### Q1.

The 2X2 matrix of mean square error for the training and test for 100 and 10 training test points is shown below.

From the below table it is pretty clear that the training and the test error depend on the training set size.

The training set error for 10 data point is less as it can train better on the small data set, where as on the 100 data points it has a slightly higher error as there are more points to train over.

The test error is slightly higher on the 10 data points as the trained model calculated from the training set is more generalised to the variation in the training set, as compared to the 100 training points, where the model is better generalised as compared to the 10 training points.

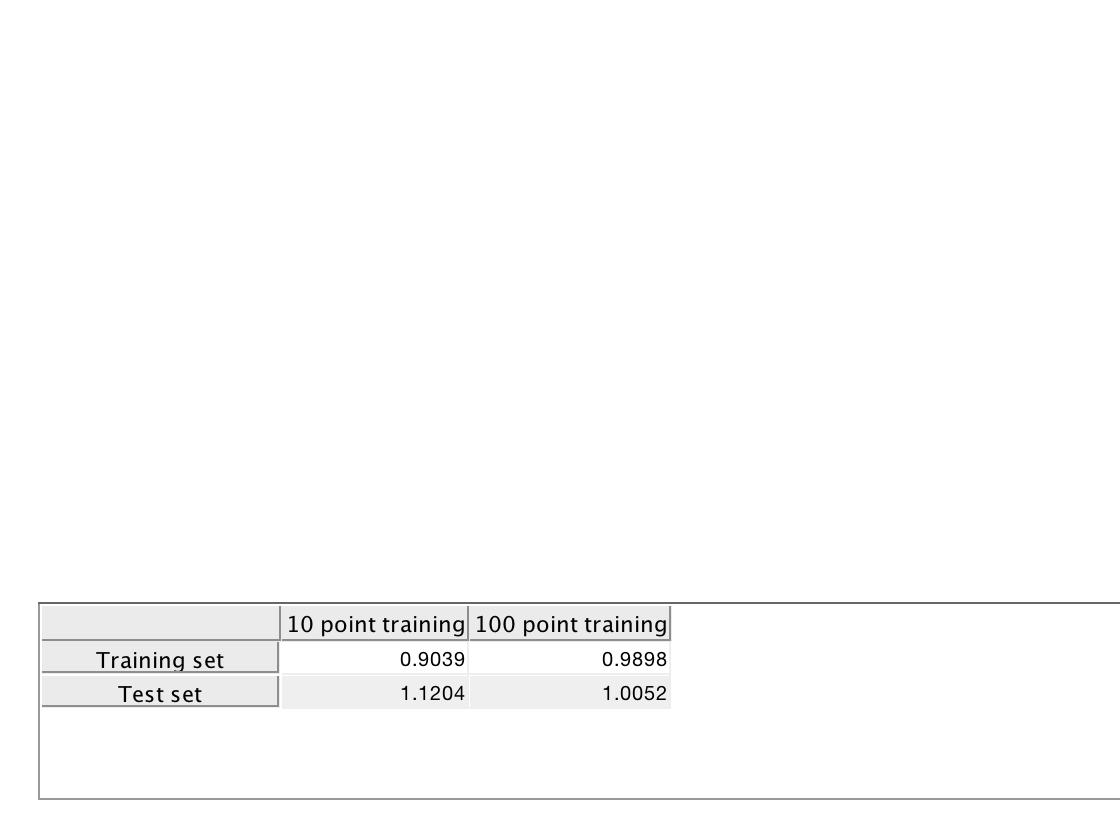
Because of this change in how the model is getting trained the test error for higher number of training set is lesser than that on the smaller set.

Fig 1 : figure of the 2x2 matrix for training and test error on 10 and 100 data points

### Q2.

The 2X2 matrix of mean square error for the training and test for 100 and 10 training test points over 10 dimensions is shown below.

From the below table it is pretty clear that the training and the test error depend on the training set size and dimension.

The training set error for 10 data point on the 10 dimensions is almost zero. As the model was able to fit to the small data provided almost perfectly. Where as for a larger set the training error is more as there are more number of variations in the data and the model was not able to generalise

As well as it did on a smaller dataset

The test error is very high for the 10 datapoint model as the model has overfit over the training set and is not able to predict better. Where as for the 100 datapoint model the test error is pretty close compared to the 10 point error as the model was able to generalise better over the training set better and was able to do better predictions over the test set.

