CS5691 - Assignment2

LINEAR REGRESSION AND BAYESIAN CLASSIFICATION

LINEAR REGRESSION

The goal of regression is to predict the value of one or more continuous target variables t given the value of a D-dimensional vector $\mathbf{x}$ of input variables.

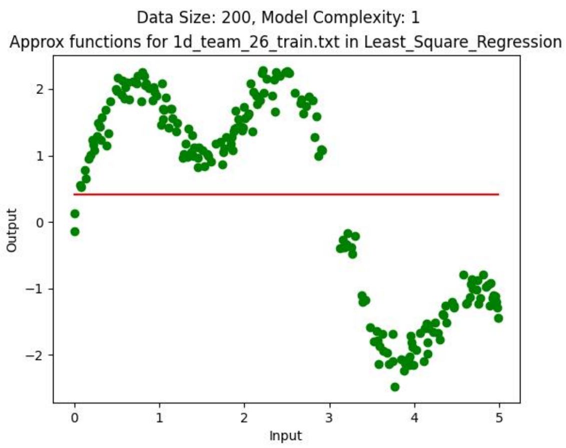
1) For 1D input data :- we shall fit the data using a polynomial function of the form $y(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots + w_M x^M$

A) Least Square Regression:- In Least square Regression error function is

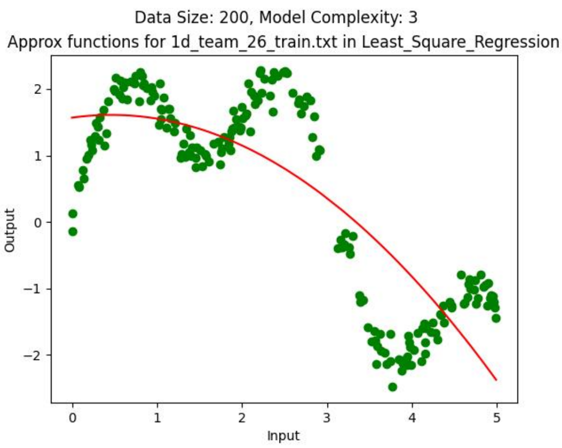
$$E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{n=N}(y(x_n,\mathbf{w}\ ) - t_n)^2 \text{ (Where } N \text{ is size of data)}$$

Closed form solution is given by $E(\mathbf{w}_{LS}) = [(\phi^t\phi)^{-1}\ \phi^t]\ \mathbf{t}$

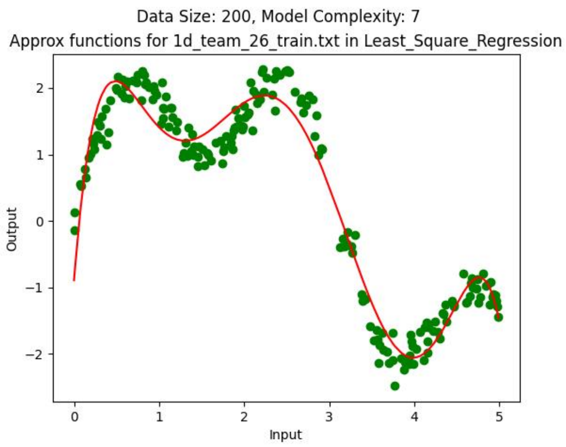
1) Dependence of Performance on Order M:-



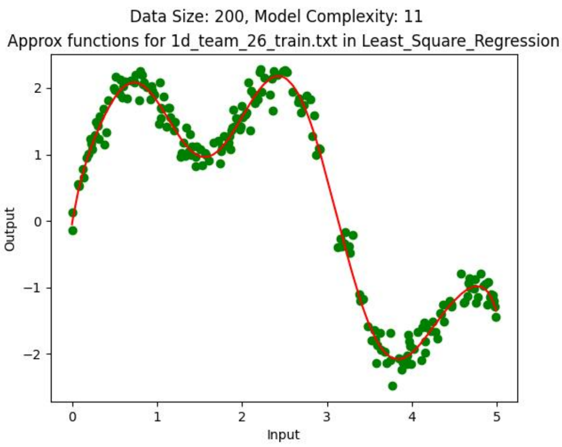
(a) $M = 1$



(b) $M = 3$



(c) $M = 7$



(d) $M = 11$

Figure 1: Plots of polynomials having orders M,shown as red curves. Green dots are given data points for Training Data.

