Regression on 1-Dimensional Data

1.Least Square and ridge regression :

For 1 dimensional data using linear regression $y = w_0 + w_1 x + w_2 x^2 + ... + w_m x^m$ and calculate Φ as

$$\Phi=\begin{bmatrix} 1 & x_1 & x_1^2 & \dots & x_1^M \\ 1 & x_2 & x_2^2 & \dots & x_2^M \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_N & x_N^2 & \dots & x_N^M \end{bmatrix}$$

n=10 :

we first plot the graphs for different values of m for n=10

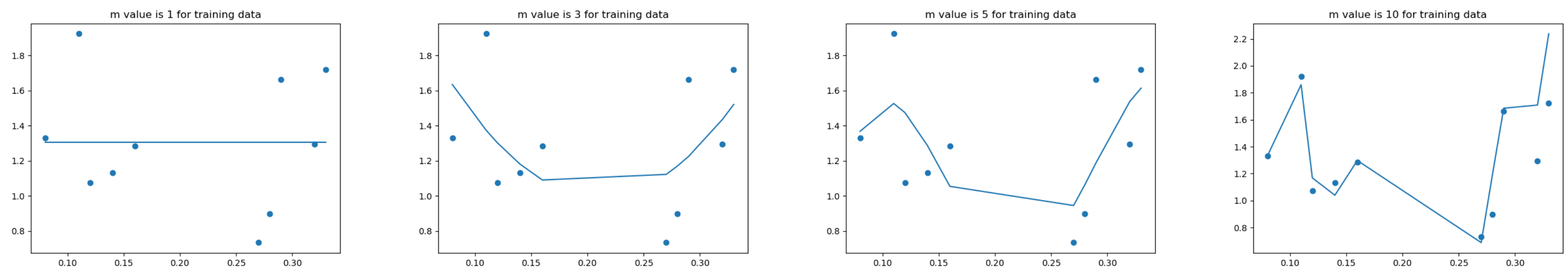

Fig 1:plots for different value of m for n=10 and m=1,3,5,10

Here we can see that m=1 is not a good fit but m=5 is a good fit and m=10 is a over fit ,we are claiming this by looking at E_{RMS} values for different values of m.

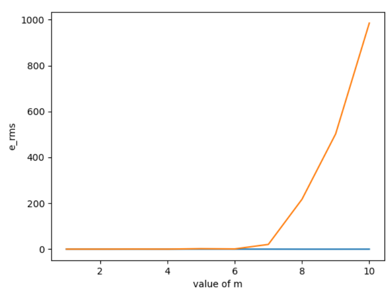

Fig 2: E_{rms} vs m for n=10

as we can see that even though error is less for training data it is huge for test data so we classify it as over fit to reduce the over fit one we can do is increase the data size we see the plots for different plots for m=10.

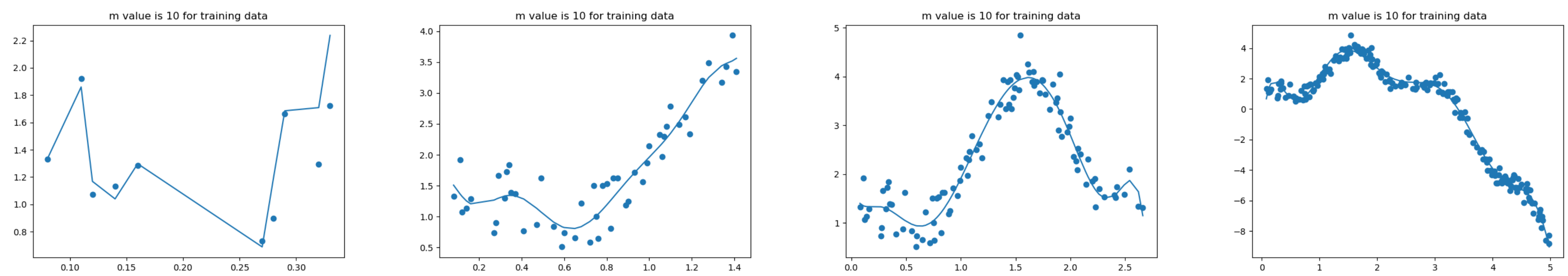

Fig 3:plots for different value of n=10,50,100,200 for m=10

as data size increases the error reduces for a given value of m,so increasing data set reduce over fit problem,the second way is to do ridge regression by adding $\lambda|w|^2$ to the error value , now we vary λ and see the changes of the weight vector $W^{\star}$. The table is as follows

$W^\star$	$ln(\lambda) = -\infty$	$ln(\lambda) = -20$	$ln(\lambda) = -10$	$ln(\lambda) = -0$
$w_0^\star$	-2.65426600e+03	-1.87440494	2.30108986	1.14502739
$w_1^\star$	1.36217901e+05	82.70274521	-10.48737584	2.04412430e-01
$w_2^\star$	-3.00674313e+06	-656.9312796	16.14493745	5.28977878e-02
$w_3^\star$	3.76270492e+07	1750.0057065	20.51814577	1.68255083e-02
$w_4^\star$	-2.95170496e+08	-661.46990324	12.54656403	5.71951155e-03
$w_5^\star$	1.50946402e+09	-1425.62235572	5.91938524	1.96534383e-03
$w_6^\star$	-5.04360735e+09	-1091.21299532	2.45547332	6.72064959e-04
$w_7^\star$	1.06364148e+10	-616.289161	0.94622229	2.28048979e-04
$w_8^\star$	-1.28629521e+10	-297.31385138	0.34825698	7.68309268e-05
$w_9^\star$	6.57835523e+10	-325.45267730	0.24823725	9.371236217e-05

Table 1: values of $W^\star$ for different values of $ln(\lambda)$ for n=10

as we can see that as the value of $ln(\lambda)$ increases the coefficients become smaller and becomes have less oscillations or vary less between the points.