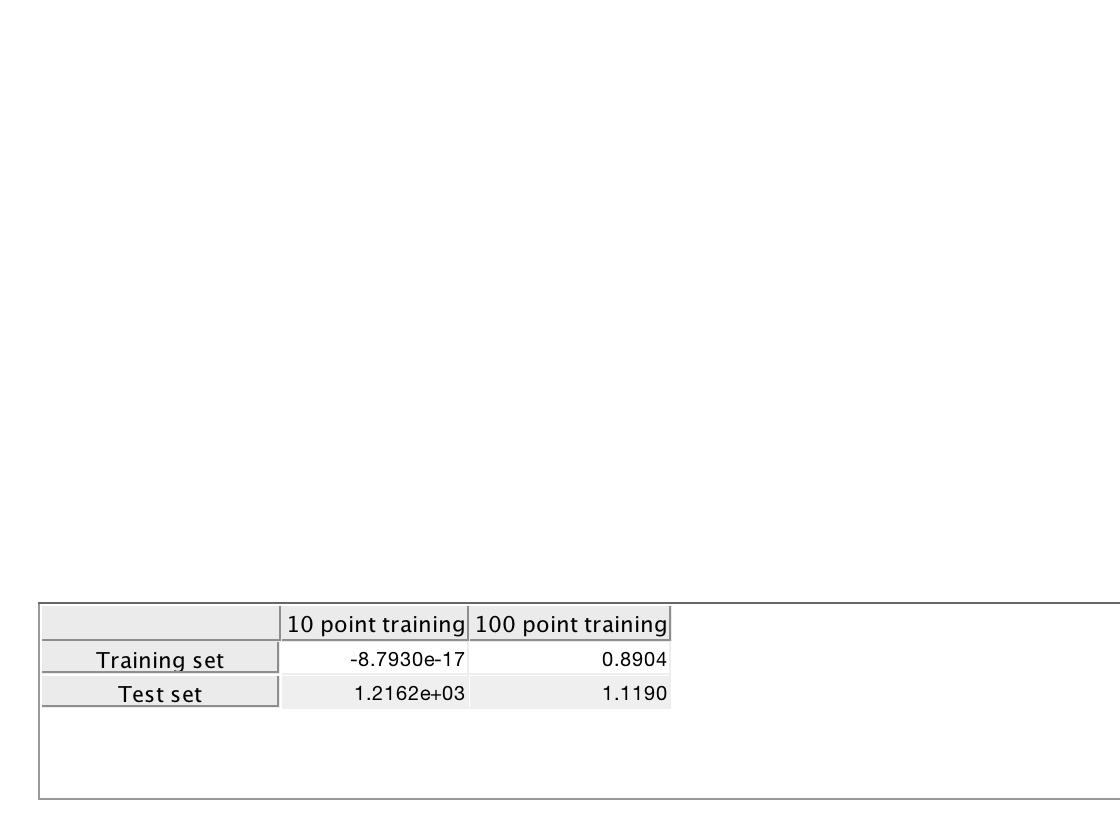
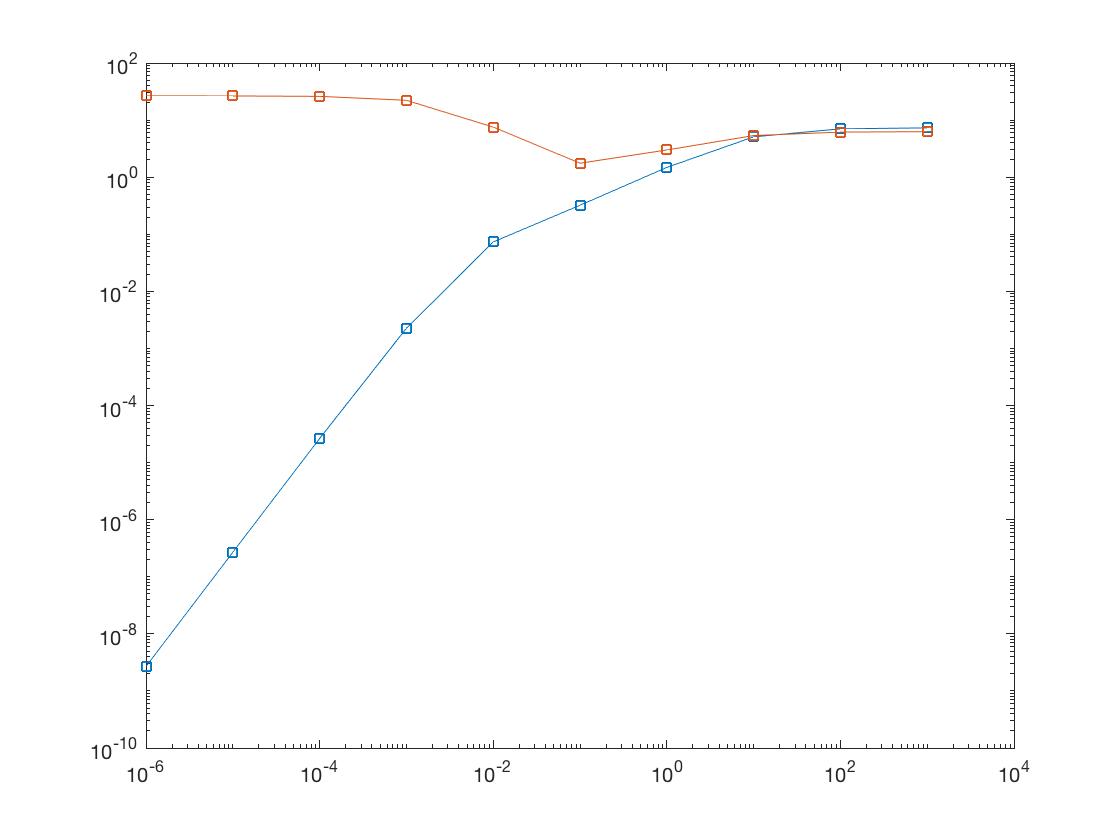
Because of this change in how the model is getting trained the test error for higher number of training set is lesser than that on the smaller set.

Fig 2 : figure of the 2x2 matrix for training and test error on 10 dim data

### Q3.

Fig 3 : the above figure gives the solution for part a and b of question 3

### Q4.

Fig 4.1: train & test error on 100 training points Fig 4.2: train and test error on 10 training points