Inference from above graphs :- $M = 1$ and $M = 3$ polynomials are giving poor fit. $M = 7$ and $M = 11$are seems to be giving the better fit. But, This is not enough to conclude that $M = 5$ or $M = 7$ gives always best fit.

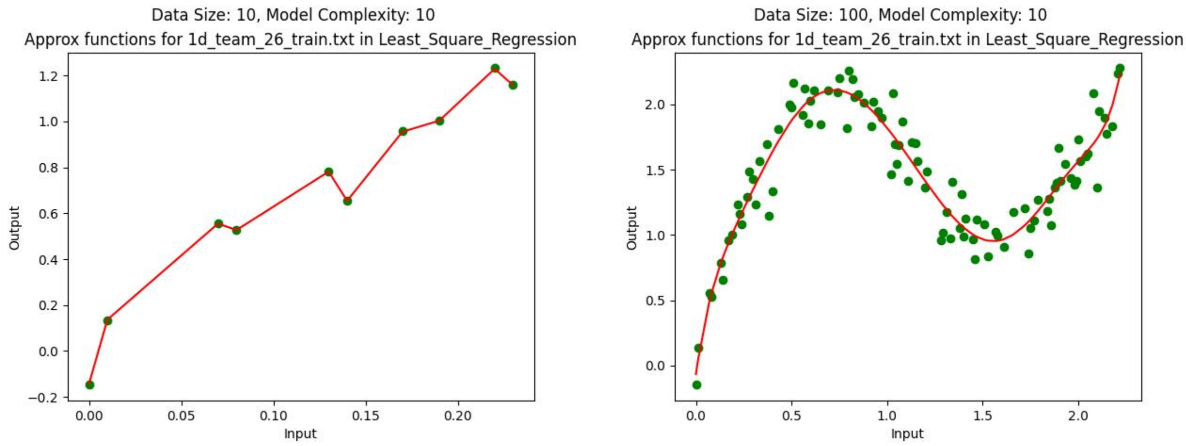
Therefore, as M increases, there is better fit in Training Data and at some point there will be overfit .

weights	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	
$M = 3$	1.09	0.79	-0.31									
$M = 5$	-0.13	7.68	-8.53	2.91	-0.18							
$M = 7$	0.07	4.94	-0.55	-5.22	2.46	0.309	-0.214					
$M = 13$	3.89	1.66	-59.57	305.88	-786.07	1161.31	-950.65	263.65	250.94	-296.99	136.65	-

Inference from the table :-

As M(order of polynomial) increases the magnitude of coefficients gets larger. For large values of M the model increasingly tunes to the random noise in data.

2) Varying N (datasize) for a fixed model order M:-



(a) $N = 10$ (b) $N = 100$

Figure 2: Plots of polynomials having orders $M = 10$,shown as red curves. First graph is corresponding to $N = 10$, second is $N = 100$

Inference from above table :-

We can observe that for a given model complexity, as the size of dataset increases the problem of overfit becomes less severe. Larger the dataset, we can afford to fit data with larger model complexity.

3) Variation of Error with Model Complexity M :- Instead of considering $E(\mathbf{w}_{LS})$ error, We consider

$$E_{RMS} = \sqrt{\frac{2E(\mathbf{w}_{LS})}{N}}$$

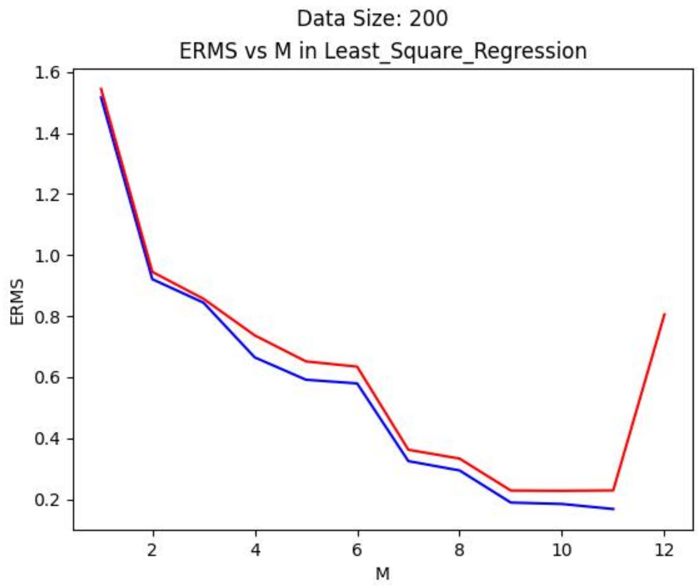


Figure 3: $RMSerrorvsM$

Inference from graph :-

As, M increases in the case of training data the value of RMS error is giving small errors .But For the case of development data the RMS error suddenly increase after reaching certain complexity.