we also see the graphs for different values of $ln(\lambda)$

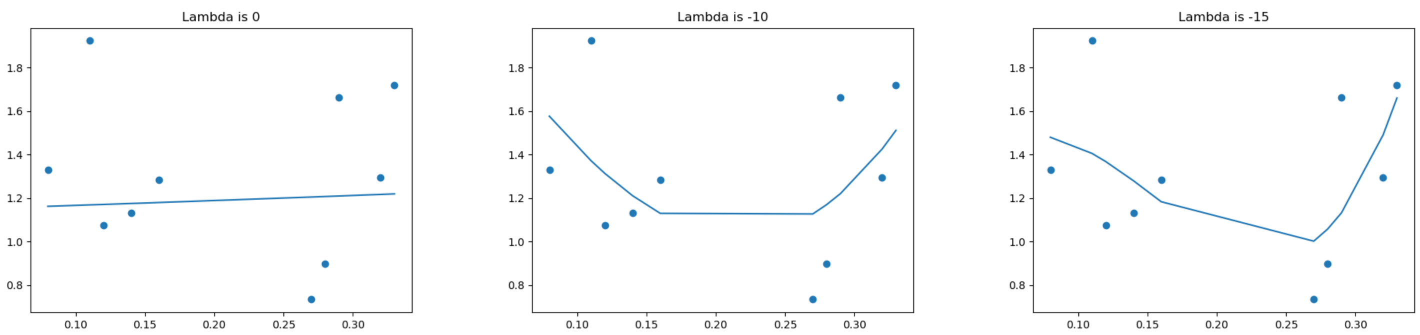

Fig 4:graph for different values of $ln(\lambda)$=0,-10,-15 for m=10

as we increase the value of $ln(\lambda)$ we get the curve which is more of the original curve and get better results for test data too,below is the graph for E_{rms} for training and test data for different values of $ln(\lambda)$.

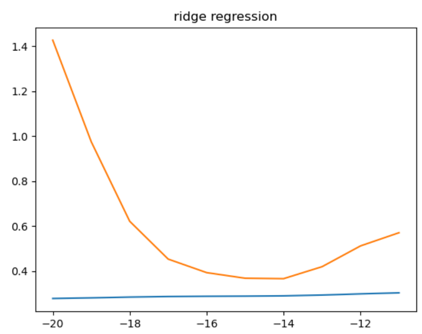

Fig 5:$ln(\lambda)$ vs E_{rms} for m=10

we can see that error for test data first decreases and then again increases because as we further increase $ln(\lambda)$ the error term is dominated by $\lambda|w|^2$ and we not get a good fit.

n=20 :

we do the same analysis for n=10,first we draw for different values of m for n=20

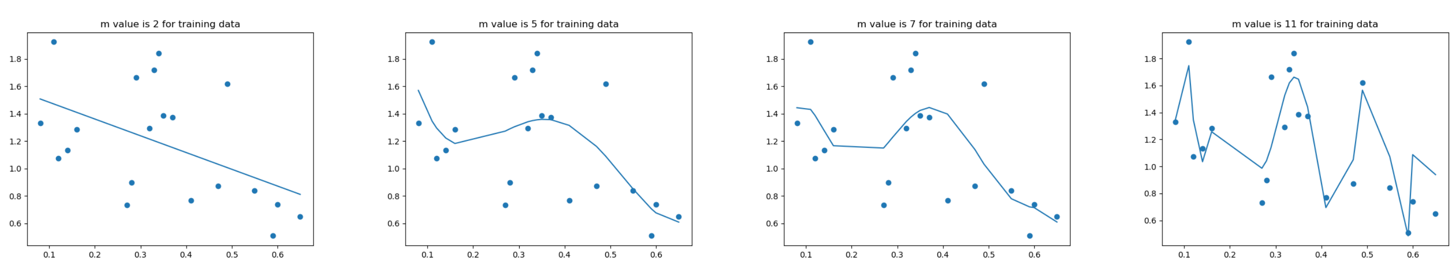

Fig 6:plots for different value of m=2,5,7,11 for n=20

now we draw the graph of E_{rms} vs m for n=20 and observe the trend

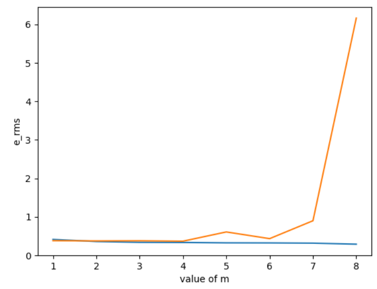

Fig 7: E_{rms} vs m for n=20

we can see that here also error is less for training data and error is high for test data one way is to increase the data size

