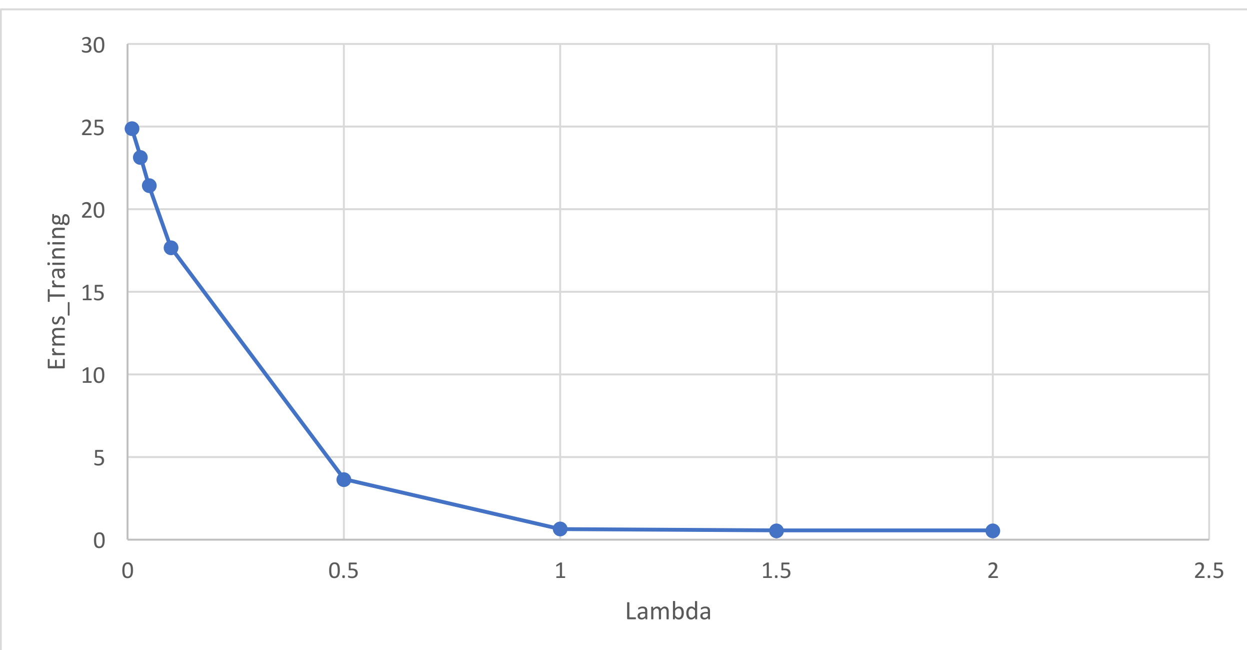


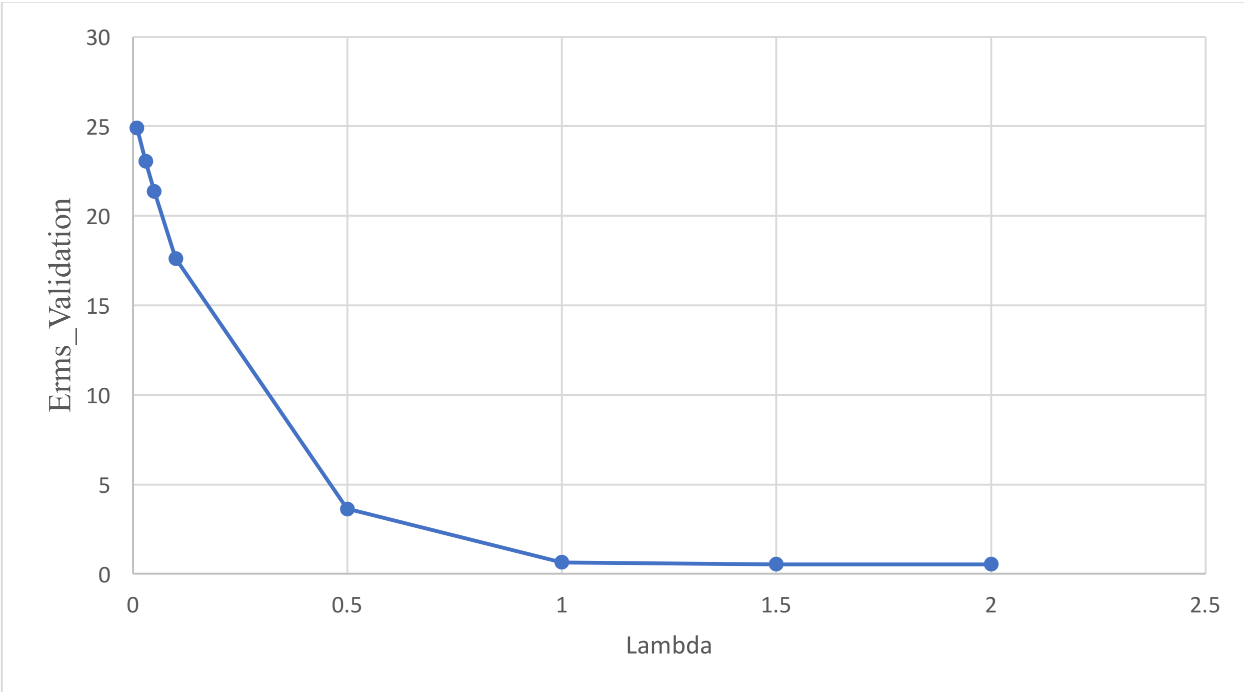


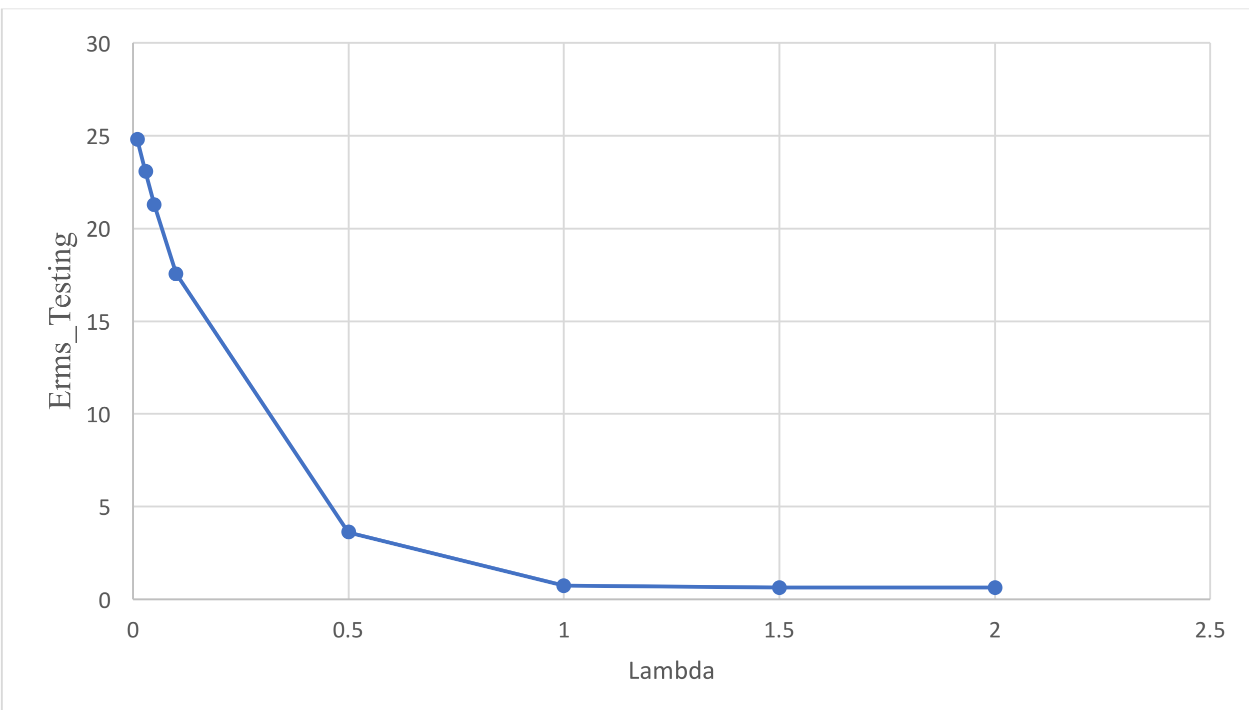


With the value of Lambda increasing the E_rms_training value keeps increasing whereas both the E_rms_validation and the E_rms_testing keep decreasing but by a very less difference. So, for the closed form solution it is better to take higher value for the lambda value.

For stochastic gradient descent solution:







With the value of Lambda increasing, the E_rms_training, E_rms_validation and the E_rms_testing values keep decreasing with a very large difference. So, for the stochastic gradient descent solution it is better to take higher value for the lambda value.

No. of clusters:

Let us change the value of M (no. of clusters) by keeping the values of λ for closed form solution, stochastic gradient solution and the learning rate for stochastic gradient descent constant at 0.03, 2 and 0.01 respectively.