Blue Color :Training Data

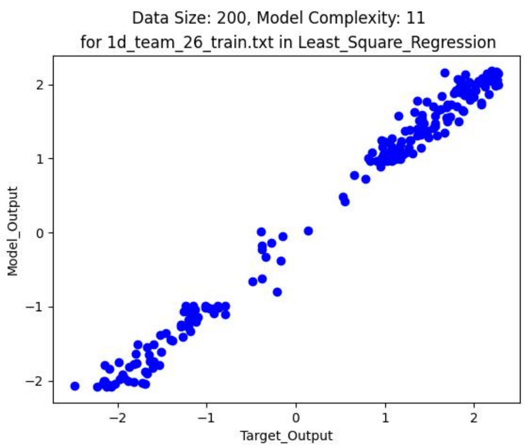
Red Color :Development Data

Best Performing model for whole data:-

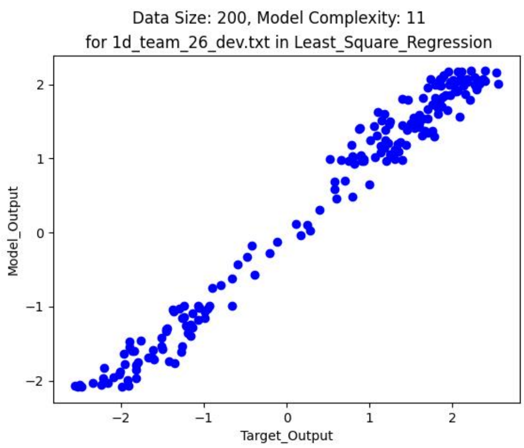
Model Complexity :- polynomial of degree **11**

RMS Error for training data :- **0.1676517**

RMS Error for development data :- **0.22852377**



(a) Training data



(b) Development data

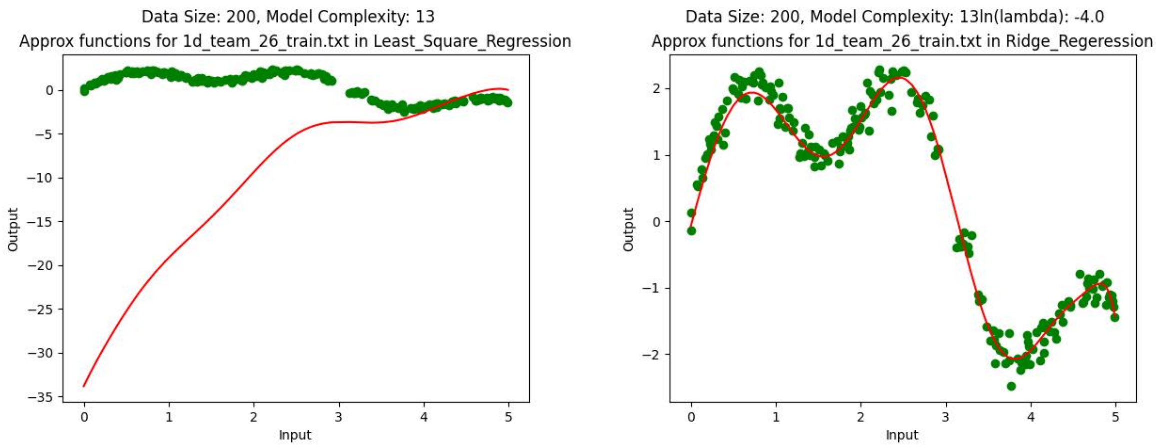
Figure 4: Model Output vs Target Output

B) Ridge Regression:- Ridge Regression is used to control magnitude of weights.
In Ridge Regression error function is $E(\mathbf{w}) = \frac{1}{2}\sum_{n=1}^{n=N}(y(x_n,\mathbf{w}\ ) - t_n)^2 + \lambda||\mathbf{w}^2||$