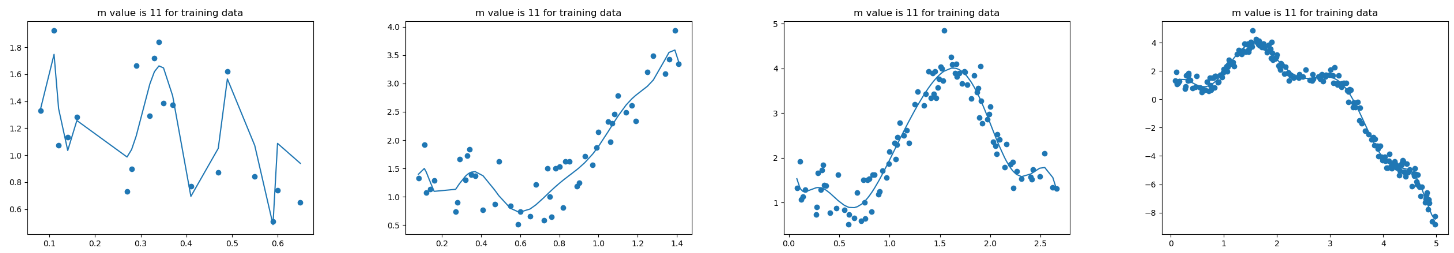

Fig 8:plots for different value of n=20,50,100,200 for m=11

second one is add error value λ term

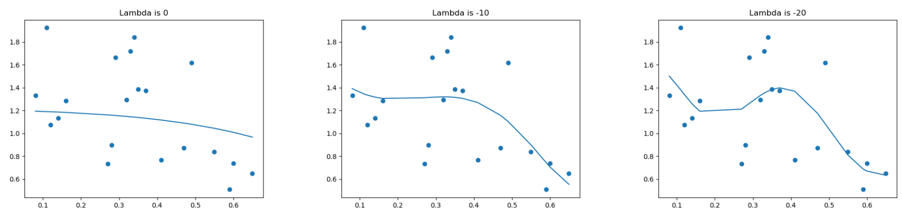

Fig 9:graph for different values of $ln(\lambda) = 0, -10, -20$ for m=11

now we see the W for different values of $ln(\lambda)$

$W^\star$	$ln(\lambda) = -\infty$	$ln(\lambda) = -20$	$ln(\lambda) = -10$	$ln(\lambda) = -0$
$w_0^\star$	-3.99040986e+02	1.26154404e+00	1.61413921	1.20083859
$w_1^\star$	1.77727860e+04	1.37810415e+01	-3.82917631	-0.06699199
$w_2^\star$	-3.31870519e+05	-2.00675830e+02	14.01583867	-0.22495122
$w_3^\star$	3.44751044e+06	9.58589183e+02	-11.07275648	-0.19193505
$w_4^\star$	-2.21601376e+07	-1.74222157e+03	-13.08274042	-0.13551096
$w_5^\star$	9.25061080e+07	5.91451705e+02	-3.99273928	-0.08932634
$w_6^\star$	-2.55097017e+08	1.10436406e+03	4.27776834	-0.05707958
$w_7^\star$	4.60812983e+08	-2.22674593e+01	8.62222499	-0.03592792
$w_8^\star$	-5.23947753e+08	-7.60033921e+02	9.75916384	-0.02245103
$w_9^\star$	3.39847789e+08	-4.68547942e+02	9.07820064	-0.01398534
$w_{10}^\star$	-9.58145346e+07	4.48928033e+02	7.63492288	-0.00870377

Table 2: values of $W^\star$ for different values of $ln(\lambda)$ for n=20

as we can see that the value of $ln(\lambda)$ increases coefficients become smaller,we will also see that graphs for different values of $ln(\lambda)$.

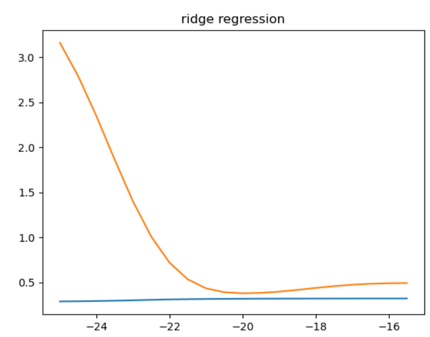

Fig 10: $ln(\lambda)$ vs E_{rms} for m=11

similarly we can see the graphs for n=50,n=100,n=200 for we will plot the best curves and plot $E_{RMS} vs m$

n=50:

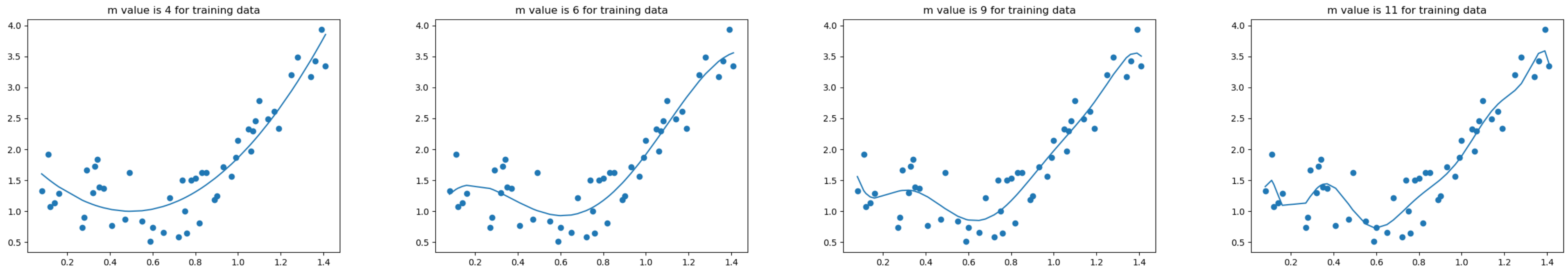

Fig 11:plot of different values of m=4,6,9,11 for n=50

as m values increase we can see that the curve gets more fit and gets less errors for bigger m,here m=11 is a good fit

we will see the error plots for the given values of m

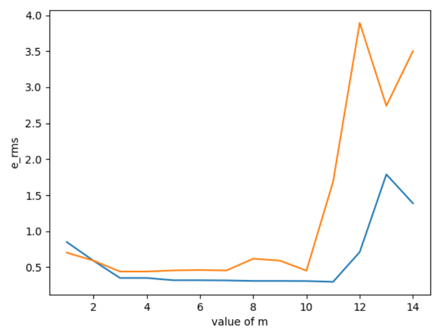

Fig 12:plot of E_{RMS} and m for n=50

after m=11 we can see that the error values for both the training data and test data increases and we can say that m=11 is best fit

n=100:

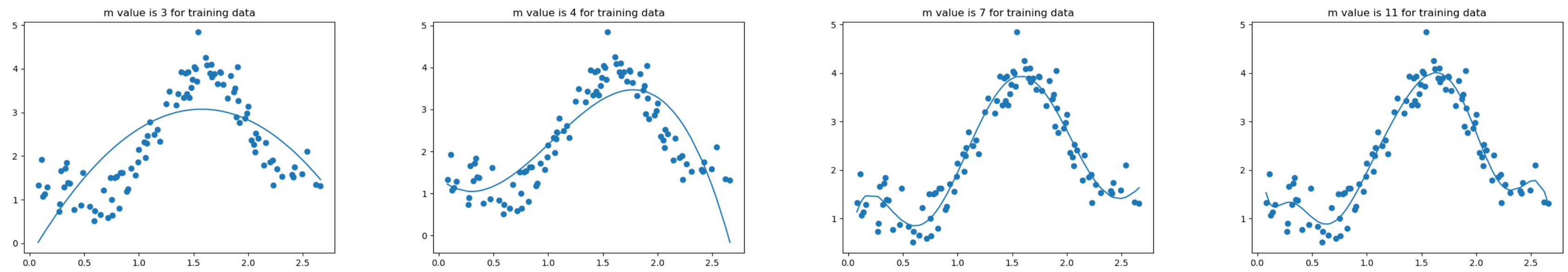

Fig 13:plot of different values of m=3,4,7,11 for n=100

similar to above we will also plot the E_{rms} vs m for n=100

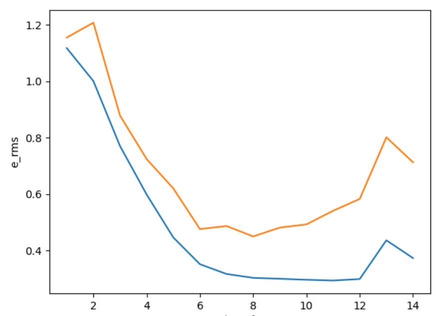

Fig 14:plot of E_{RMS} and m for n=100

Here also we can see that m=11 is also the best fit and the error values increases as the value of m increases further

n=200:

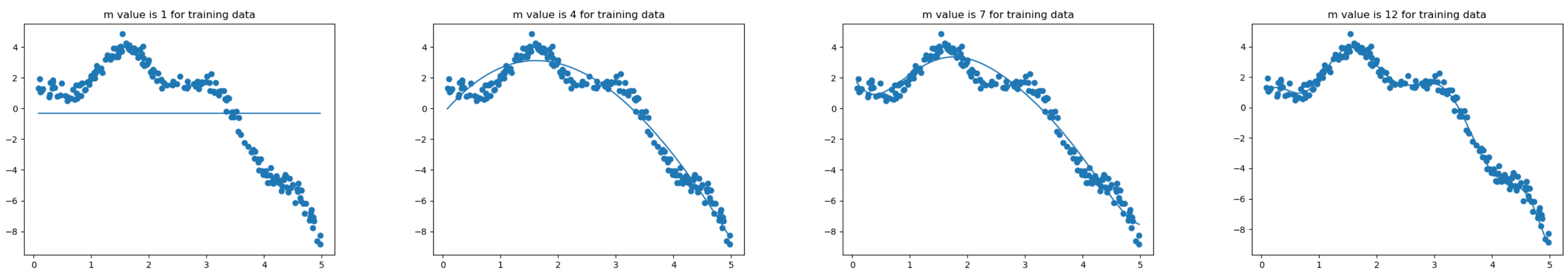

Fig 15:plot of different values of m =1,4,7,12for n=200

the error graph is

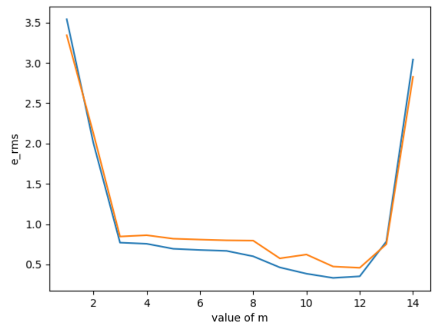

Fig 16:plot of E_{rms} and m for n=200

now for the best fit curves for Y vs T where Y is calculated data and T is the given data,ideally we need to get the straight line $y = x$ since ideally calculated data is the given data for both train and test data

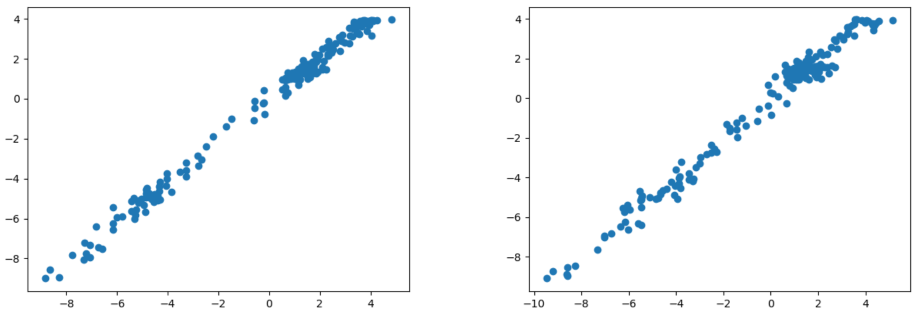

Fig 17:plot of y vs t for n=200 for both test and train data left one is train data and the right is test data.

Regression on 2-Dimensional Data

Least Square and ridge regression :

In linear regression for 2-D data we will use the following equation to calculate output where $y = w_{00} + w_{10}x_1 + w_{11}x_2 + w_{20}x_1^2 + w_{21}x_1x_2 + x_{11}x_2^2 +$ .... to calculate y and find Φ.

n=50 :

Now for the 2-D data as we can see we will not plot 3-D plots it is difficult to give an angle to imagine in 2-D so we will only present the data and 2-D graphs,the first one E_{rms} vs m for n=50

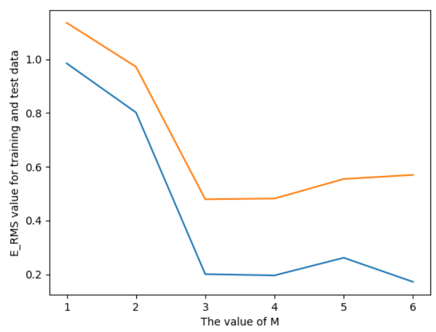

Fig 2.1:Graph of E_{rms} vs m for n=50

we can see that the error value for test data goes up for the values of m>=6 so we can see that there is ridge regression when we do the ridge regression here we can decrease the over fit curve and fit the curve for test data, we will see how the $ln(\lambda)$ vs E_{rms} varies for different values of λ.

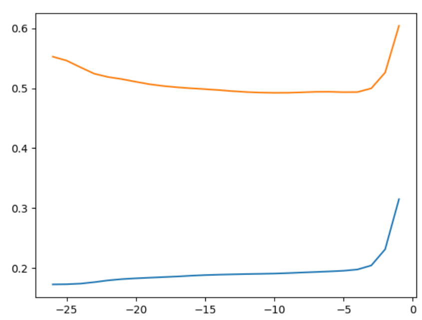

Fig 2.2:graph of E_{rms} vs $ln(\lambda)$ for n=50,m=6

we will also see the W values when λ values vary and how the decrease of values of w can decrease the test data error and give us a good fit to the data.