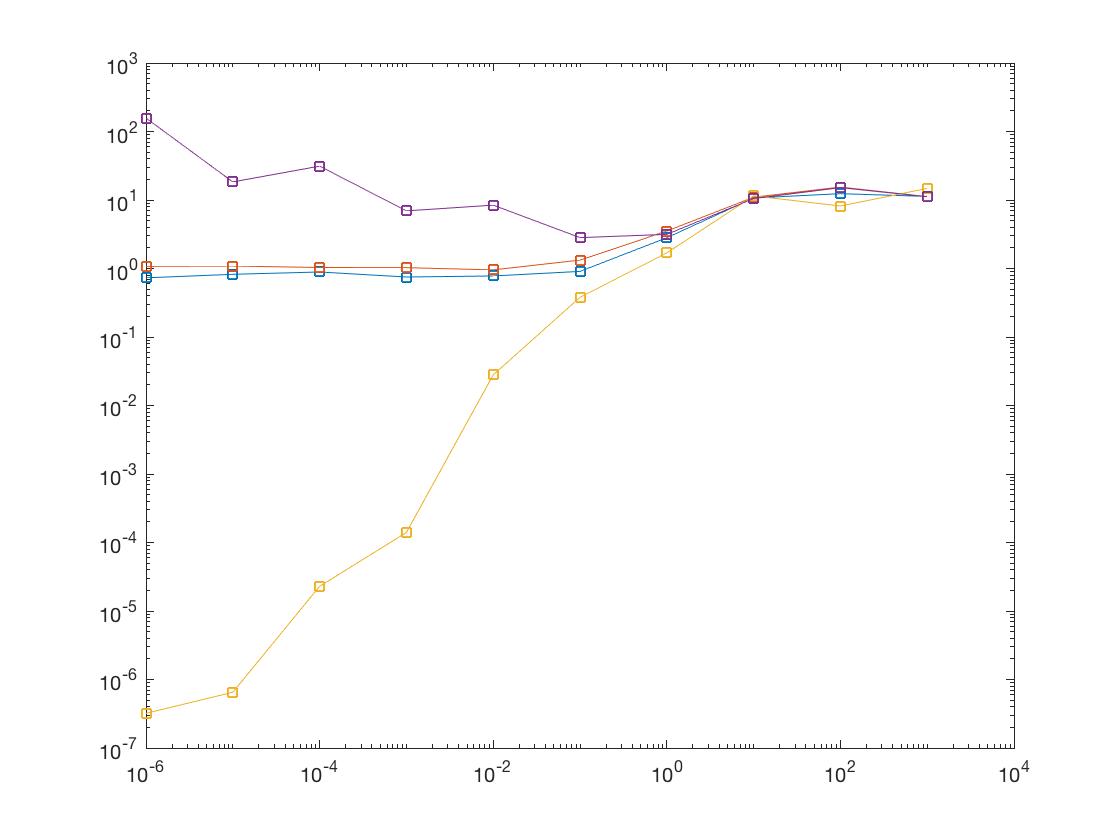
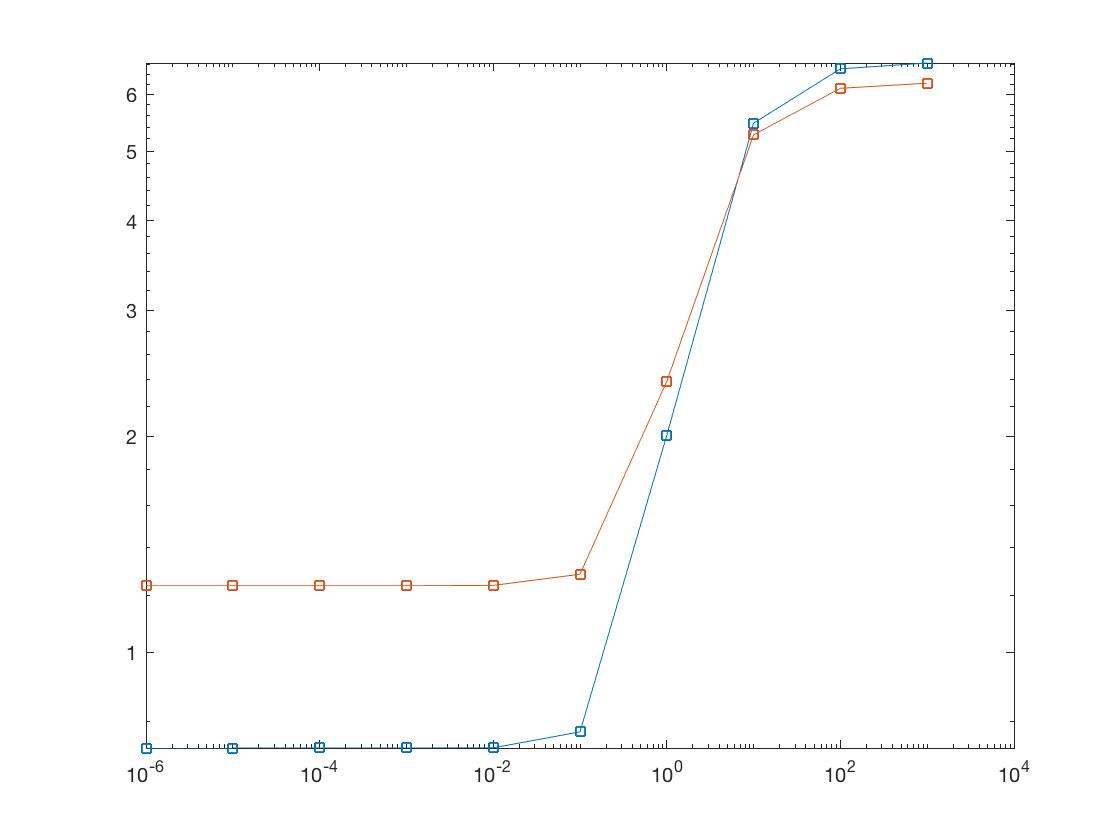


Fig 4.3 : train and test errors calculated 200 times and averaged over 10 and 100 datapoint

All the above images are plotted on a log scale.

From the above graphs it is very clear that the least test error can be achieved by choosing the least train error achieved from a range of regularisation parameter γ.

In the above example we have chosen a range of 10-6  to 10+3 . And we achieved a test and train errors as shown below for the range of regularisation parameter mentioned above.