Closed form solution is given by $E(\mathbf{w}_{LS}) = [(\phi^t\phi + \lambda I)^{-1}\ \phi^t]\ \mathbf{t}$. .

weights	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9	w_{10}	w_{11}	w_{12}	w_{13}
$M = 3$	1.09	0.79	-0.31										
$M = 5$	-0.16	5.39	-4.39	0.23	0.38								
$M = 7$	0.16	4.54	-0.10	-2.39	-0.69	1.71	-0.43						
$M = 13$	0.21	4.08	-0.72	-1.34	-0.83	-0.319	0.31	0.35	0.08	-0.15	-0.04	0.04	-0.006

Inference from above table :- Comparing with the table of Least square regression, the values are smaller in case of ridge regression



(a) $ln(\lambda) = -\infty$ (b) $ln(\lambda) = -4$

Figure 5: Plots of polynomials having orders $M = 13$,shown as red curves.First graph is corresponding to $ln(\lambda) = -\infty$, second $ln(\lambda) = -4$

Variation of Error with ln(λ) for Model complexity $M = 11$:-

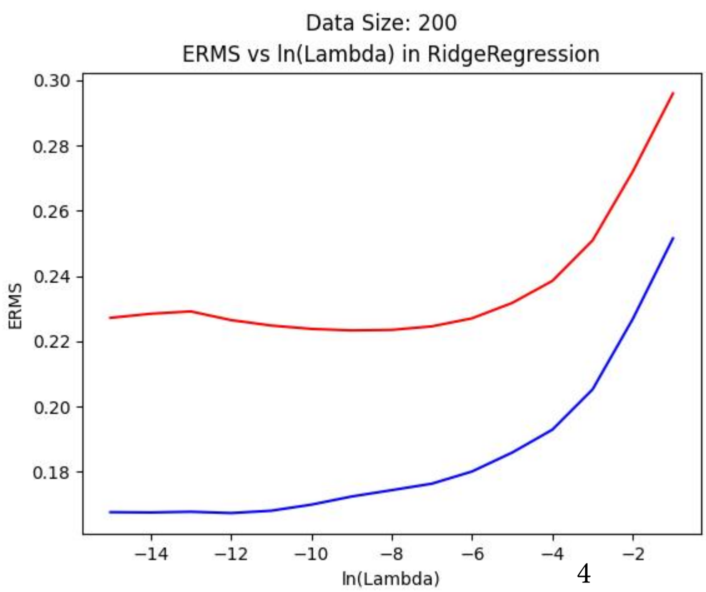


Figure 6: RMS error vs ln(λ)

Blue Color:- Training Data
Red Color:-Development Data

Inference from graph :- λ can also control the effective complexity of the model. λ determines the degree of overfitting. Choosing the best λ is also very important and crucial in deciding overall model

Variation of weights with ln(λ) for Model complexity $M = 9$:-

weights	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	w_9
$ln(\lambda) = -20$	0.33	-0.83	26.79	-56.25	48.63	-21.40	5.04	-0.61	0.029
$ln(\lambda) = -10$	0.29	-0.31	25.05	53.81	46.86	-20.69	4.88	-0.58	0.028
$ln(\lambda) = -5$	-0.28	8.42	-5.79	-8.81	13.47	-6.98	1.72	-0.21	0.01
$ln(\lambda) = 1$	1.11	0.77	-0.27	-0.44	0.10	0.38	-0.26	0.06	-0.004

Inference from the table :-

As λ increases the value of coefficients gets smaller.

