Assignment2

cs19b026 - sravan , cs19b030 - aryan

March 3, 2022

1 REGRESSION

1.1 MOTIVATION

To understand polynomial linear regression for 1D and 2D data points with regularisation

1.2 1D DATA

1.2.1 EXPERIMENTAL RESULTS

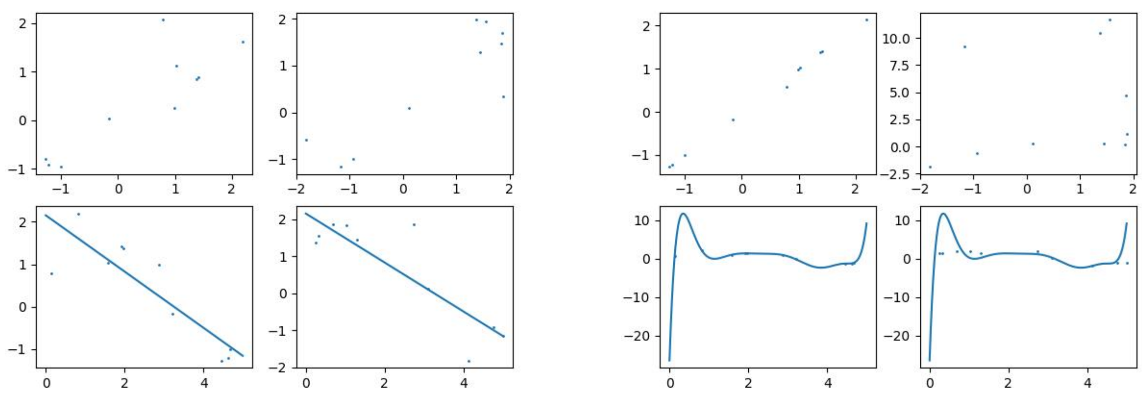


Figure 1: N = 10,degree = 1 and 9,lambda = 0

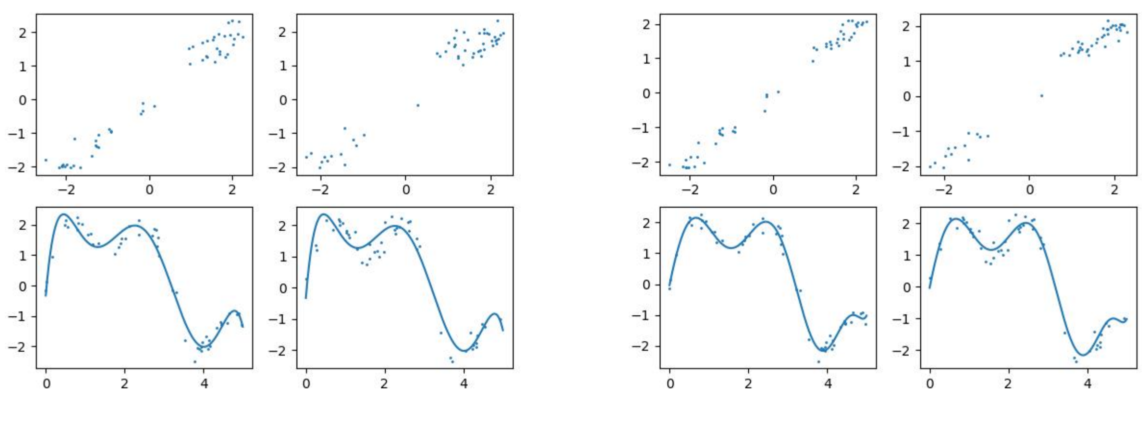


Figure 2: N=50,degree=6 and 9,lambda = 0

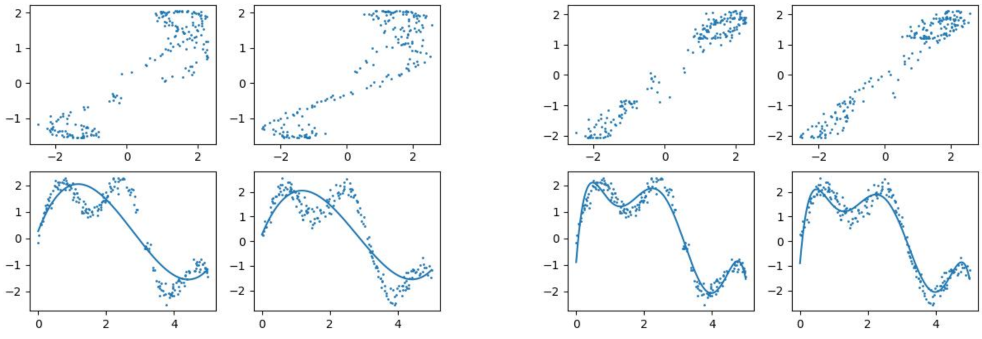