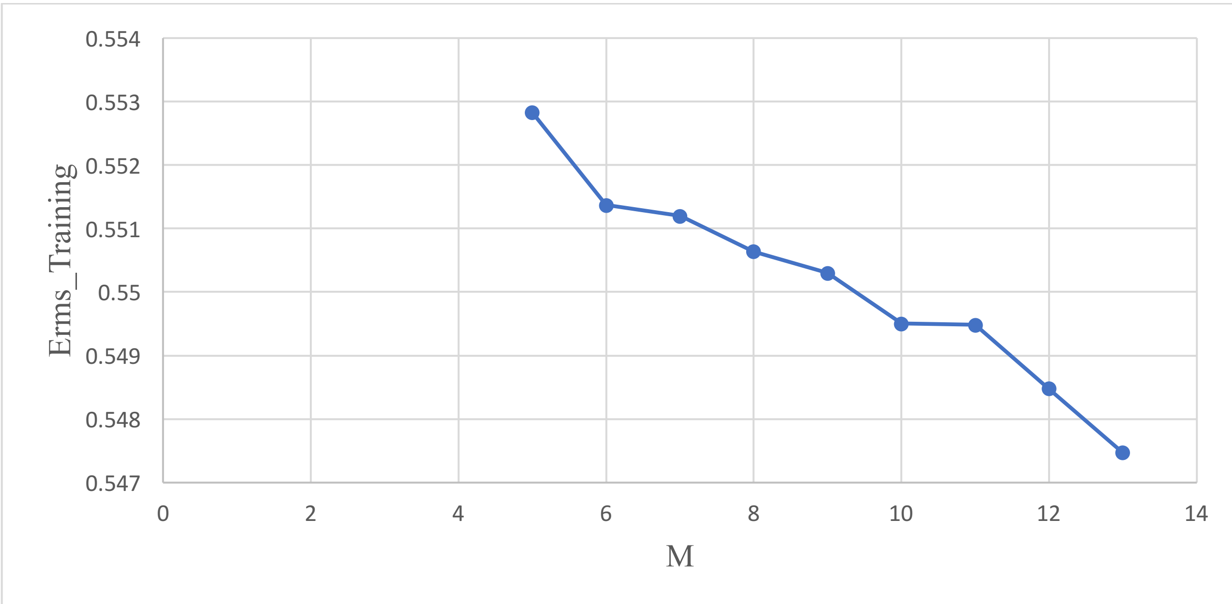
Considering the values of M between 5 and 13 the E_rms values of training, validation and the testing sets for the closed form solution and gradient descent solution are

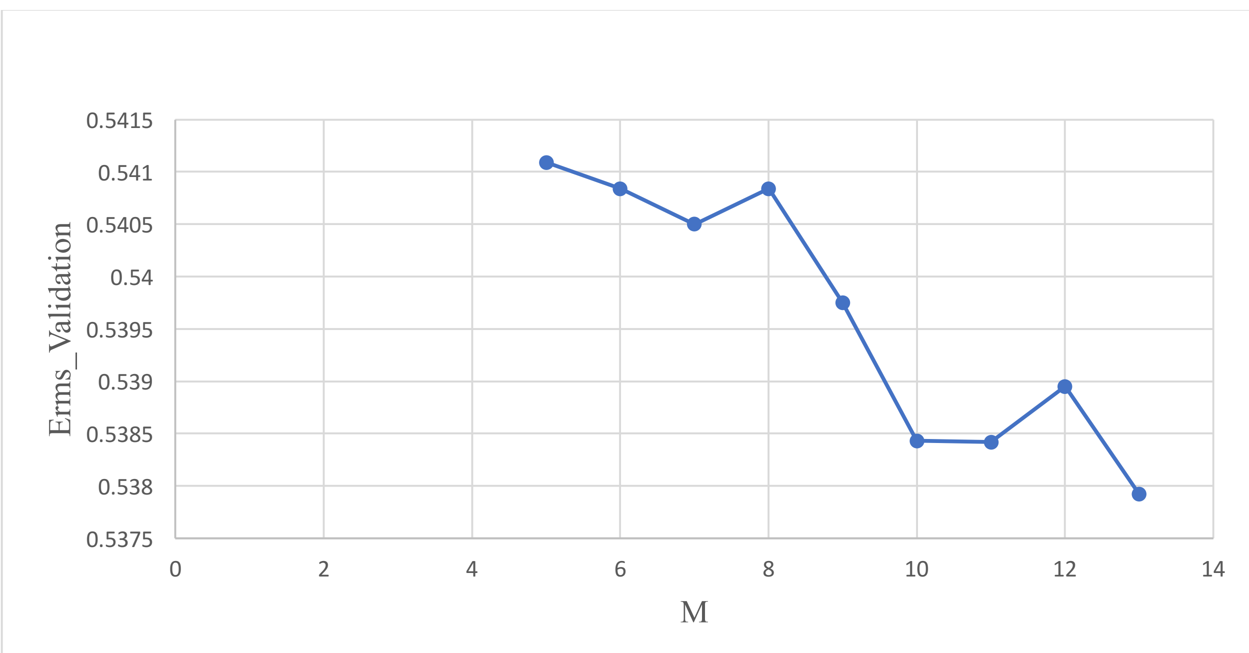
M	λ	E_rms_Training	E_rms_Validation	E_rms_Testing
5	0.03	0.55283	0.54109	0.63113
6	0.03	0.55137	0.54084	0.6284
7	0.03	0.5512	0.5405	0.6289
8	0.03	0.55064	0.54084	0.62924
9	0.03	0.5503	0.53975	0.62923
10	0.03	0.5495	0.53843	0.62798
11	0.03	0.54948	0.53842	0.62798
12	0.03	0.54848	0.53895	0.62929
13	0.03	0.54747	0.53792	0.62772