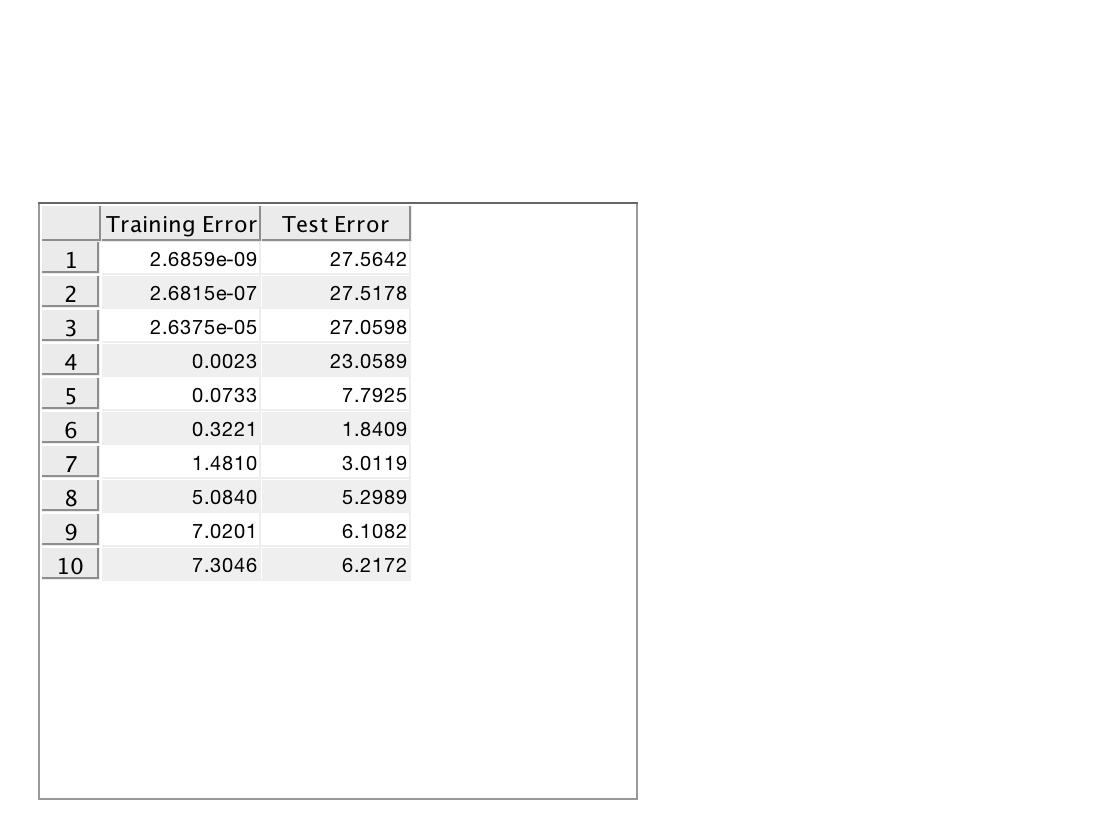
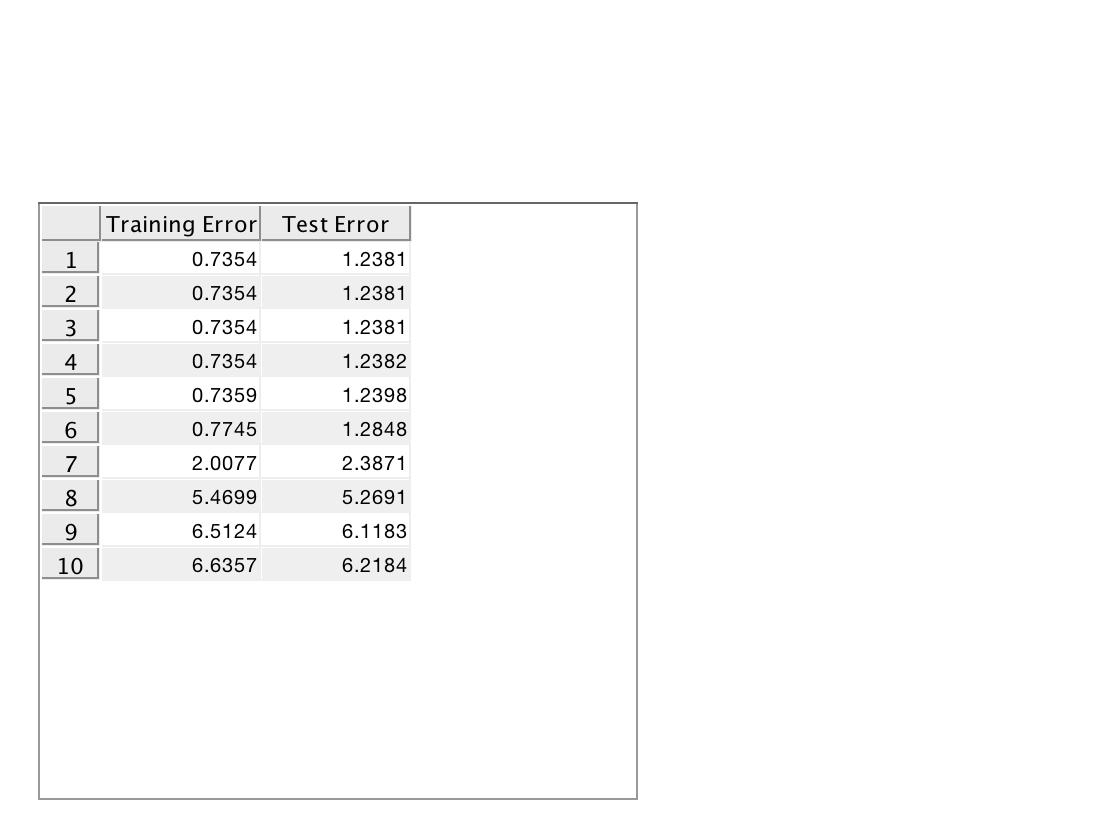


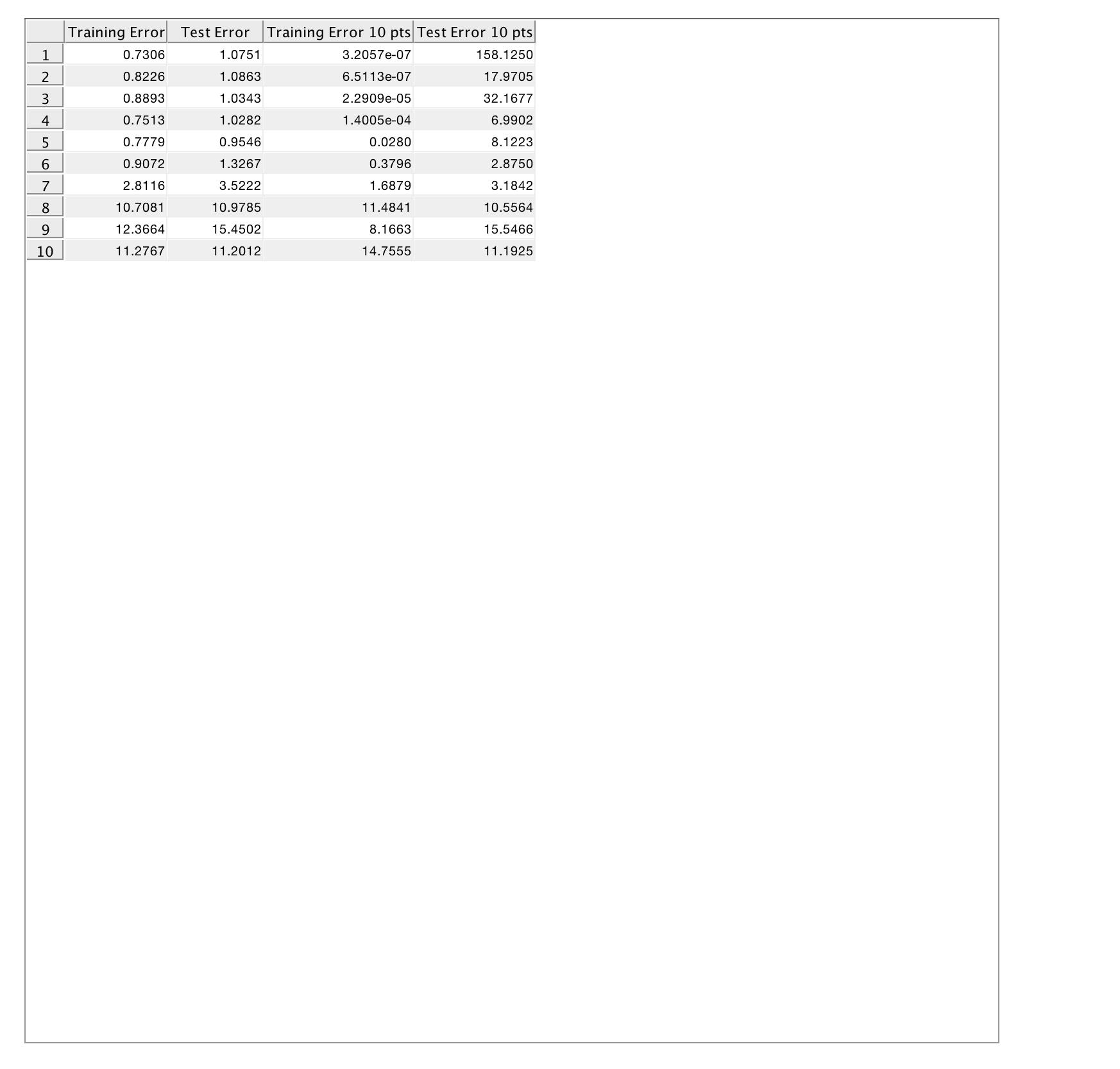
Fig 4.4 : training and test error on 100 points fig 4.5 : training and test error on 10 points