$W^{\star}$	$ln(\lambda) = -\infty$	$ln(\lambda) = -20$	$ln(\lambda) = -10$	$ln(\lambda) = 0$
$w^{\star}_{00}$	-1.55241753e+04	-2.91070840e+02	-8.87216754	-1.84589787
$w^{\star}_{10}$	-4.98806904e+04	-4.79398666e+02	0.07034758	1.16409569
$w^{\star}_{11}$	1.06047227e+05	6.91981527e+02	-2.10110291	0.20289254
$w^{\star}_{20}$	-3.47831750e+04	1.50743657e+02	3.63711479	-0.54793822
$w^{\star}_{21}$	-8.55675090e+03	1.47790792e+03	-2.38730018	-0.0864899
$w^{\star}_{22}$	4.06416616e+04	-4.33310850e+02	-2.56334676	-1.04165984
$w^{\star}_{30}$	-7.15618564e+05	3.08433670e+02	-3.79309479	-0.00855165
$w^{\star}_{31}$	-2.68371129e+04	9.98377948e+01	3.34150736	-0.01808198
$w^{\star}_{32}$	-2.98366996e+01	-1.24419067e+03	-2.17611866	0.72987683
$w^{\star}_{33}$	6.37283640e+04	-1.05186446e+02	-2.31780427	0.01093407
$w^{\star}_{40}$	-5.05603949e+05	-3.33378091e+02	-0.82124271	0.51080501
$w^{\star}_{41}$	-2.80279984e+04	-1.47870397e+03	1.23043885	0.11201192
$w^{\star}_{42}$	-5.66468531e+01	-1.17090421e+03	-1.84388795	-0.44848912
$w^{\star}_{43}$	-2.26978755e+01	-2.16964522e+02	-9.11791427	0.03802608
$w^{\star}_{44}$	2.26615469e+04	-1.93209768e+01	1.62577986	-0.39842605

Table 2: values of $W^{\star}$ for different values of $ln(\lambda)$ for n=50

as we can see as the value increases then we can see the oscillations decrease and get the perfect fit and increasing from there dominates the original error and we do not get a good fit.

n=1000:

Here also we will see the graphs for n=1000 for 2-D regression first we give the E_{rms} vs m for different values of m.

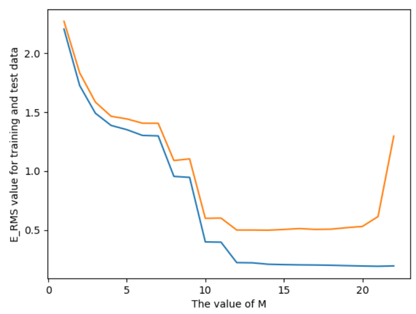

Fig 2.3:graph of E_{rms} vs m for n=1000

as we can see that the error for train set decreases and the test set increases after some m as we can see that it is because of over fit of the data for higher values of m . we can reduce the error by adding the λ term we will see the error values in the below graph.

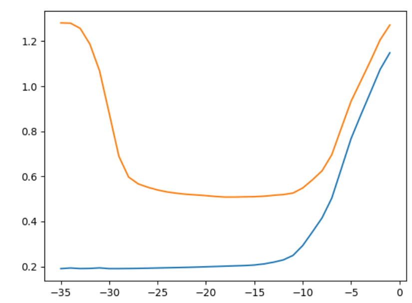

Fig 2.4: graph of E_{rms} vs $ln(\lambda)$

we can see that the error first decreases for test data decreases as $ln(\lambda)$ and then increases because of the reason mentioned above,similarly for n=100,n=500 the graphs are given below