Figure 3: N = 200,degree = 3 and 6,lambda = 0

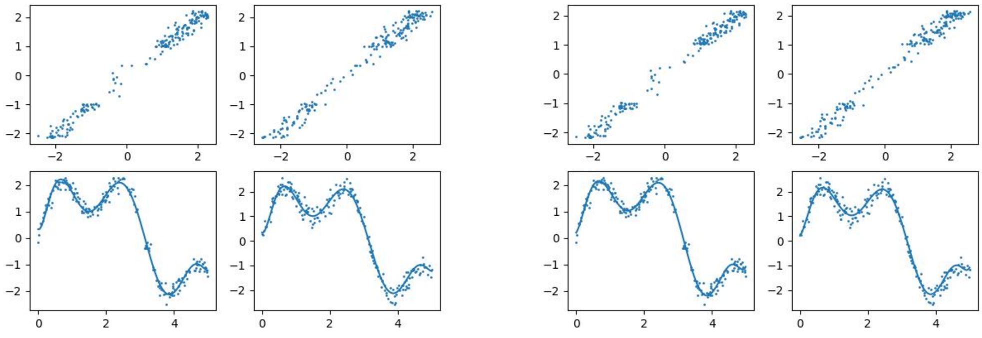


Figure 4: N = 200,degree = 8 and 9,lambda = 0

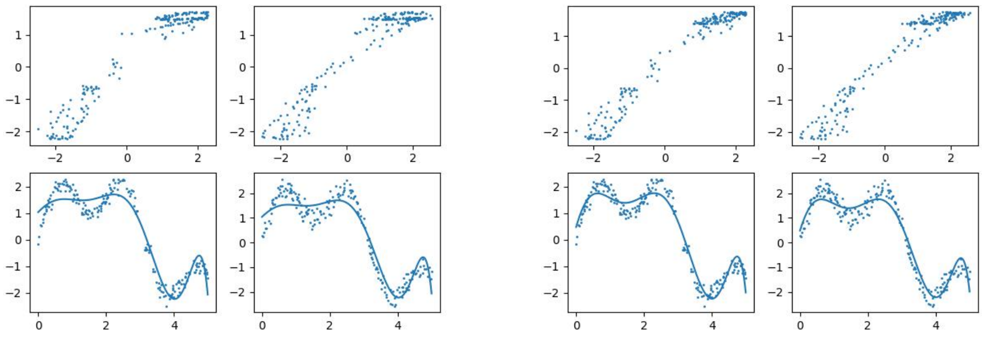


Figure 5: N = 200,degree = 8,lambda = 1 and 0.1

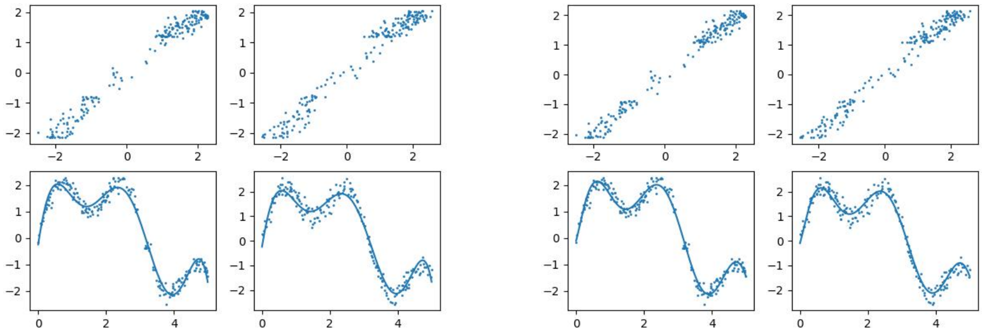


Figure 6: N = 200,degree = 8,lambda = 0.01 and 0.001

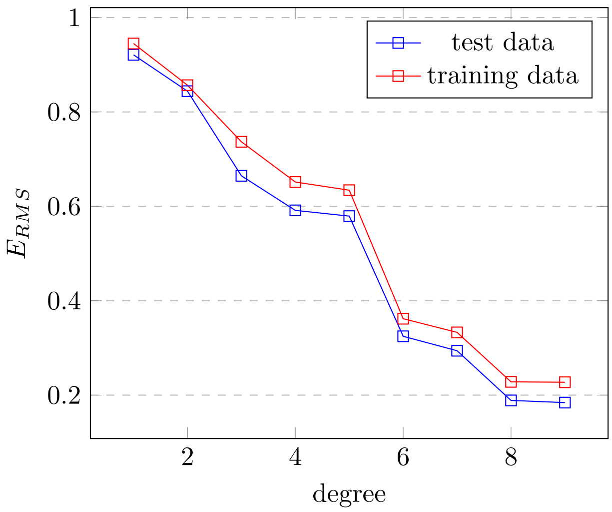


Figure 7: RMS error versus degree with lambda = 0 for 200 samples

