Pattern Recognition and Machine Learning
Assignment -1
Roll No. ME17S301
Name : Rana Pragneshkumar Rajubhai

Q:1

Model Build-UP:
1.Divided data into 70:15:15 propotion
2.Taking 70% of the data, bifurcated the data class wise and calculated the prior probability = (Data in the calss/Total data in the training set)
3.Estimated the class conditional density using maximum likelyhood. For that, mean ,co-variance matrix is calculated and proper function are defined for that.
4. All class conditional density are multiplied with prior and addition of all gave Evidence
5. density*prior/Evidence for specific class gave me posterior of class.
6. Each entry of training dataset is evaluated by posterior of each class and as we have here 3-class so, loss function is defined here for minimization of loss. So, for every entry after evaluation of posterior (1X3) is multiplied with loss matrix(3X3) which gives 3 values and out of which minimum value and its corresponding class give label or class to the given entry.

1. Classification Accuracy:
Code is randomly selecting the Training and validation points so every time while running the accuracy v aries but Training set and Validation set shpws quite consistent result .

Dataset-1	**Training Set**	**Validation set**
Model :1	0.9492063492063492	0.9466666666666667
Model :2	0.9650793650793651	0.9614814814814815
Model :3	0.9638095238095238	0.9674074074074074
Model :4	0.9650793650793651	0.9629629629629629
Model :5	0.9641269841269842	0.9733333333333334

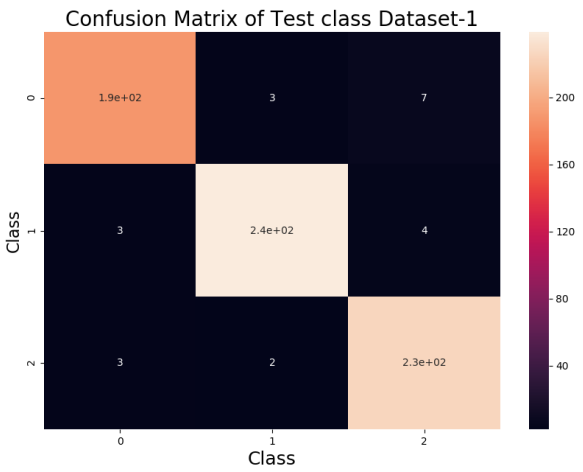
Dataset-2	**Training Set**	**Validation set**
Model :1	0.5365079365079365	0.5407407407407407
Model :2	0.9292063492063493	0.9392592592592592
Model :3	0.9295238095238095	0.9407407407407408
Model :4	0.9298412698412698	0.9377777777777778
Model :5	0.9304761904761905	0.9422222222222222

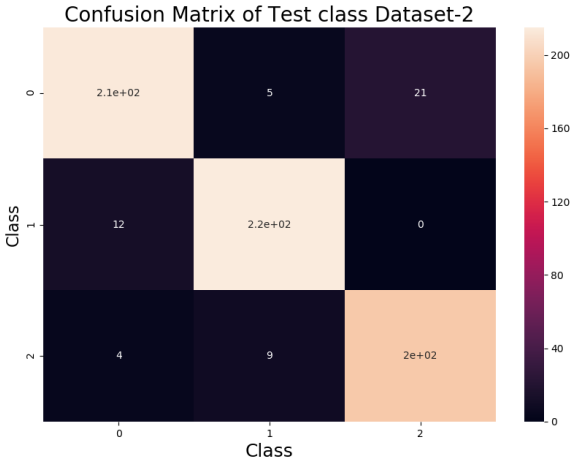
Note :For Model 2 and 4 common covariance Matrix is generated by whole dataset

Result:
Here, Model-5 has good accuracy out of all model for both the dataset. Sometimes model 2 and 4 gives better accuracy than metioned but those results are not consistent. Even for many number of trial on both dataset Model-5 gives consistently moderate result. For Datase -1, Variation in accuracy is low for different models but as dataset is changed the predominat effect of models are observed.