Fig 4.6 : train and test error for 100 and 10 points averaged over 200 times

From the above 3 figures it is clear that as the number of data points increase the lower the train error the lower is the test error. But if we take a lower number of data points there is still an issue of overfitting but not as drastic as the previous method. The overfit test error is far less compared to previous results

### Q5.

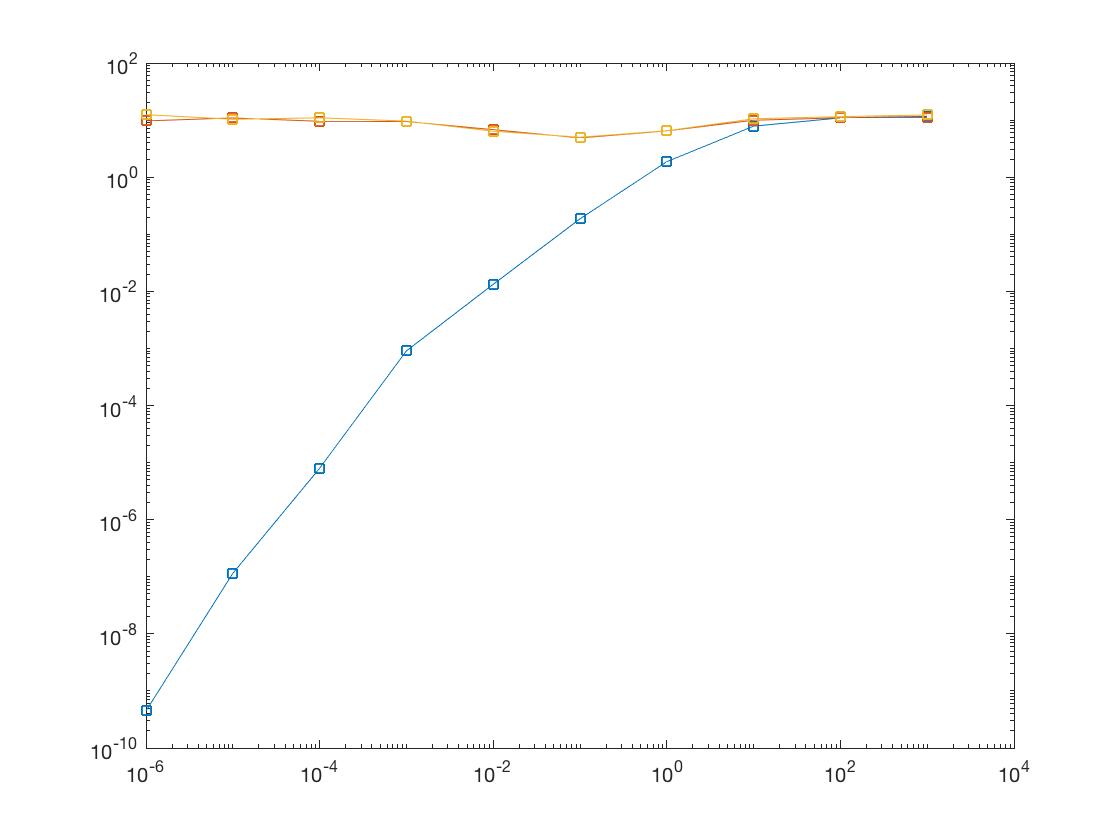
a/b

Fig 5.1: train, test, val errors for 100 data points Fig 5.2 : train, test, val error for 10 data points

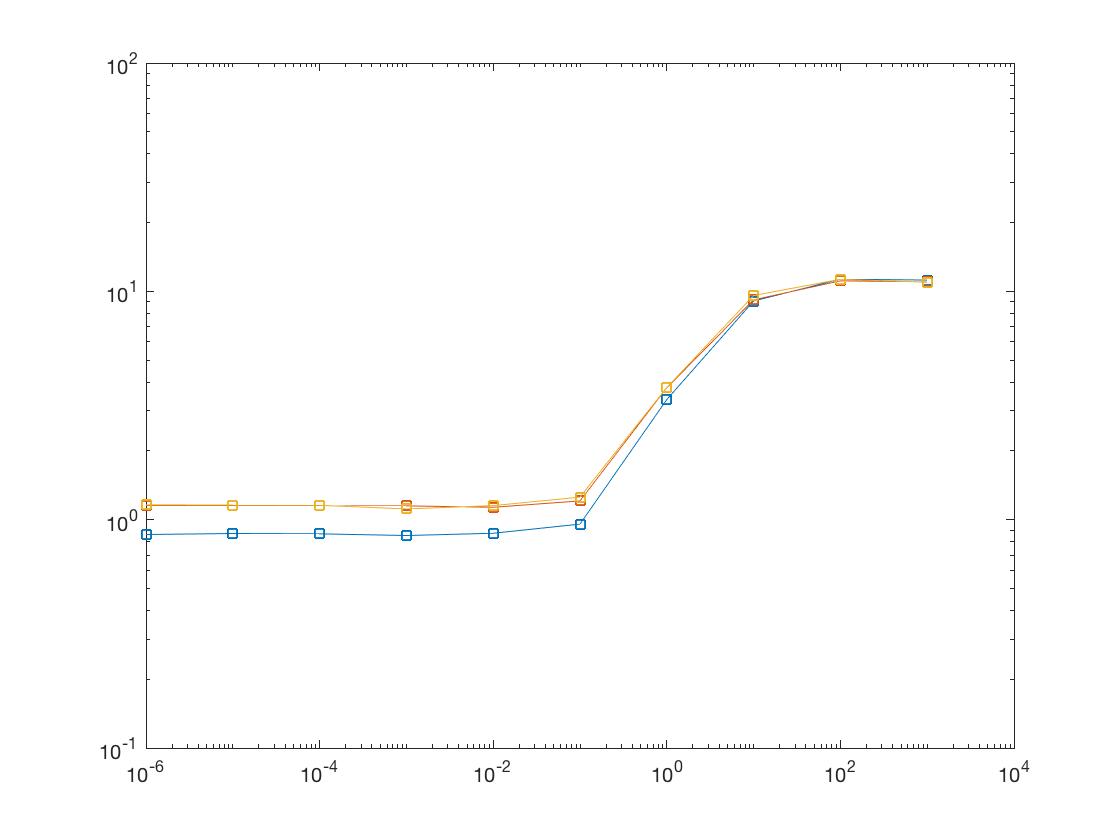
From the above graphs the min validation error for both 100 and 10 training sets were chosen and the corresponding regularisation parameter was selected and the following test errors were calculated.