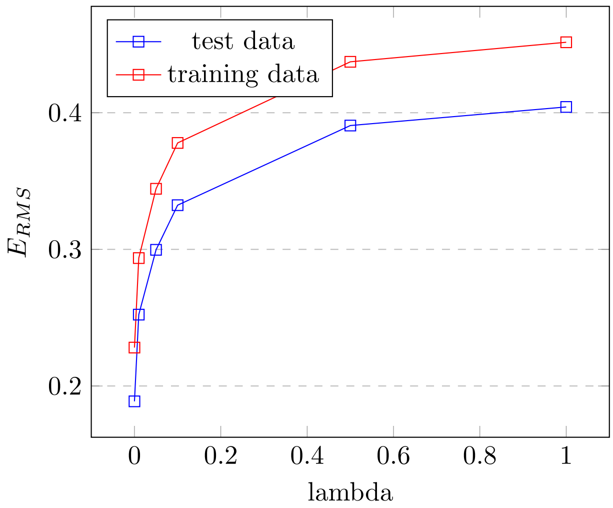


Figure 8: RMS error versus lambda with degree = 8 for 200 samples

	λ = 0	λ = 0.01	λ = 1
w_0	0.33649	-0.24138	1.03923
w_1	-0.83885	8.41218	1.30632
w_2	26.79831	-6.67786	-0.61027
w_3	-56.25941	-6.74411	-0.70131
w_4	48.63127	11.5655	0.336554
w_5	-21.40401	-6.0899	0.361705
w_6	5.04454	1.504338	-0.27951
w_7	-0.60668	-0.1777	0.064407
w_8	0.02927	0.00805	-0.00489