2) For 2D input data :-

A) Least Square Regression:-

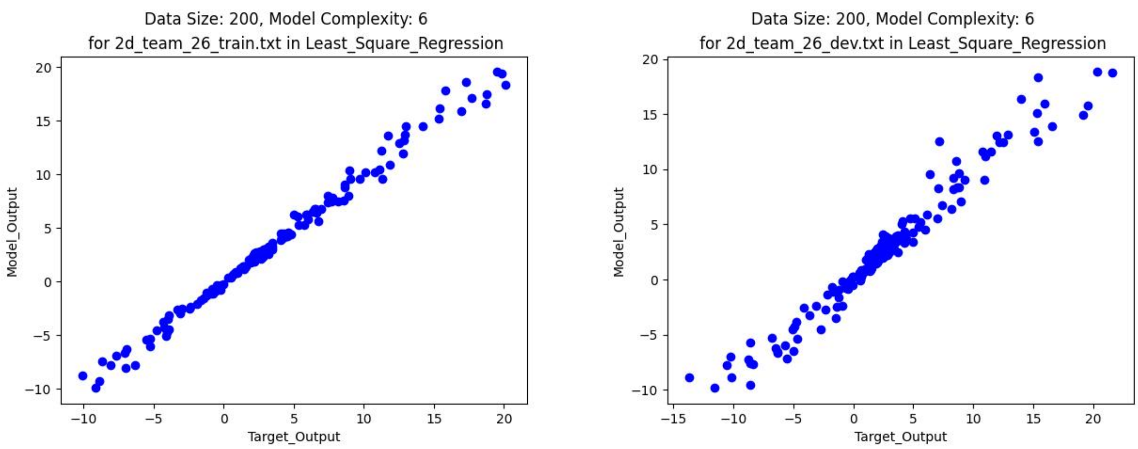
Best Performing model for whole data:-

Model Complexity :- polynomial of degree **6**

RMS Error for training data :- **0.54375782**

RMS Error for development data :- **1.13407802**

Target output vs Model output for best model:-



(a) Training Data (b) Development Data

Figure 7: Target Output vs Model Output.

Variation of RMS error with model complexity :-

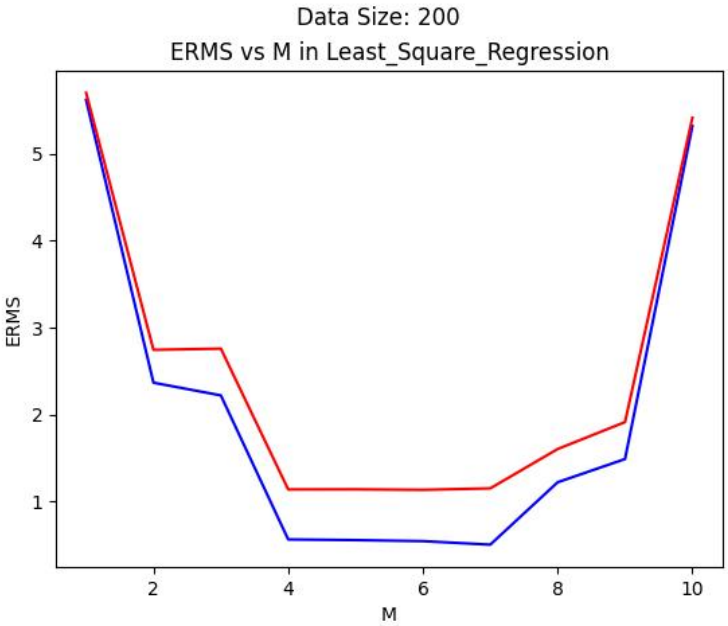


Figure 8: RMS vs M

Blue Color:- Training Data
Red Color:-Development Data

Inference from graph :-

As M increases RMS Error is decreasing upto some point and then it is increasing for both training and development data. The least value is reached at M=6

B) Ridge Regression

Variation of RMS error with λ for model complexity 9:-

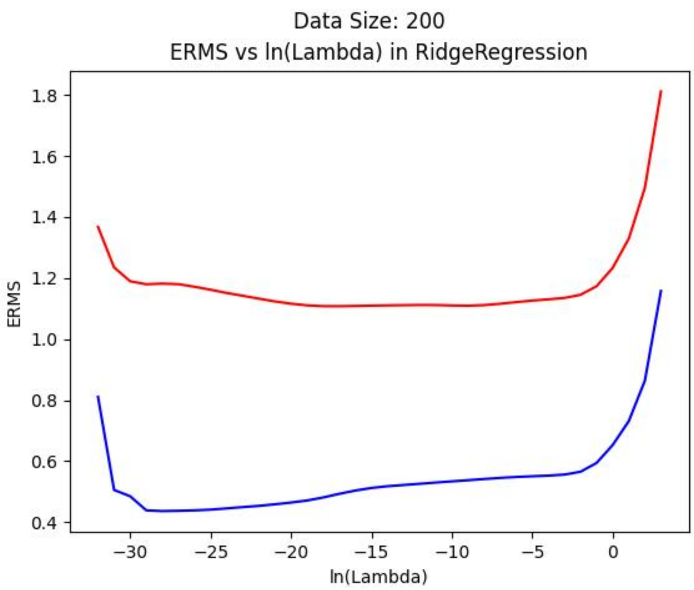


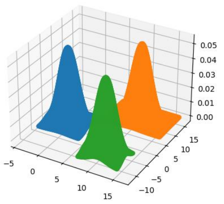
Figure 9: RMS error vs ln(λ)