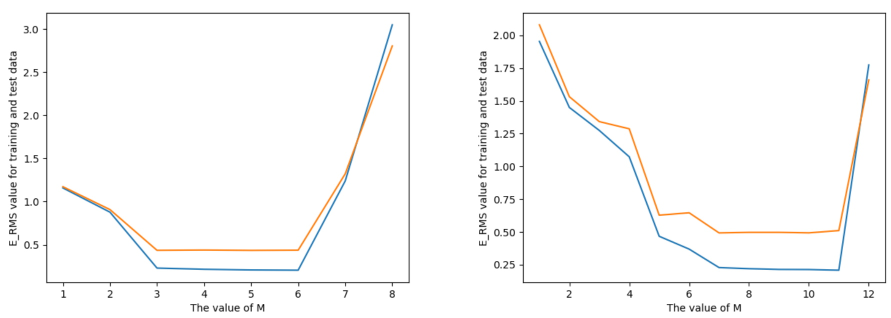

Fig 2.5:graph for E_{rms} vs m for n=100,500

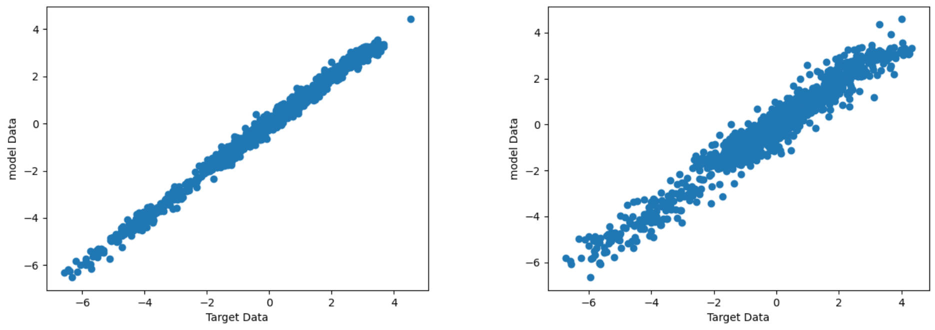

Fig 2.6:graph for y vs t for train and test dat

Classification for linearly seperable data

probability distribution functions for five cases

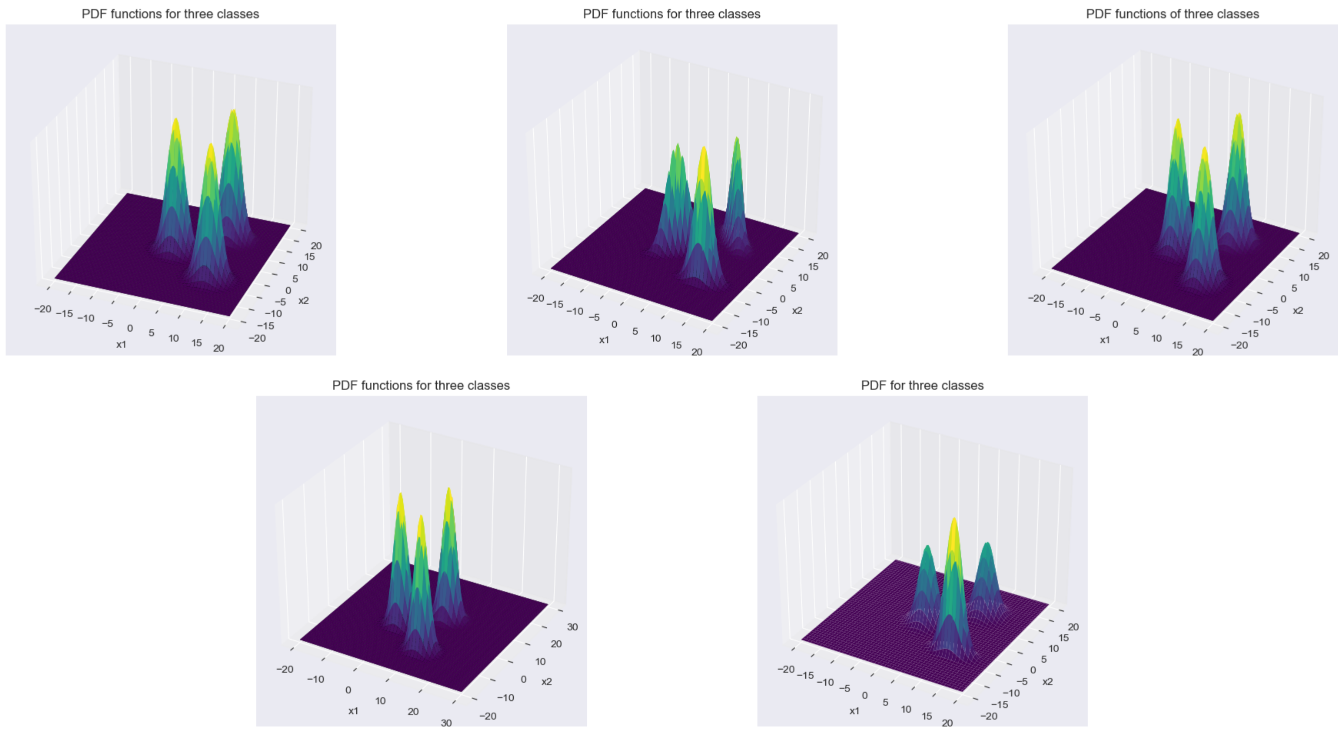

Fig 1:PDF plots for different values of Σ

now we plot the eigen vectors and contours of each case

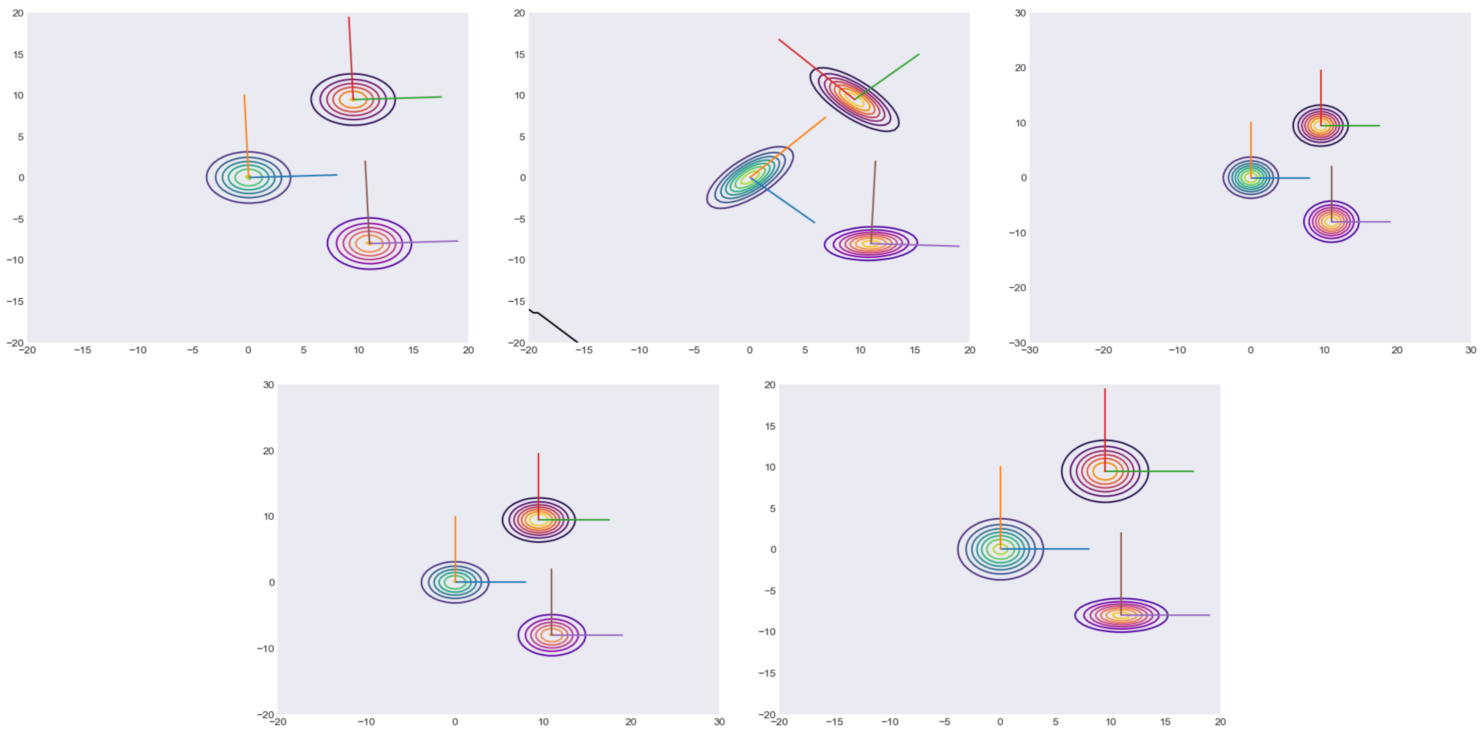
Fig 2:PDF plots for different values of Σ

we can see that for different values of Σ we can see the directions of eigen vectors and whether they are parallel to the axes or not.