2. Confusion Matrix on Test Set foe Best Model-5 :

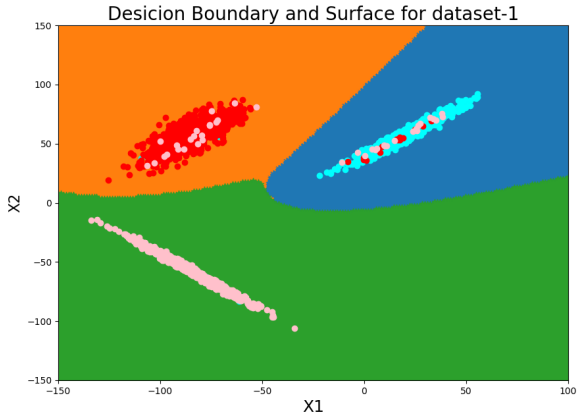
Confusion Matrix result shows the Poper classifiaction and missclassification.
As model is working properly misclassification rate is very low. 10 entries of class-0 are misclassified into wrong class like 3 points in class-1 and 7 points in class-2. Same interpretation can be observed for class 1 and 2 for other dataset also.

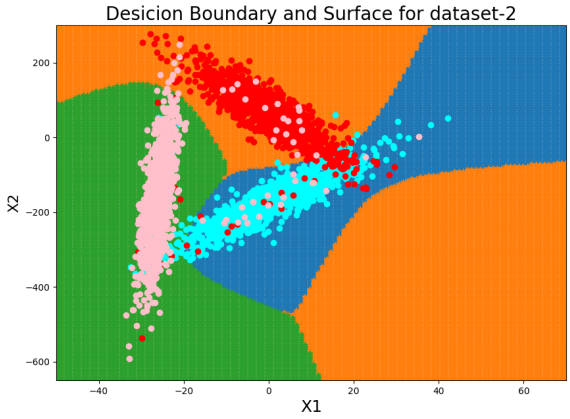




3. Decision Boundary and Surface for Best model-5 along with datpoints

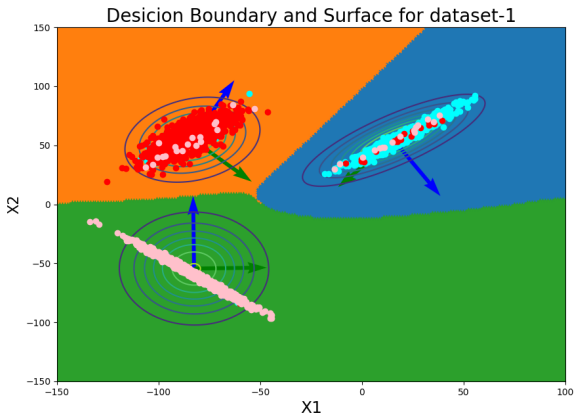
Decision Boundary shows the most probable region of each class

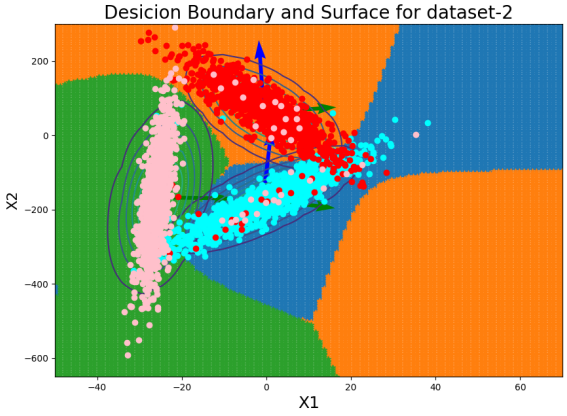




4. Contour curves and Eigen vector of Covariance Matric

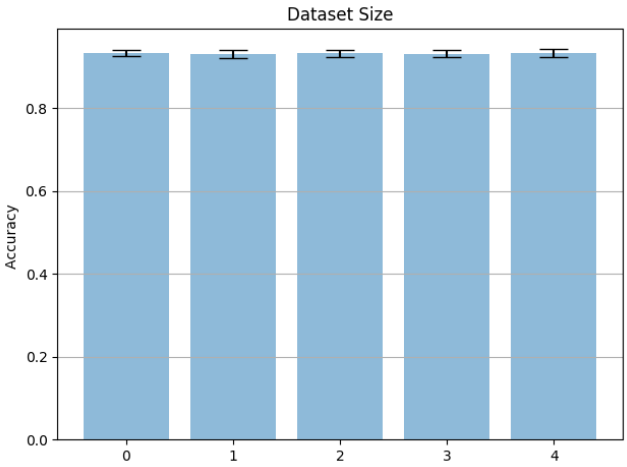
Eigen vector of Covariance Matric suggestes the spred along the major axis and contour curves the probable region in which the data falls.