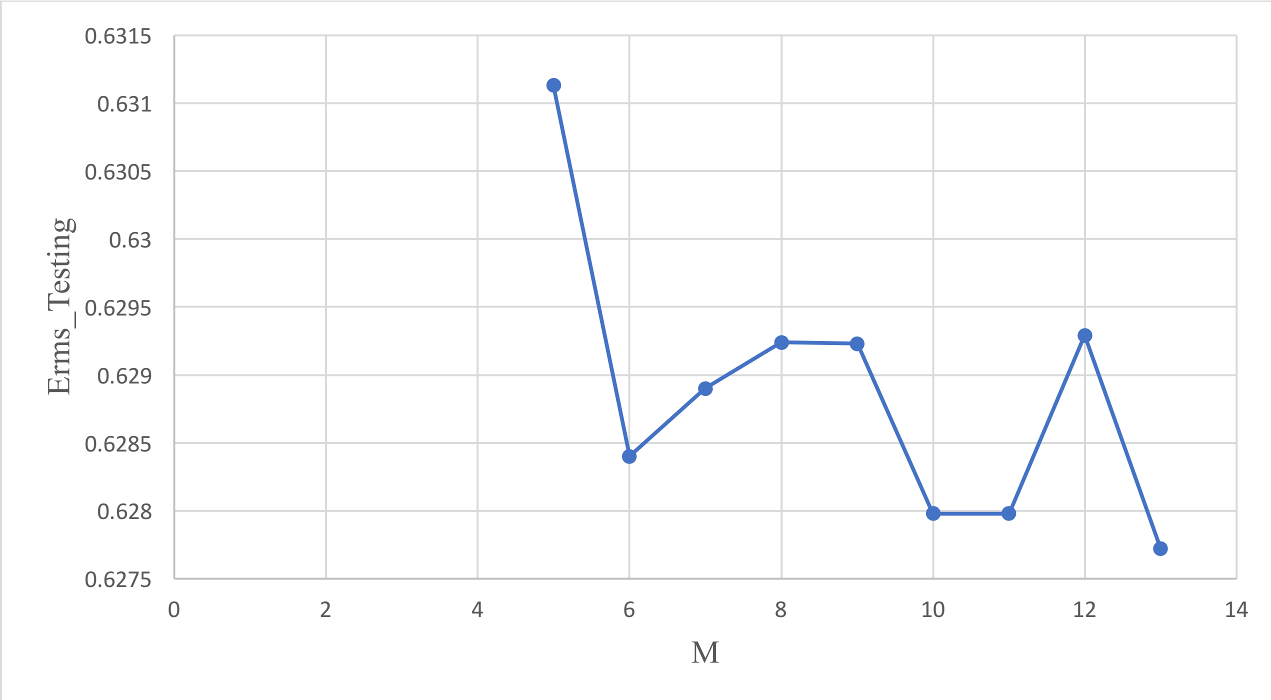
M	λ	Learning rate	E_rms_Training	E_rms_Validation	E_rms_Testing
5	2	0.01	0.55324	0.5414	0.62702
6	2	0.01	0.55166	0.54097	0.6245
7	2	0.01	0.55133	0.54047	0.62482
8	2	0.01	0.55068	0.54084	0.62489
9	2	0.01	0.55039	0.53983	0.62482
10	2	0.01	0.54964	0.53846	0.62372
11	2	0.01	0.54967	0.53847	0.62365
12	2	0.01	0.54867	0.53917	0.62496
13	2	0.01	0.54757	0.53805	0.62369

Graphs:

For closed-form solution:

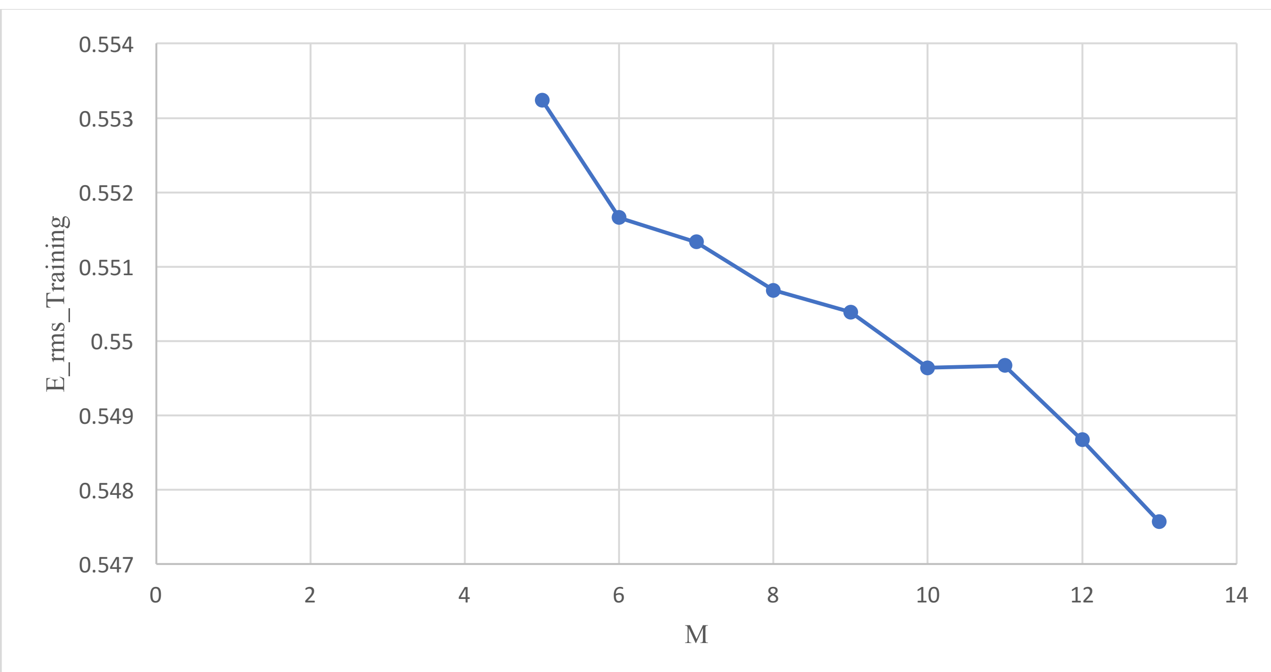


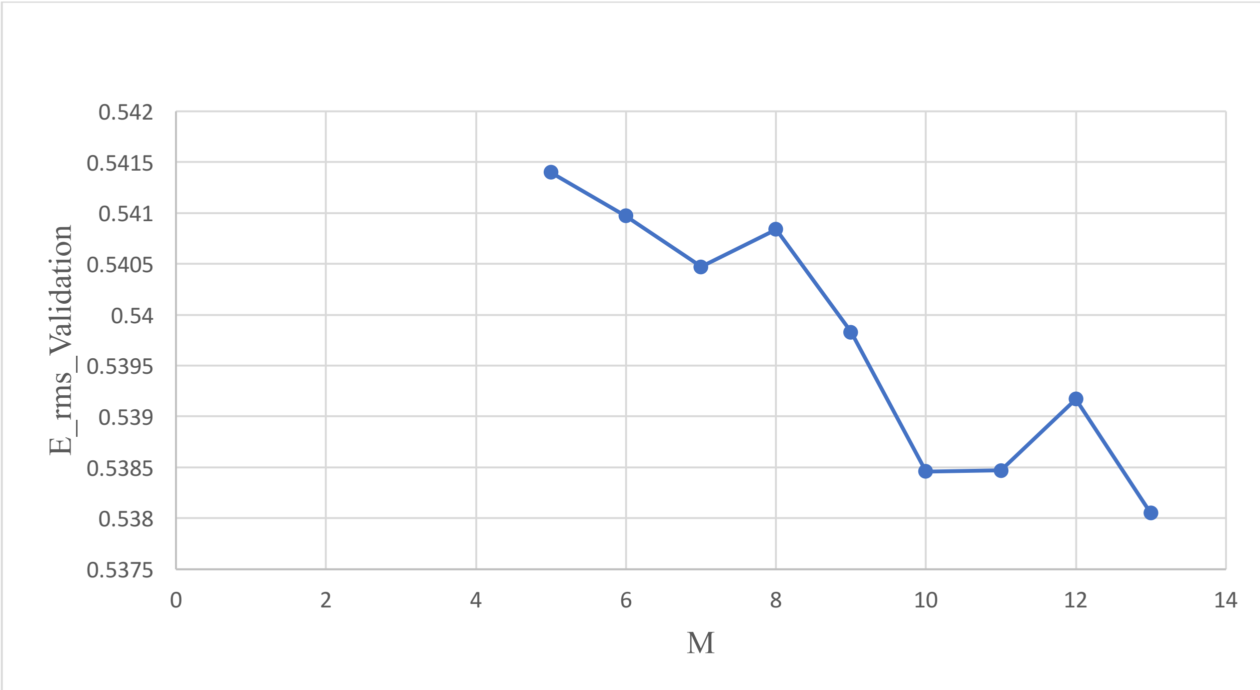


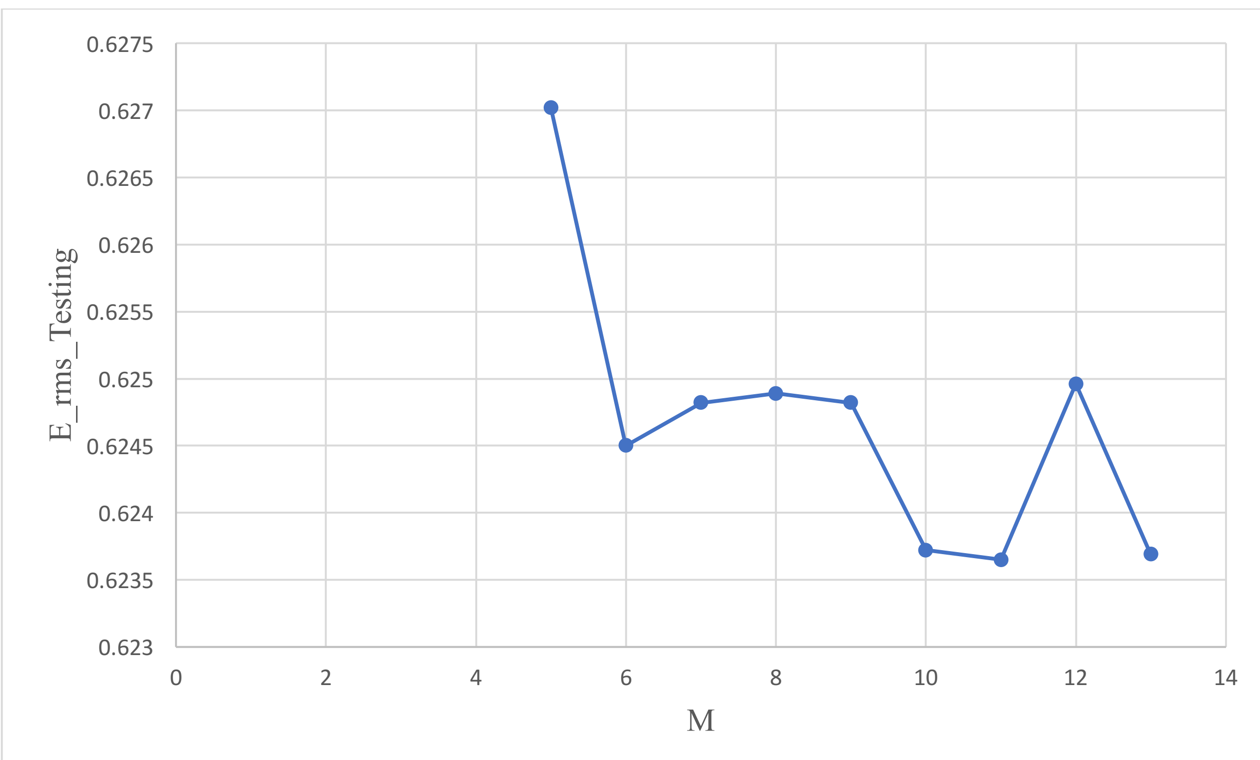


With the value of M increasing the E_rms_training, E_rms_validation and the E_rms_testing keep decreasing but by a very less difference. So for the closed form solution it is better to take the higher value for the number of clusters M.

For Gradient Descent solution:







With the value of M increasing the E_rms_training, E_rms_validation and the E_rms_testing keep decreasing but by a very less difference. So, for the gradient descent solution it is better to take the higher value for the number of clusters M.