Blue Color:- Training Data
Red Color:-Development Data

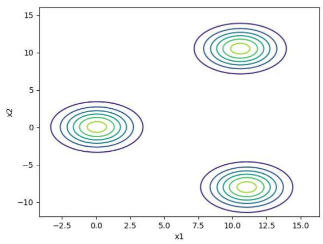
BAYESIAN CLASSIFICATION

The goal of classification is to predict a class label for given example of input data based on previous known data.

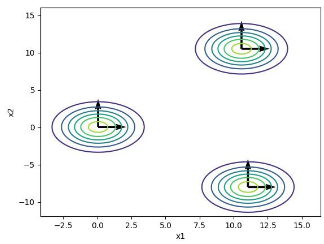
1) Linearly Seperable Data:-Accuracy for All 5 cases is :-100



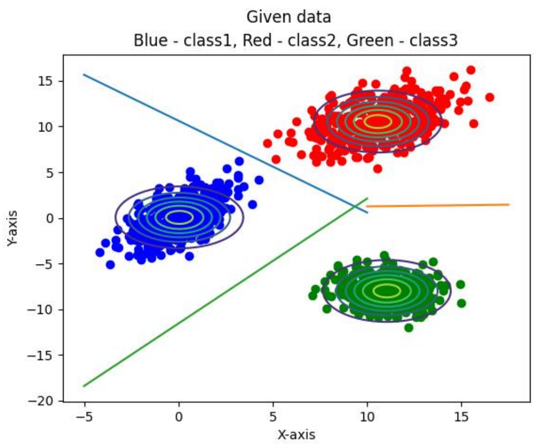
(a) PDF of given data



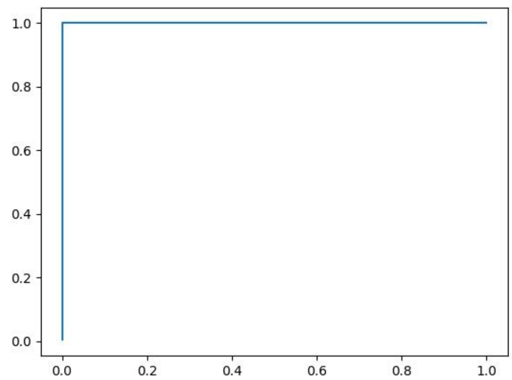
(b) Contours of given data



(c) Contours with Eigen vectors



(a) Contours and decision boundaries with Eigen vectors

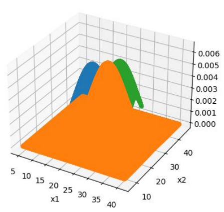


(b) ROC curve for all cases

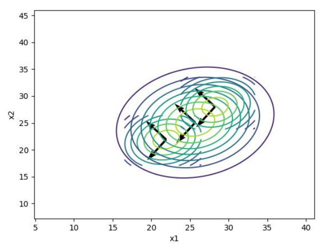
Figure 11: Different plots for Linearly Seperable Data

2) Non Linearly Seperable Data

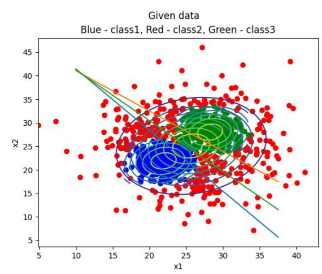
case1:-65.67 case2:-85.67 case3:-66.67 case4:-66.67 case5:-85.34