Q:2

Build_Up:



I have selected the of the training dataset of given size [100,500,1000,2000,3000] for Dataset-2 , out of that i have selected 20 data points randomly. the Accuracy and satandard deviation is plotted as here. Both are checked on testing data. As datasize increase,the increment in Standard deviation is observed, which is due to increment in random of selection of the data points. For the fix number of sampling, as Training data size increases, the repetability of points is less which gives deviates the results more and increses the standard deviation.

Q:3

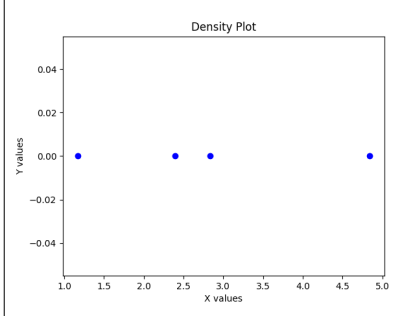
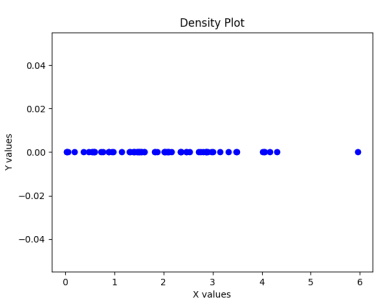
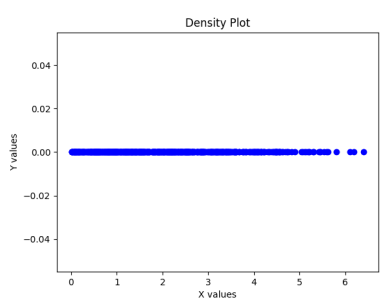
Dataset - 3			Dataset - 4		
No. of feature	Feature	Training Error	No. of feature	Feature	Training Error
1	x1	0.03428571428571429	1	x1	0.0033333333333333335
1	x2	0.06571428571428571	1	x2	0.0019047619047619048
1	x3	0.028095238095238097	1	x3	0.0
2	x1,x2	0.7361904761904762	2	x1,x2	**0.8204761904761905**
2	x2,x3	0.6914285714285714	2	x2,x3	0.6542857142857142
2	x3,x1	**0.7928571428571428**	2	x3,x1	0.6719047619047619
3	x1,x2,x3	0.49333333333333335	3	x1,x2,x3	0.5066666666666667

Discussion : For single feature both dataset has poor predicition but increase in feature accuracy increases, in which out of all pair one pair gives best result. For dataset-3, x3 and x1 feature and For dataset-4, x1 and x2 feature gives highest accuracy. But when I have consider all the features the accuracy has been reduced. So, increment in feature does not guarantee that accuracy will always increse. **For finite set of data, it is possible that Error might increse with increment in feature.**

Q:4

Density Plot:

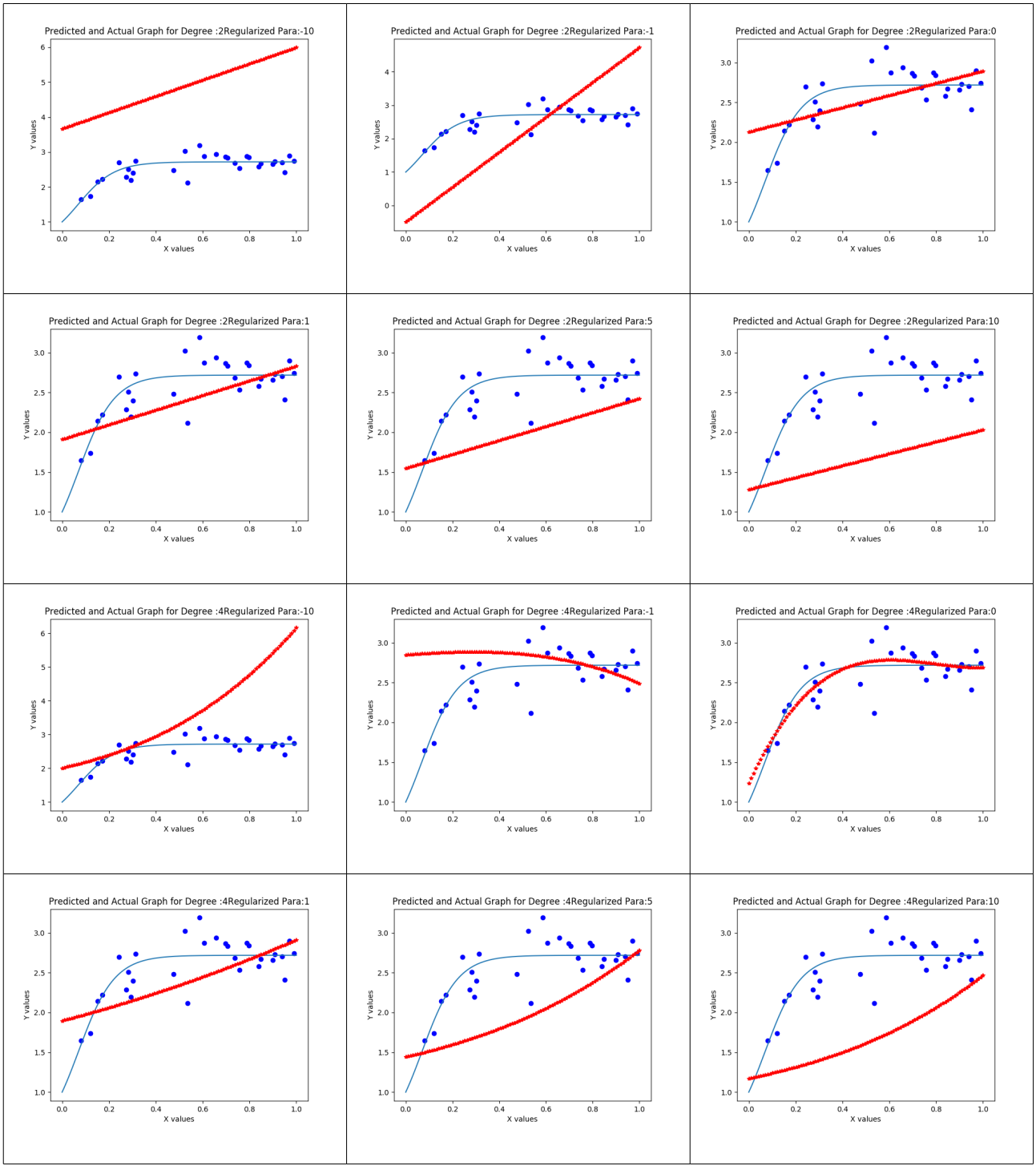
Points

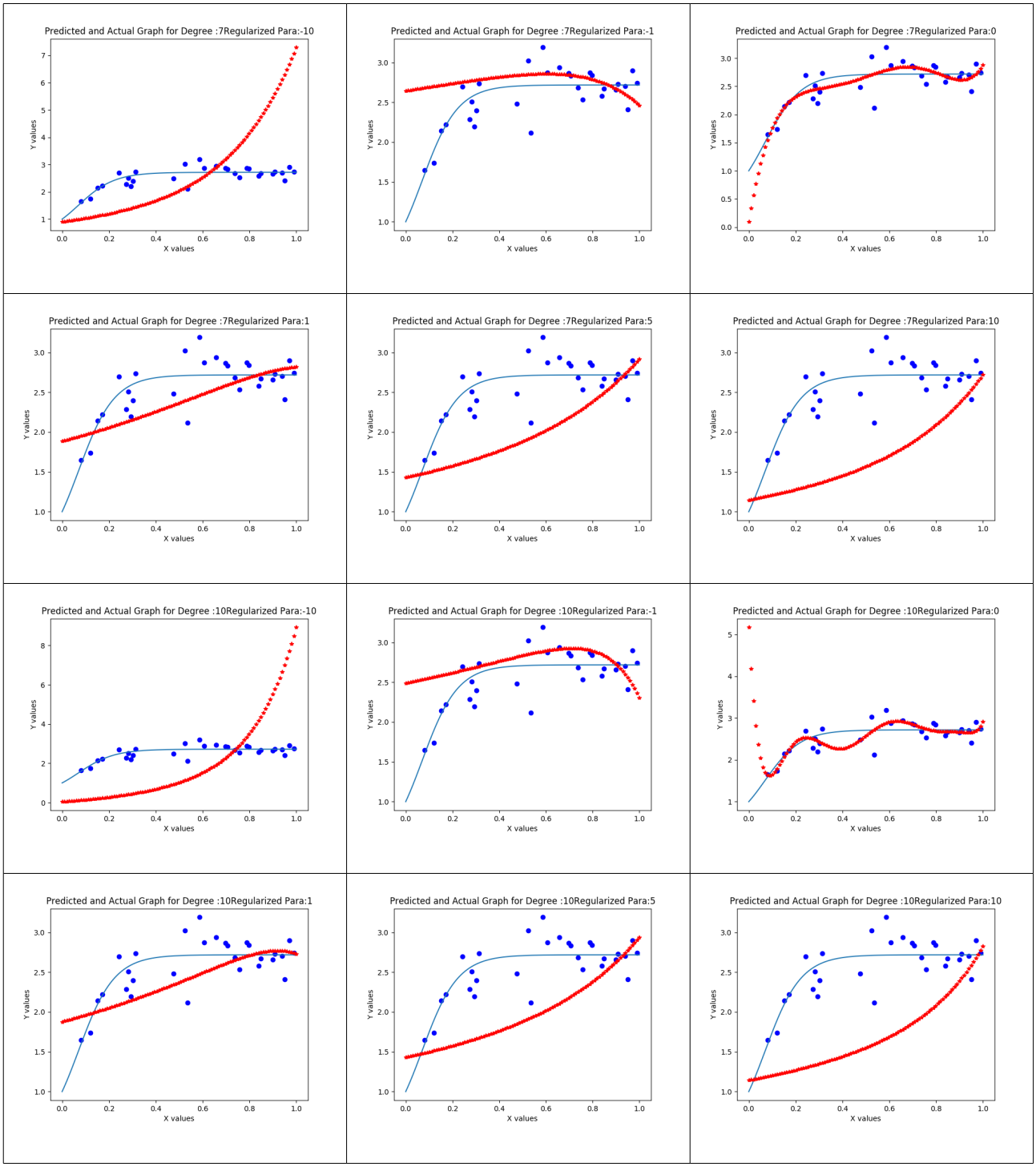
10	100	1000
		

Q:5

Graphs with Different polynomial fit and Regularization Parameter :

Here I have plotted the graphs with power [1,3,5,9]. As we can observe with regularization parameter overfitting is reduced. But very low[-10] and very high [10] regularization parameter does not improve the accuracy but give opposite effects.Here I have used regularization parameters [-10,-1,0,1,10].





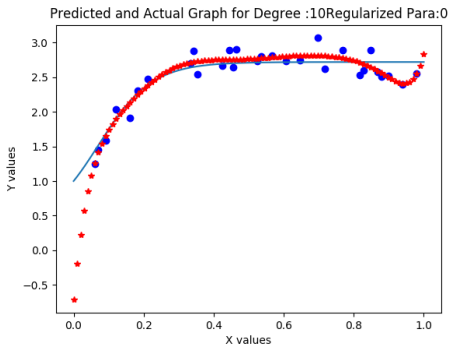
Regularization Parameter = -10				
Weights / Power	1	3	5	9
w0	3.88217386	2.25092941	1.17597719	0.44185311
w1	2.34703724	1.68562267	1.31282563	1.07414742
w2		1.39382826	1.24986422	1.17675486
w3		1.1926898	1.15148093	1.15370704
w4			1.05331481	1.09372441
w5			0.96334479	1.02375134
w6			0.88269424	0.95333477
w7				0.88608066
w8				0.82327869
w9				0.7652405