Figure 9: table with lambda and w values for N = 200,degree = 8

1.3 2D DATA

1.3.1 EXPERIMENTAL RESULTS

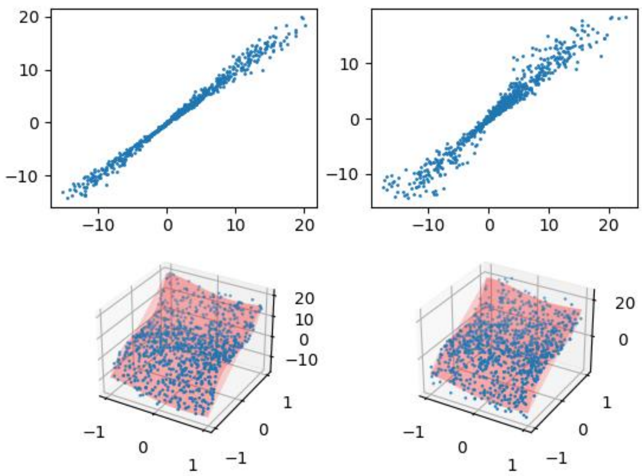


Figure 10: Best Fit for 2d Dataset,N = 1000, $\lambda = 0$,$deg = 4$

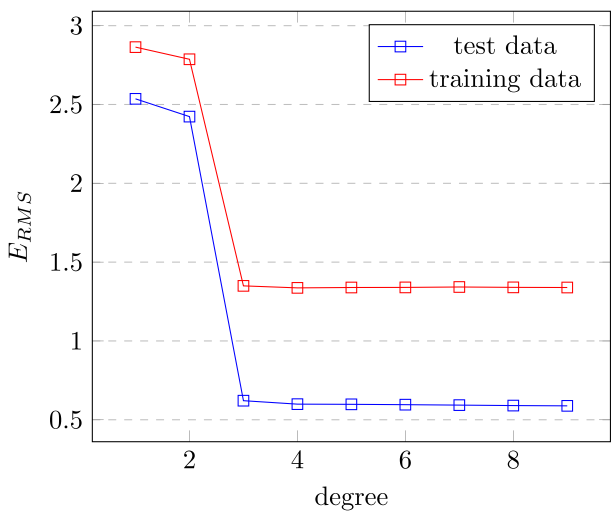


Figure 11: RMS error versus degree with lambda = 0 for 1000 samples

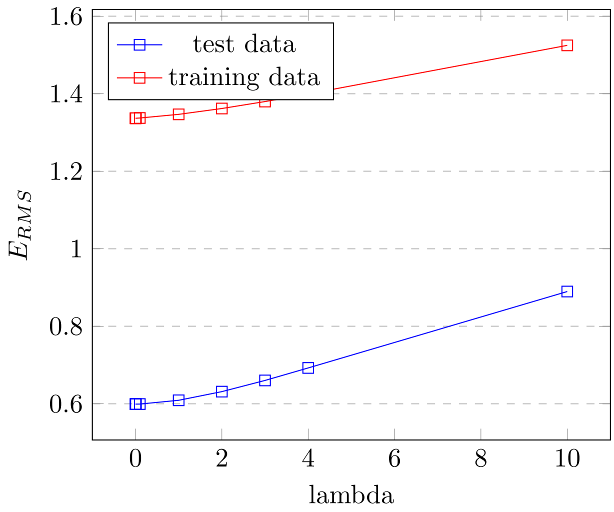


Figure 12: RMS error versus lambda with degree = 4 for 1000 samples

	λ = 0	λ = 0.01	λ = 1
w_0	1.021144492	1.02199467	1.084405211
w_1	0.1056675852	0.1093737413	0.4545769352
w_2	-0.9333326205	-0.9328794274	-0.8901681525
w_3	-0.3787874223	-0.3774691503	-0.2731217007
w_4	-0.2396312619	-0.239416329	-0.2136237283
w_5	4.063681311	4.054750165	3.387034236
w_6	14.73073886	14.72462828	14.15163367
w_7	-3.011294019	-3.010183979	-2.904332965
w_8	-0.00901908171	-0.00911139580	-0.00983381565
w_9	1.869614314	1.868448522	1.755433501
w_10	0.5653332279	0.5633761956	0.4088848189
w_11	-0.06864787093	-0.06802928268	-0.0205268703
w_12	0.1035897987	0.1045004416	0.1756363738
w_13	0.4116978307	0.4111287834	0.356527732
w_14	-2.073933782	-2.064247141	-1.342604528