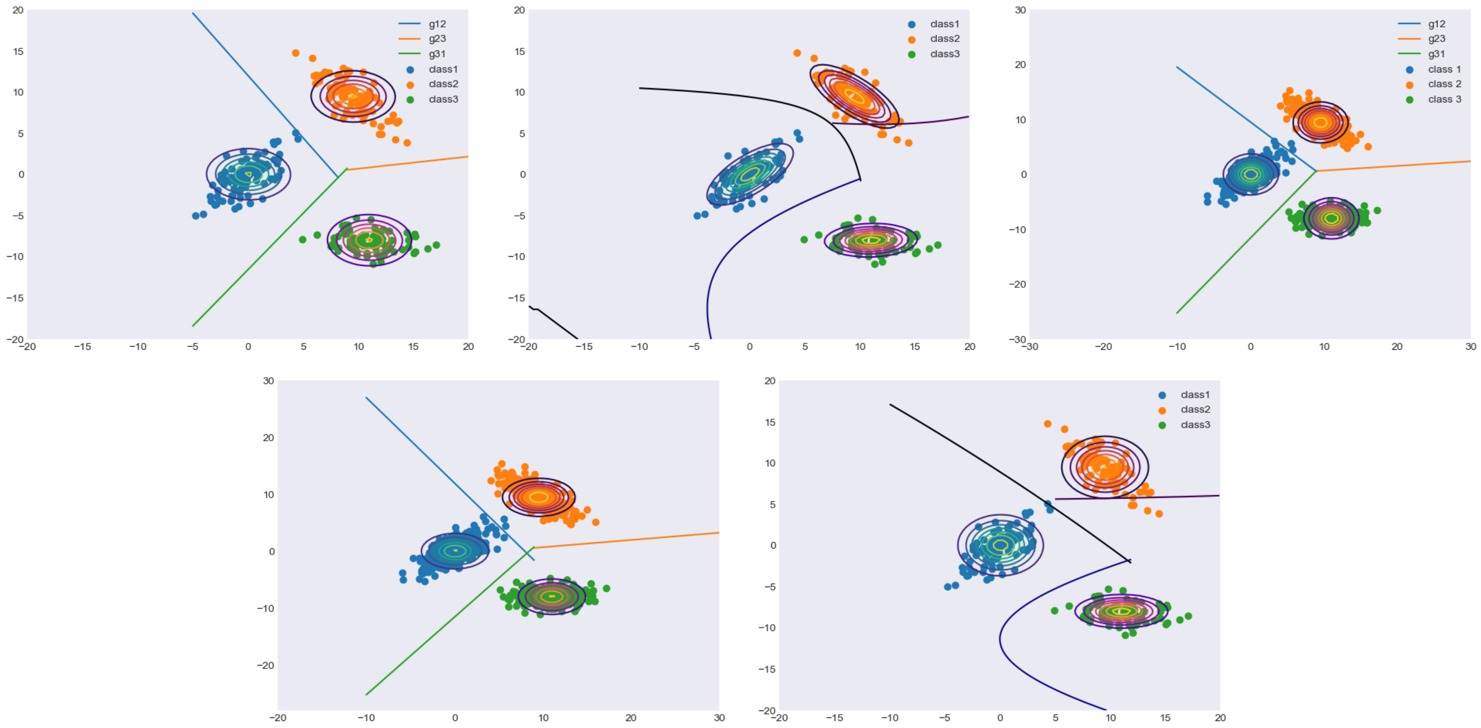

Fig 3:decision curves for different values of Σ

the roc curves and det curves for the non linearly seperable data and linearly seperable data and real data is given.