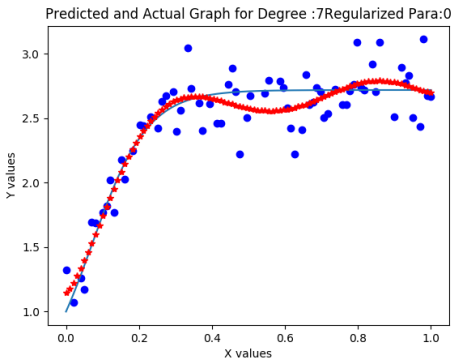
Regularization Parameter = -1				
Weights / Power	1	3	5	9
w0	-0.96020412	2.89190396	2.74355077	2.6629448
w1	6.28709657	0.107386	0.28963886	0.38068639
w2		-0.20350137	-0.01048181	0.0784462
w3		-0.26243246	-0.09348669	-0.02334935
w4			-0.12736148	-0.07458511
w5			-0.14282804	-0.10378179
w6			-0.14893732	-0.1204196
w7				-0.1291908
w8				-0.13288627
w9				-0.13328208

Regularization Parameter = 0				
Weights / Power	1	3	5	9
w0	2.13337603	1.0479474	1.17565731	1.06983211
w1	0.83717977	9.6094292	4.820096	15.0304423
w2		-16.2244394	19.01552789	-173.870253
w3		8.43215001	-85.24768058	1448.49284
w4			100.14009739	-6371.47954
w5			-31.73077964	15981.7941
w6			-5.61087171	-24094.5375
w7				21726.0287
w8				-10820.697
w9				2290.93773

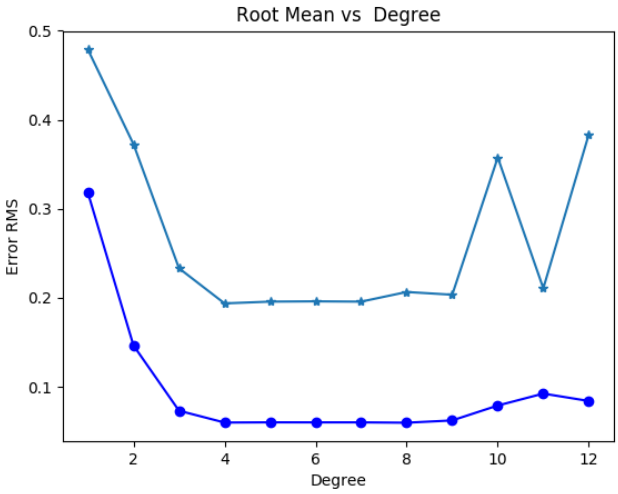
Regularization Parameter = 1				
Weights / Power	1	3	5	9
w0	1.94372082	1.93136806	1.91727358	1.90935807
w1	0.94524715	0.84158365	0.88341716	0.89007175
w2		0.21910969	0.2961192	0.3197449
w3		-0.04971006	0.04440348	0.07872255
w4			-0.06710472	-0.02670091
w5			-0.11654414	-0.07300961
w6			-0.13673221	-0.09191804
w7				-0.09764907
w8				-0.09698478
w9				-0.09329231

Likewise Reguarization parameter for +10 table can be obtained.
Plotting By for polynomial: 6 and regularization : 0 and varying the **sample size from 30(left Image) to 70(Right Image)** the result obtained as below.