Figure 13: table with lambda and w values for N = 1000,degree = 4

1.4 INFERENCE

1.4.1 1D Data

All figures from 1 to 6 have scatter plots of ypredicted vs yactual on top row and trained model drawn on training and data points in bottom row. Left figures are training data plots and right figures are Development data plots.
1) In figure-1, for N = 10, we can see the cases of underfitting and overfitting. From scatter plots we can note that both perform poorly on development data irrespective of how they perform on training data.
2)In Figure-2, At N =50, the model performs decently on both training and development data for all degrees between 6 and 9.
3)Figure - 3 shows trained data for N = 200 for degree 3 and 6.
4)In Figure - 4 1st and 2nd plots are close to x = y line, which imply The good performance of model
5)When using regularization, The plots looked like Figure 5 and 6.For this data, regularization did not help.
6)Figure 7a shows that optimal model is of degree 8.(Increasing complexity to degree 9 did not help much)
7)Figure 7b shows that lambda is 0,i.e regularization only increased the error(For both training and dev set)
8)Table of figure 8 shows values of w for different regularization parameter. We can see that for higher powers increasing lambda decreases the coefficients.(There by reducing overfitting)

1.4.2 2D Data

1)Figure 10 shows best fit for 2d data.Left are plots for training data and right are plots for dev data.
2)From figure 11, we can observe that complexity of model can be 3 or 4 and complexity need not be increased more because the error value doesn't change significantly.

3)According to 12, increasing lambda increased both training and development error indication lambda = 0 is optimal.
4)From table in figure 13 we can see effect of lambda on values of w, note that w are not in order.

2 BAYESIAN CLASSIFIER

2.1 MOTIVATION

To understand classification using Bayesian models for different kinds of data

2.2 EXPERIMENTAL RESULTS

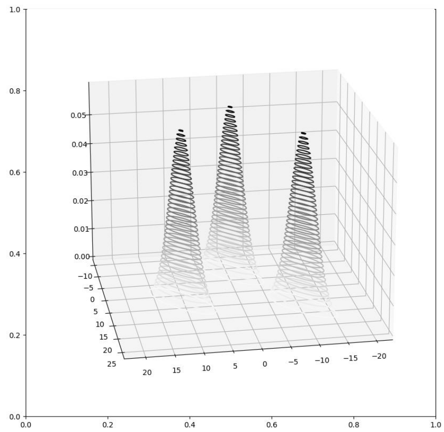


Figure 14: PDF of Linear Model(Case 1) when trained on Linearly separable Data

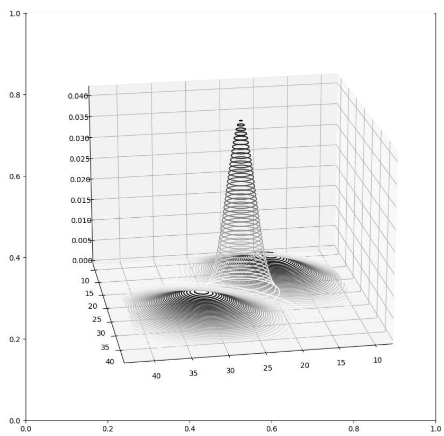


Figure 15: PDF of Case2 Model when trained on Linearly Inseparable Data

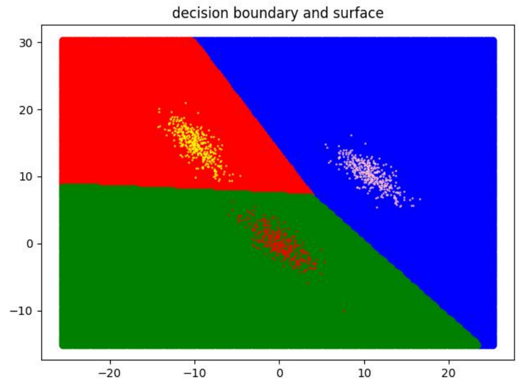


Figure 16: Surface of Linear Model(Case 1) when trained on Linearly separable Data