Learning Rate:

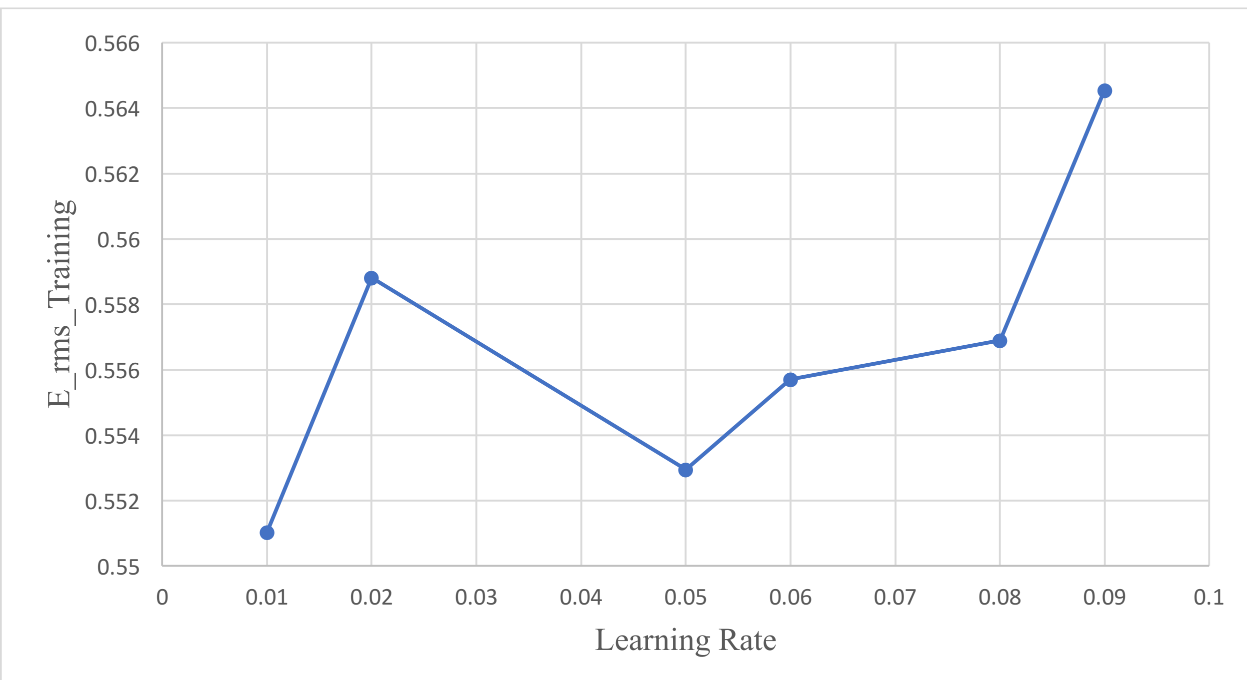
Let us change the value of learning rate by keeping the values of M and the lambda constant at 8 and 2 respectively for stochastic gradient descent solution.

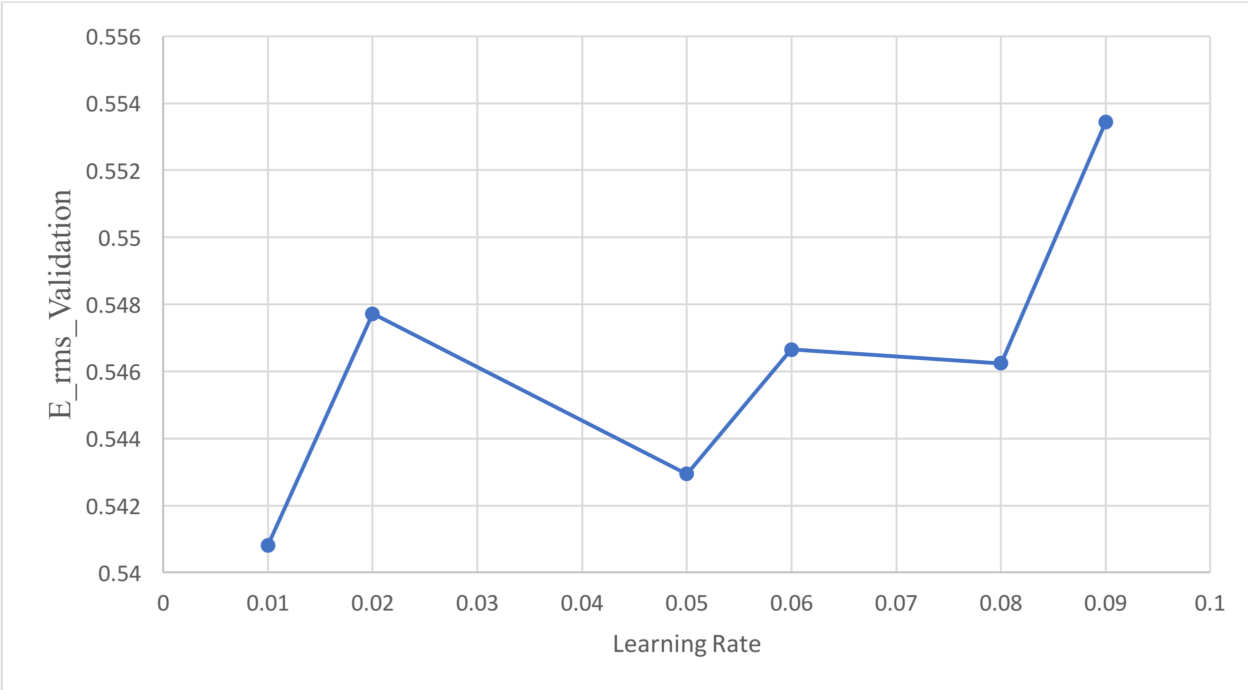
Considering the values of learning rate between 0.01 and 1, E_rms values of training, validation and the testing sets for the closed form solution and gradient descent solution are