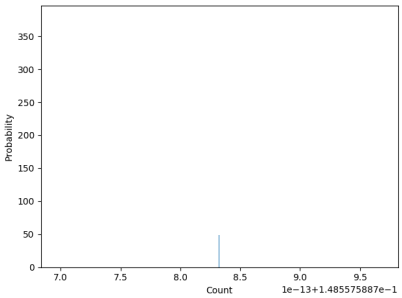
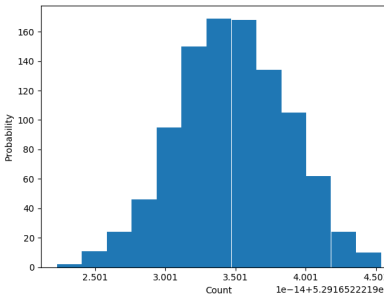
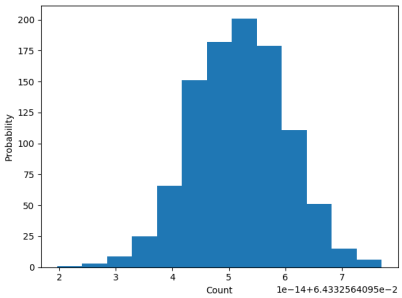
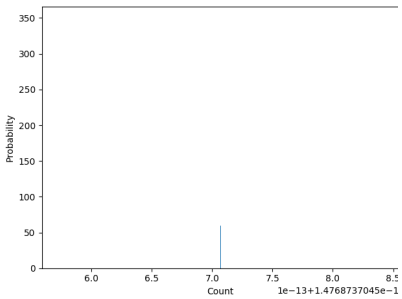
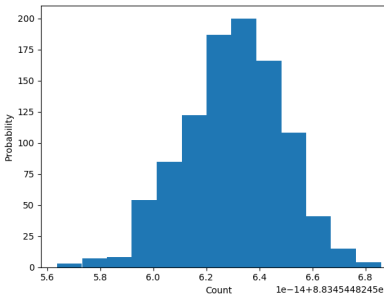
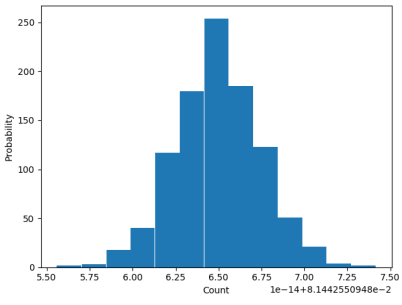


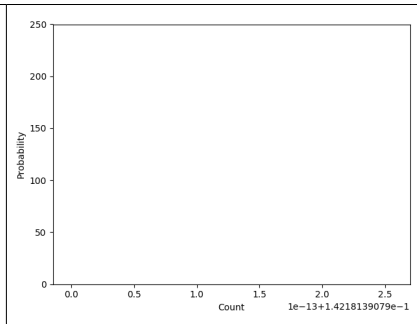
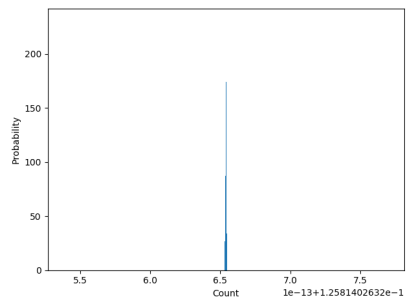
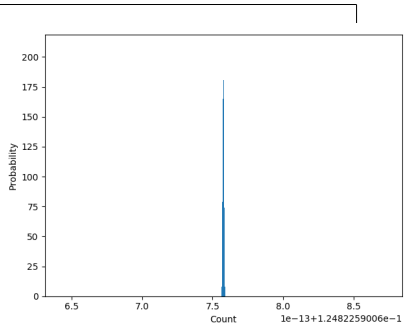
Root mean sqaure of Training and Testing varying with polynomical



As polynomial power changes, the error for traing and testing size decreases drastically but after cetain limit this Error again start increasing. This Error Behaviour varies when I run the code repetedly as each time sampleing of datapoints affetcts the ERMS.