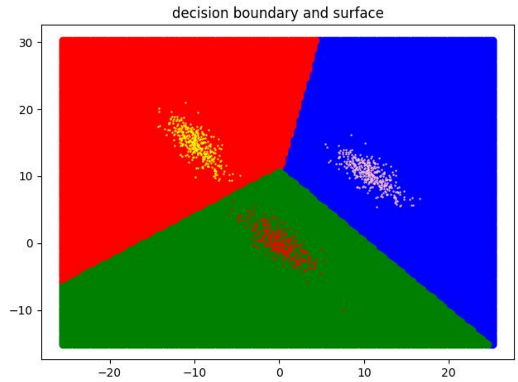


Figure 17: Surface of Case3 Model when trained on Linearly separable Data

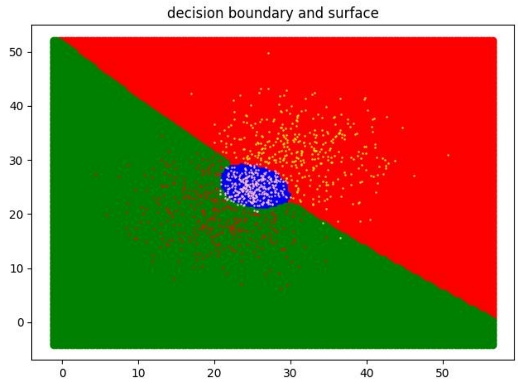


Figure 18: Surface of Case2 Model when trained on Linearly Inseparable Data,This was same as when Case5 model was plot

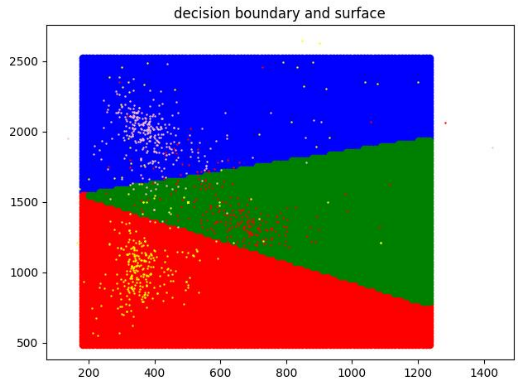
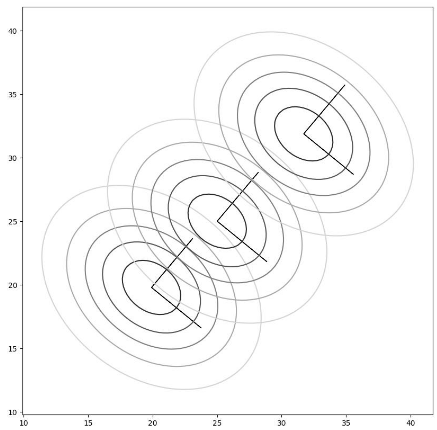
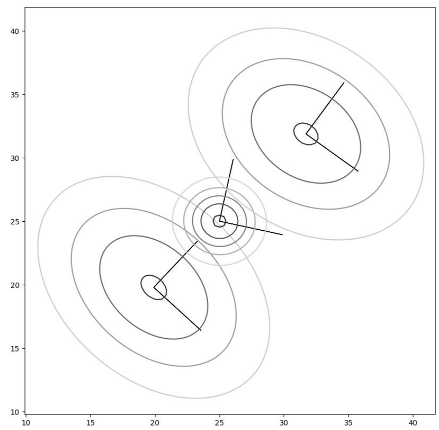


Figure 19: Surface of Case3 Model when trained on Real Data

Contour plots with eigen vectors when trained on linearly inseparable data for case 1,2 and 5 models.