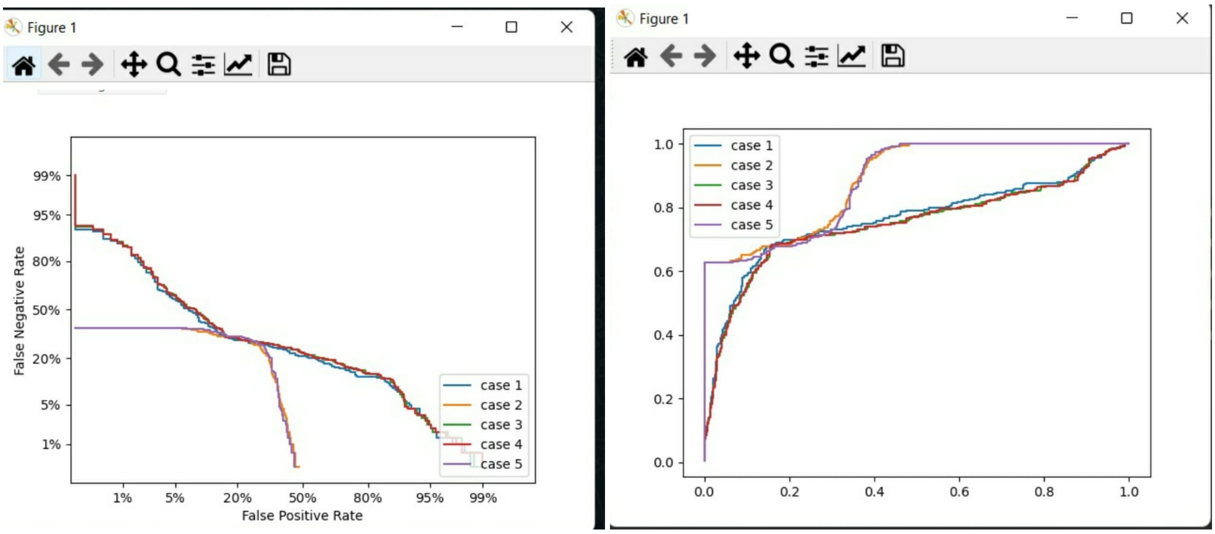
(a) PDF of given data



(b) Contours of given data



(c) Contours with Eigen vectors and decision boundaries for case1



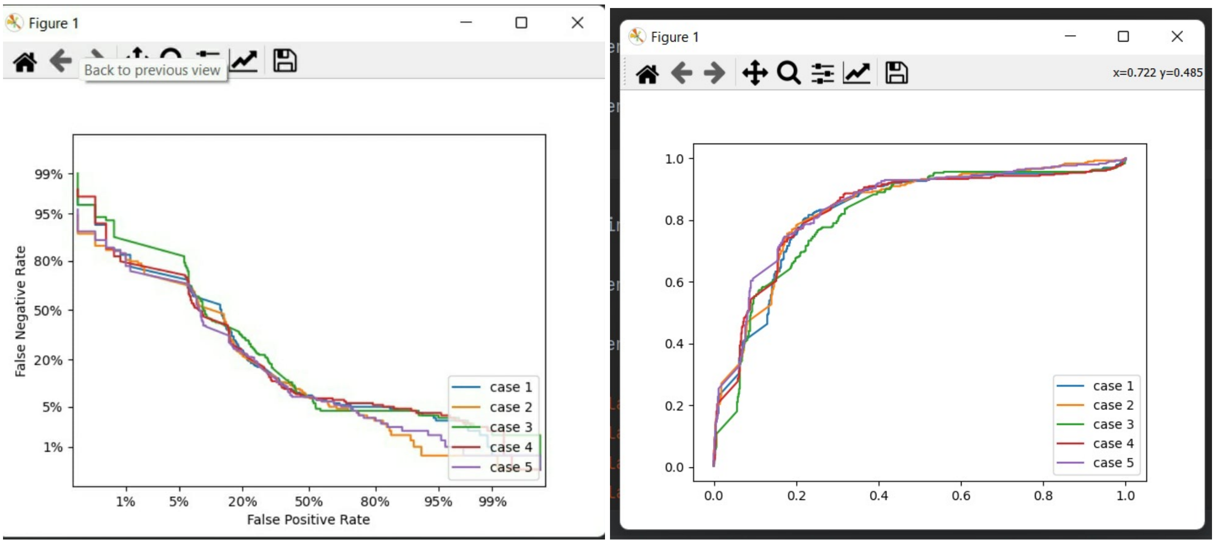
(a) DET curve for all cases

(b) ROC curve for all cases

Figure 13: Different plots for Non Linearly Seperable Data

3) Real Data:-

case1:-74.67 case2:-76.67 case3:-74.0 case4:-75 case5:-75.67



(a) DET Curve

(b) ROC curve for all cases

Figure 14: Different plots for Real Data