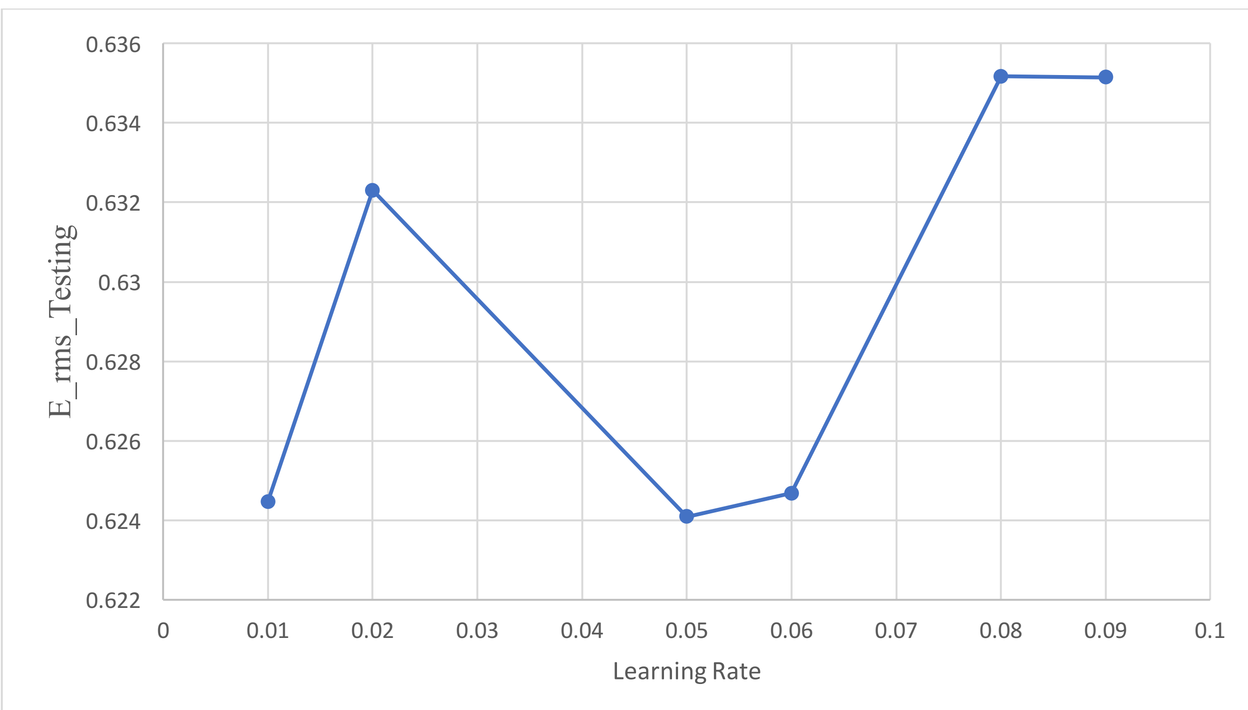
λ	Learning rate	Erms_Training	Erms_Validation	Erms_Testing
2	0.01	0.55102	0.5408	0.62446
2	0.02	0.55881	0.54772	0.6323
2	0.05	0.55295	0.54294	0.62409
2	0.06	0.5557	0.54665	0.62468
2	0.08	0.55689	0.54624	0.63517
2	0.09	0.56454	0.55343	0.63515

Graphs:

For Gradient Descent solution:







With the value of learning rate increasing the E_rms_training, E_rms_validation and the E_rms_testing keep increasing. So, for the gradient descent solution it is better to take the lower value for the learning rate.

Conclusion:

Considering the values of the λ value for closed-form solution as 0.03 and the number of clusters as 13, the E_rms values of the training, validation and testing sets were obtained as 0.54747, 0.53792 and 0.62772.
Considering the values of the number of clusters, λ value and learning rate as 13, 2 and 0.01 respectively, the E_rms values of training, validation and testing sets for the stochastic gradient descent solution are 0.54757, 0.53805 and 0.62369.