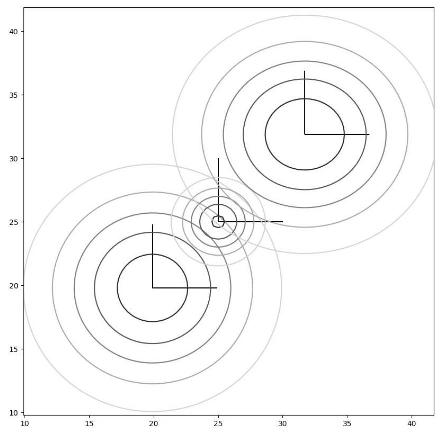


Figure 20: case1,case2,case5

	1	2	3	total % correct
1	100.0	0.0	0.0	100.0
2	0.0	100.0	0.0	100.0
3	0.0	0.0	100.0	100.0
total % correct	100.0	100.0	100.0	100.0

	1	2	3	total % correct
1	100.0	0.0	0.0	100.0
2	1.0	99.0	0.0	99.0
3	1.0	0.0	99.0	99.0
total % correct	98.0392156862745	100.0	100.0	99.33333333333333

Figure 21: Confusion Matrices for Case 1 for linearly inseparable data,Case 5 for real data,

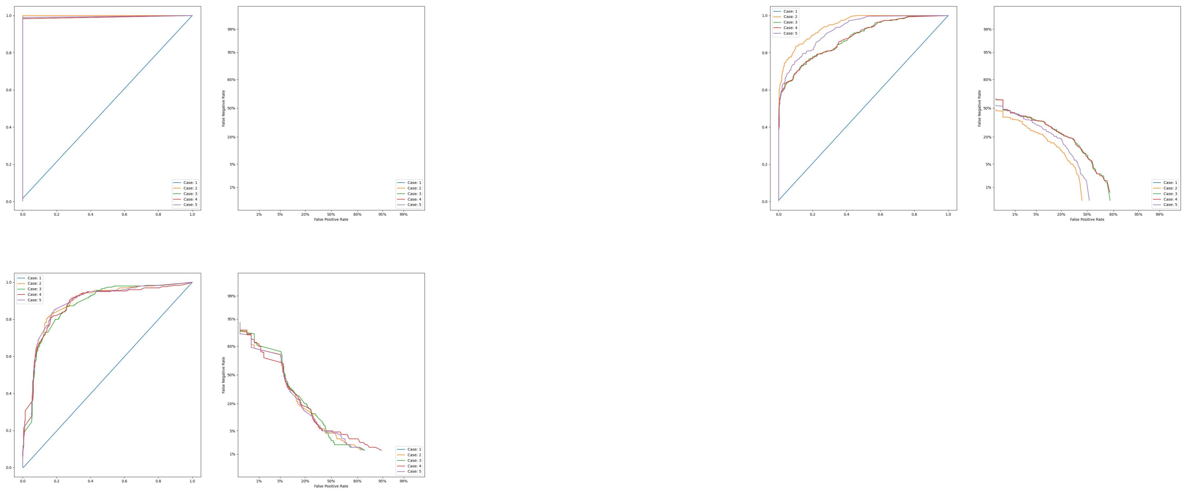


Figure 22: ROC and DET curves For Data 1,2,3

2.3 INFERENCE

1) For linearly separable data, all the models performed well(Accuracy 100) on development data.Figure 14 shows pdf of case 1 on linearly separable data.
2)PDF of non linear data(2) is shown in Figure 15 when case2 Model is used.(It was the best fit along with case5 model).
3)Figure 15 - 18 show decision boundaries of best fit models for the corresponding data.
4)All the plots for Linearly inseparable data 2 when trained assuming model 2 were similar to those of model 5, indicating that the features in data 2 are independant.
5)From eigen vector plots we can see that, eigen vectors of sigma are normal to contours of pdf.
6)Confusion matrix for corresponding best fit models of different data are in figure 22 - 24.