|  |  |
| --- | --- |
|  | Test Error |
| 100 points | 1.1144 |
| 10 points | 3.8516 |

The validation errors were pretty close to the test errors so selecting the least validation error helped us to train the model to get the lowest possible test error

c.

|  |  |
| --- | --- |
|  | Mean Gamma |
| 100 points | 0.0383 |
| 10 points | 102.0421 |

From the above table we can see that the larger dataset like the 100 training points with 10 dimensions has a lower regularisation parameter as compared to the 10 dimensional 10 data points

This is because the 10 data point training model will overfit and have a high validation error in the beginning and as the regularisation parameter is increased the validation error reduces to a minimum value and hence the test error, so they have a high regularisation parameter

In the 100 data point 10 dimensional training, the lower regularisation parameter leads to a lower validation error hence lower mean regularisation parameter

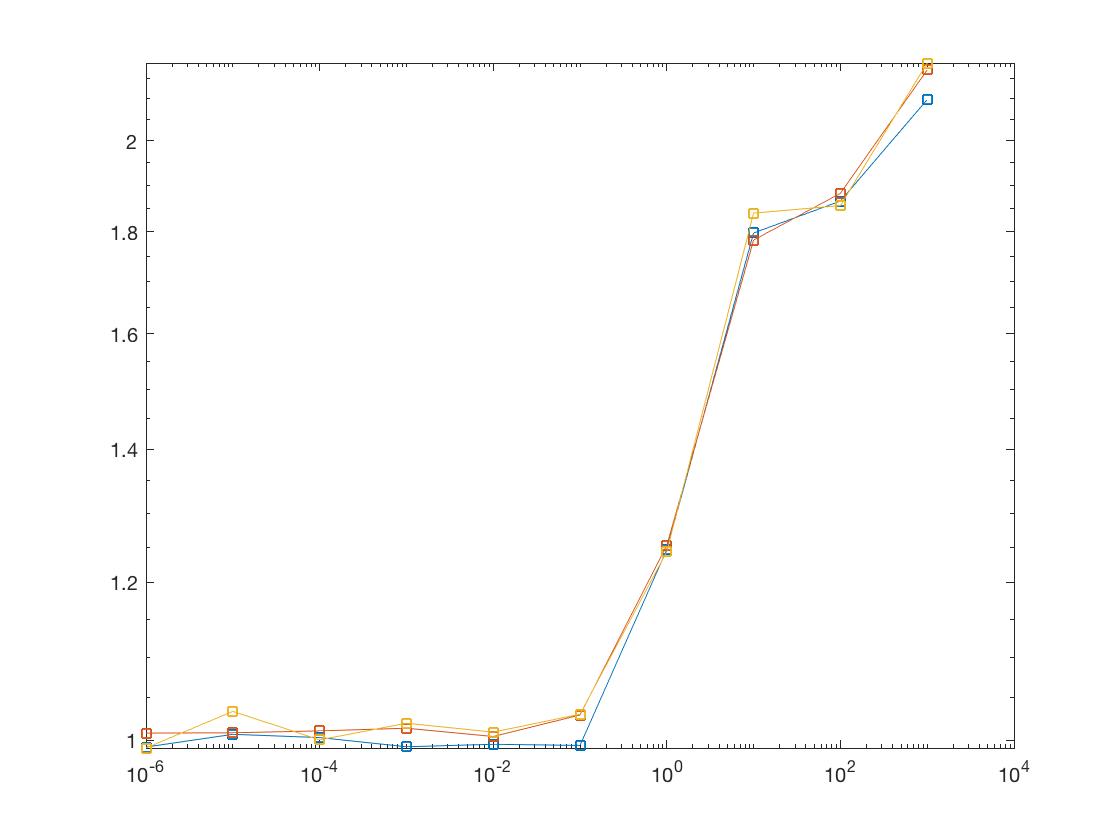
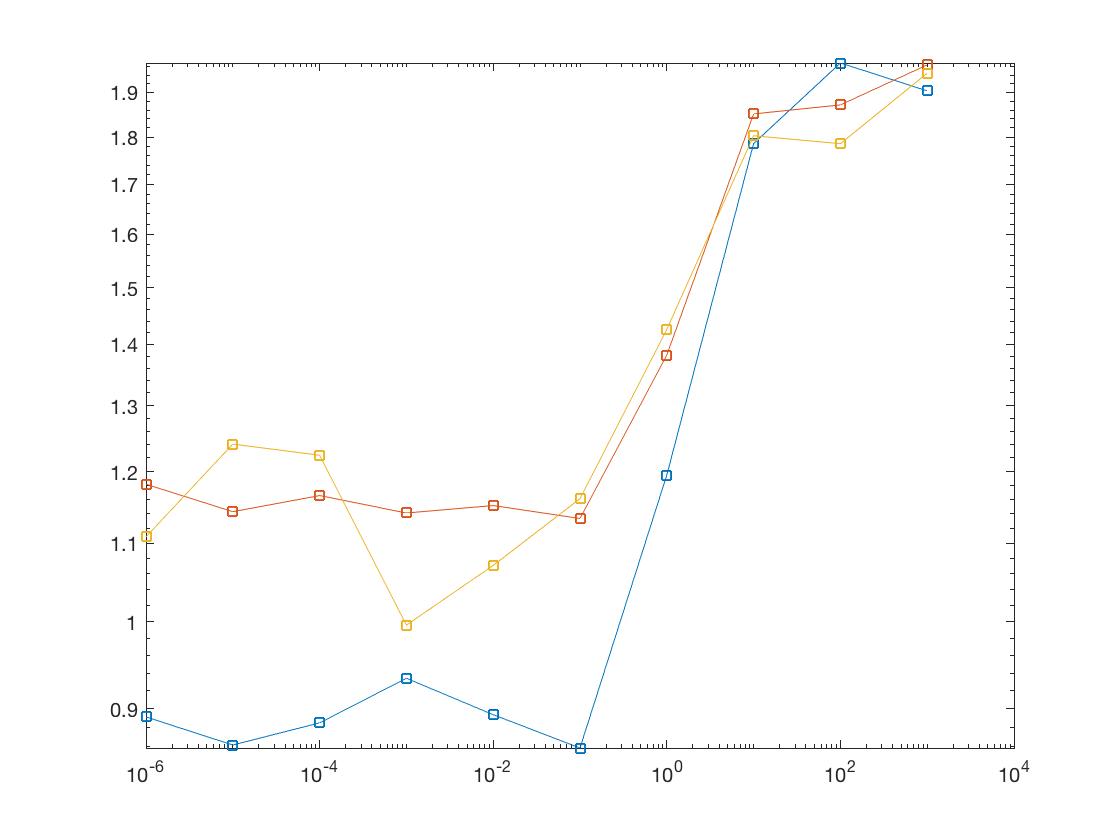
d.