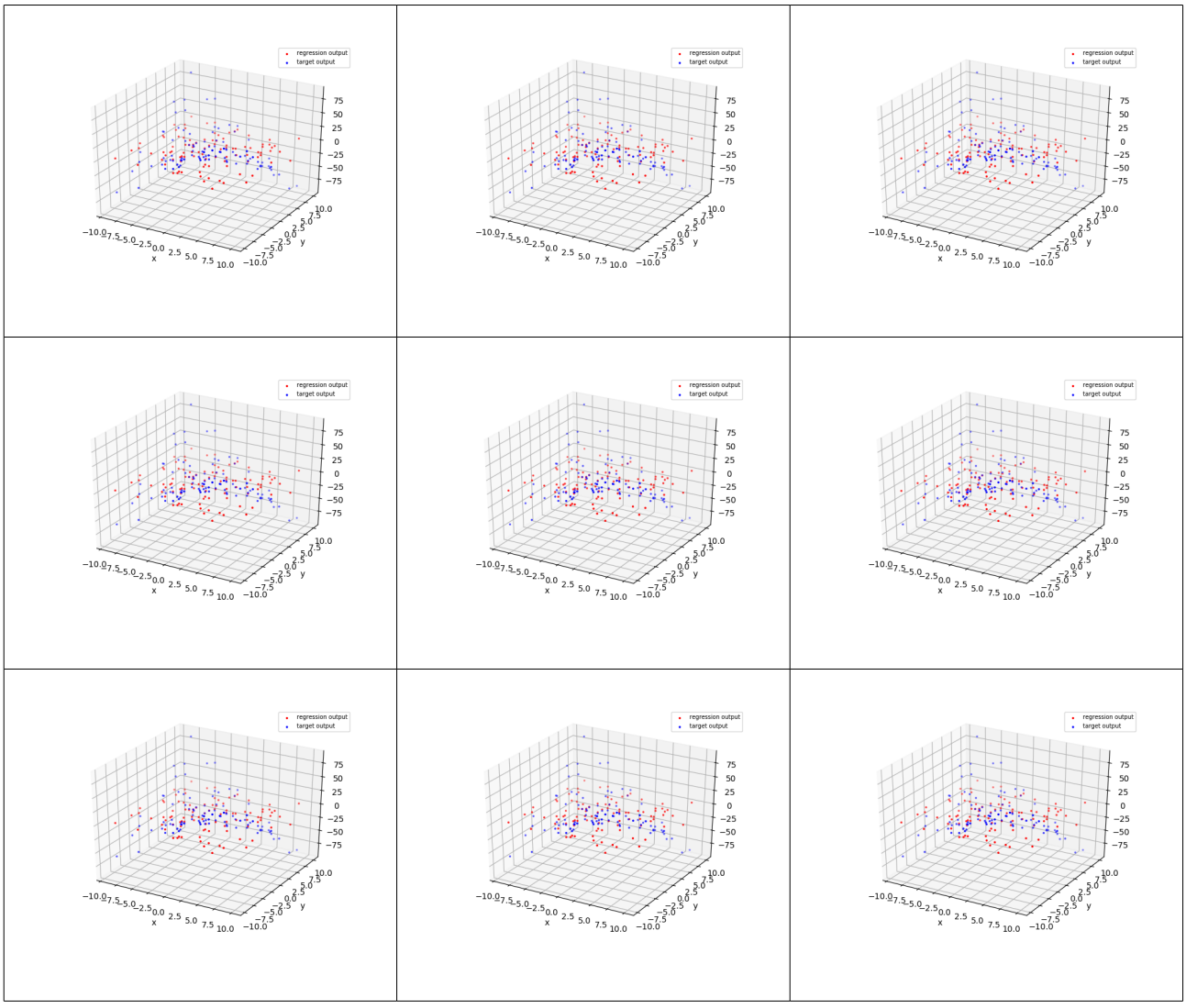
Q:6

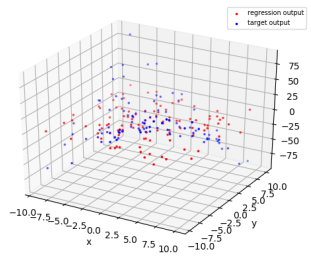
Lambda/degree	1	5	9
0.001			
0.01			

0.1			

The repetative process is useful to analyse tha bias variance trade off. The curve shows that very thin region shows the biasness of the data points where as wide spread shows variance. More biasness in the result shows under fitting where as wide spread is shows overfitting. To make compromise between this two one has to select the optimum model, Comparable between bias and variance.

Q:7





Lambda	Training Error	Testing Erro
-100	27.43857338	22.41826989
-1	27.42891881	21.54817334
-0.5	27.42775508	21.55625434
0.01	27.42663305	21.56426853
0.05	27.42654769	21.56488761
0.1	**27.42644152**	**21.56565955**
1	27.42462627	21.57920137
10	27.4133012	21.68560738
100	27.40469556	21.97367082
1000	27.40923996	22.09968909

Here, By gaussian basis function dimentional reducation is made then regression analysis done on bais function, which showed overfitting. Overfitting fitting of basis is controlled by regularization parameter which is mentioned in the table. Out of which 0.1 parameter shows quite good result of error.