Fig 5.1: train, test, val errors for 1D-100 data points. Fig 5.2 : train, test, val error for 1D-10 Data points

|  |  |
| --- | --- |
|  | Test Error |
| 100 points | 1.0064 |
| 10 points | 1.1296 |

Similar to problem a/b, the validation errors were pretty close to the test errors so selecting the least validation error helped us to train the model to get the lowest possible test error

### Q6.

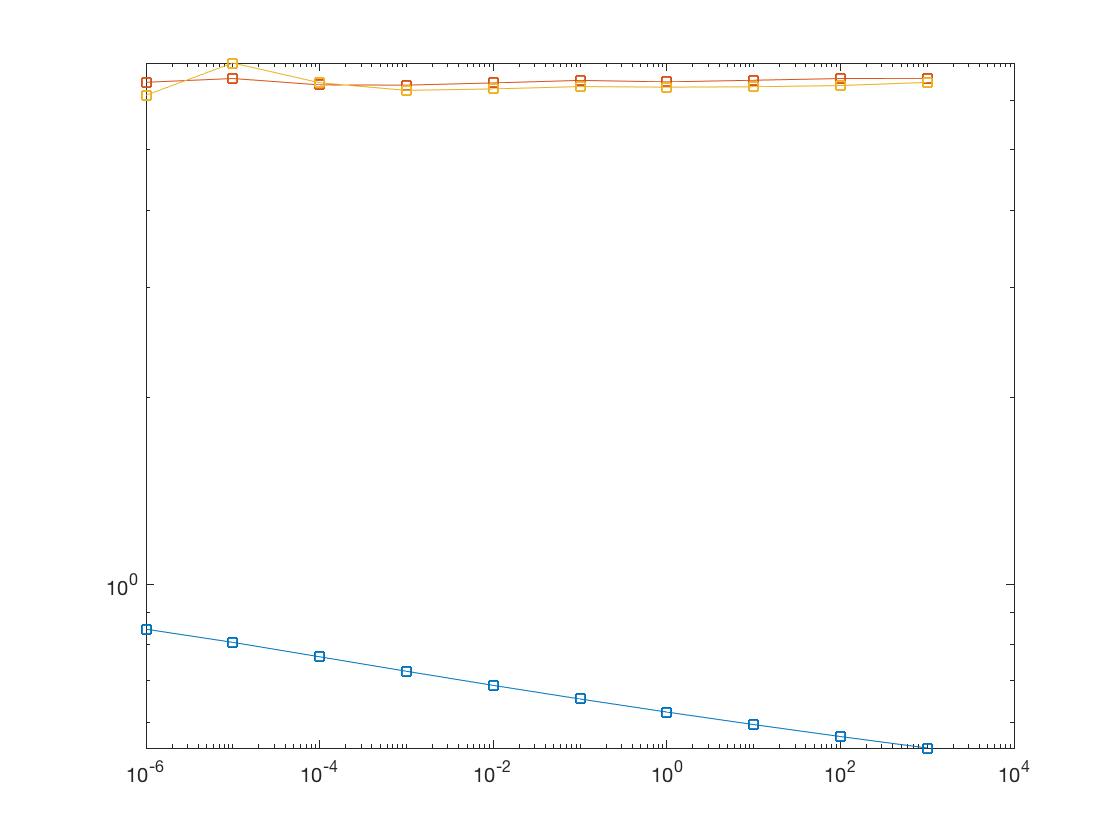
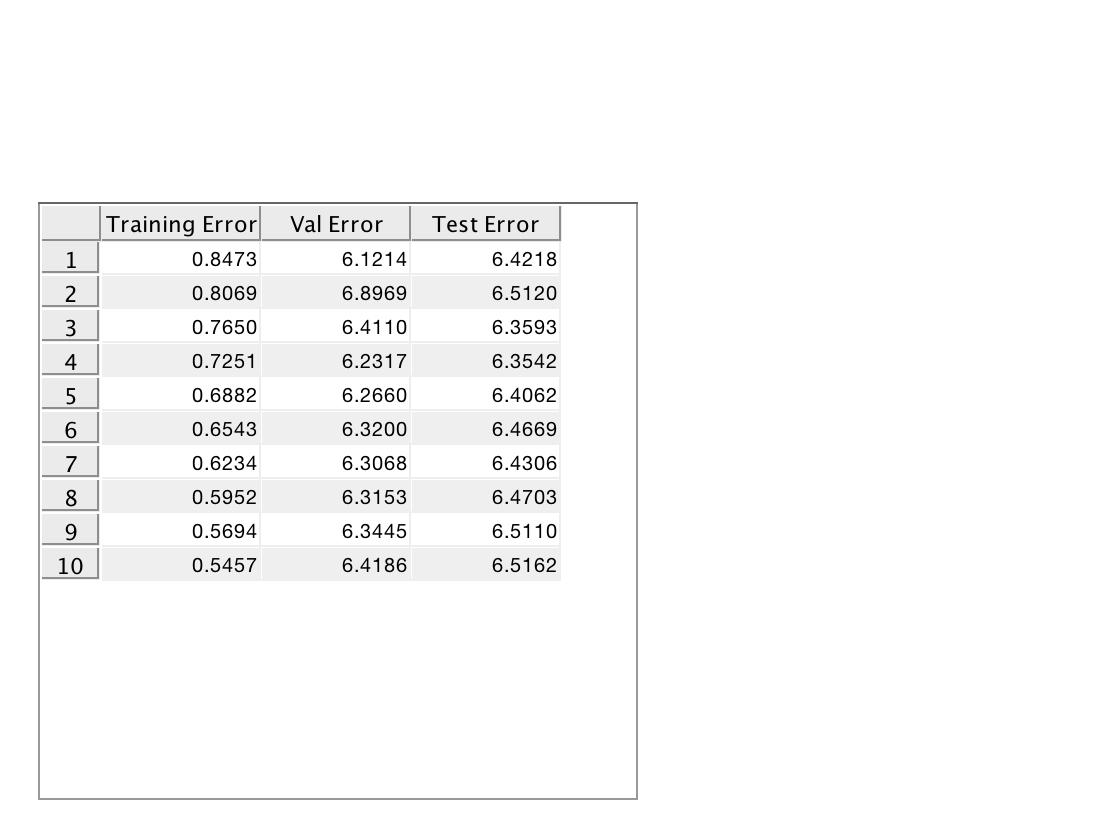
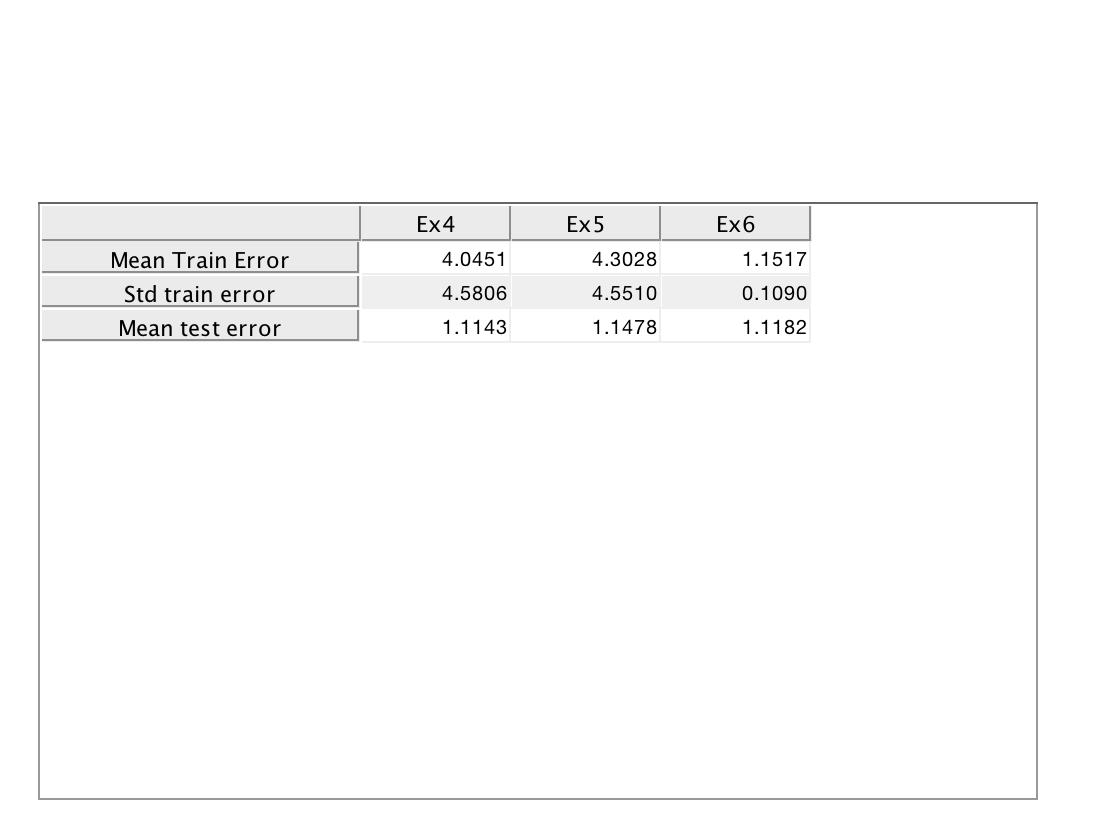
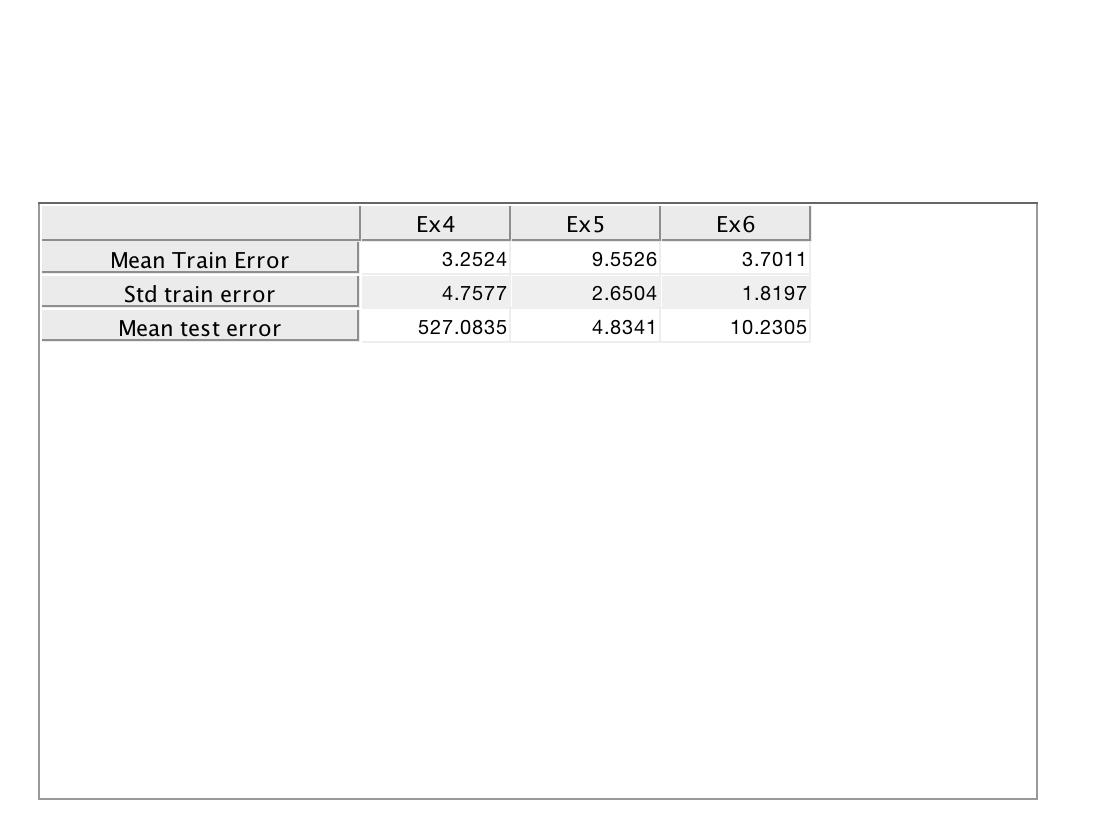
Fig 6.1 table for 100 data points after cross validation. Along with the graph for different values of gamma

Fig 6.2 table for 10 data points after cross validation. Along with the graph for different values of gamma

### Q7.

1. For 100 data points the mean test error, mean train error and the standard deviation of the train error are shown below

b. For 10 data points the mean test error, mean train error and the standard deviation of the train error are shown below



### Q9.