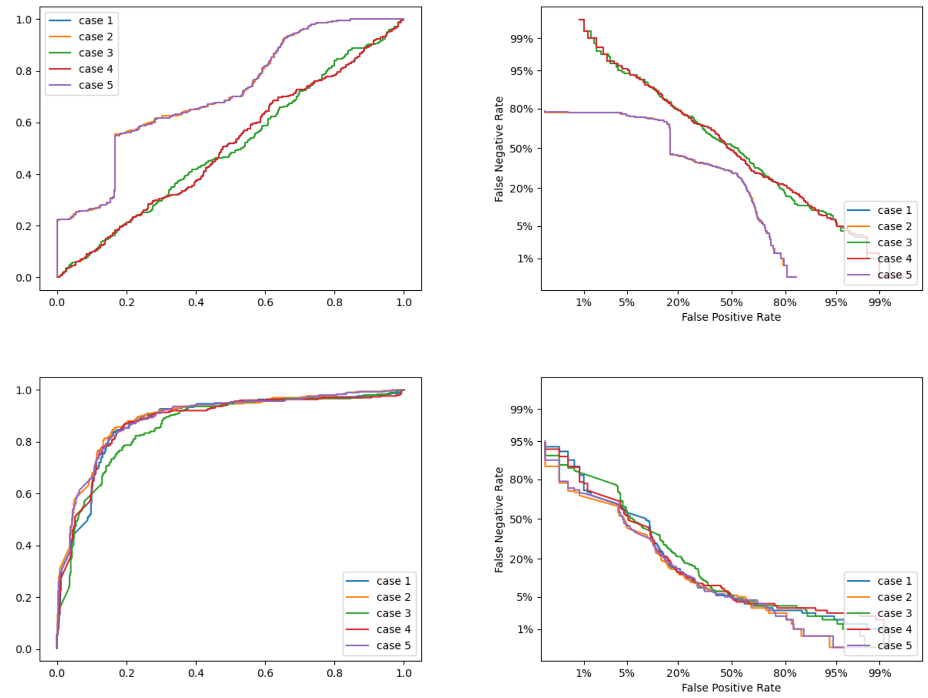

Fig 4:roc and det curves for non seperable and real data

PDF for real data

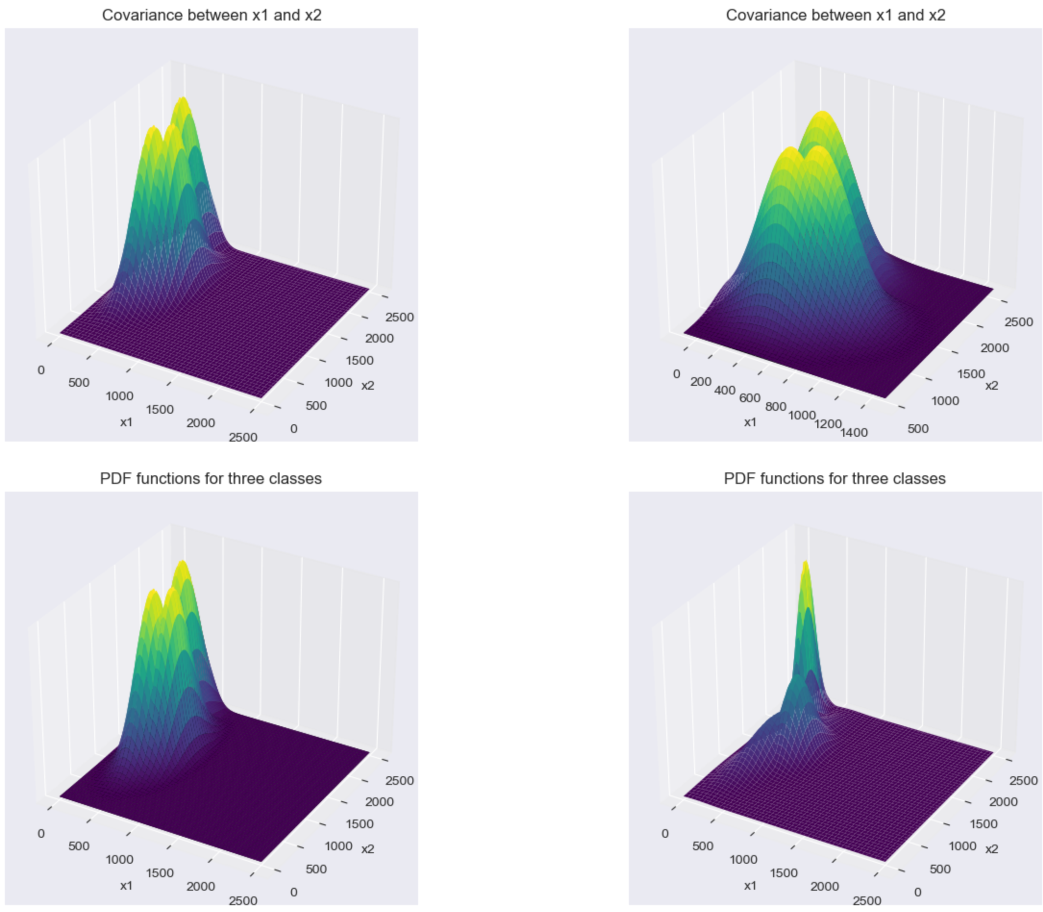

Fig 5:decision curves for different values of Σ

now let us look at the confusion matrices for different classes using different classifiers

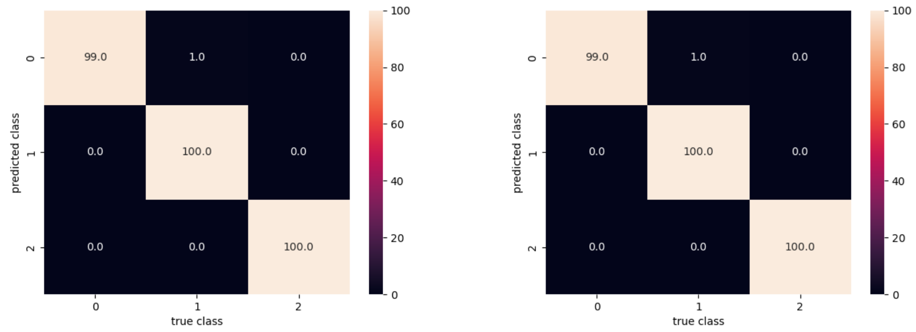

Fig 6:confusion matrices for linearly seperable data

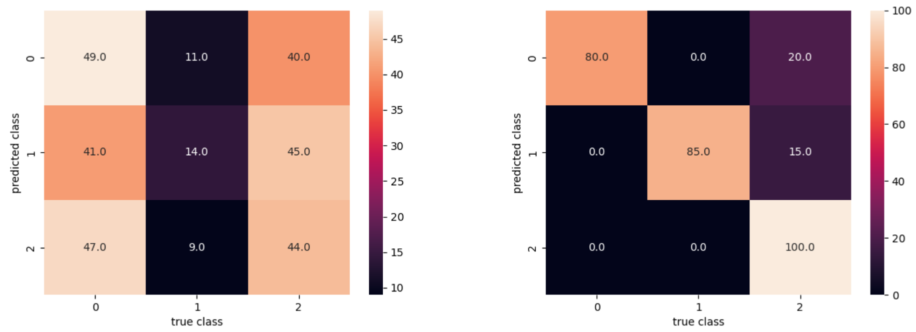

Fig 7:confusion matrices for non linearly seperable data

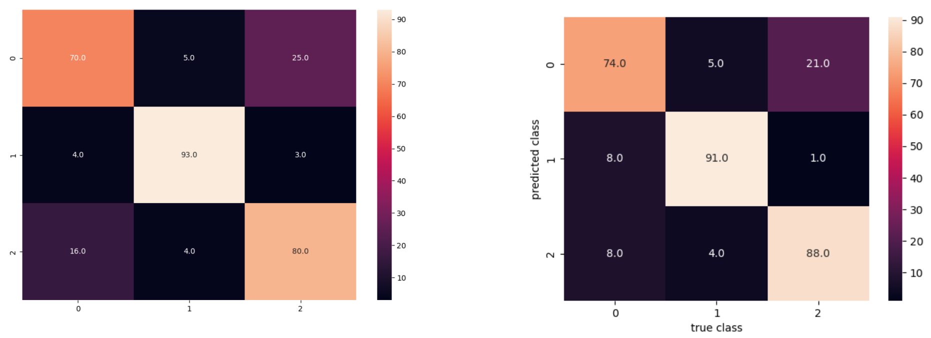

Fig 8:confusion matrices for real data

now we will also see the eigen vectors and values for real data and non linearly seperable data

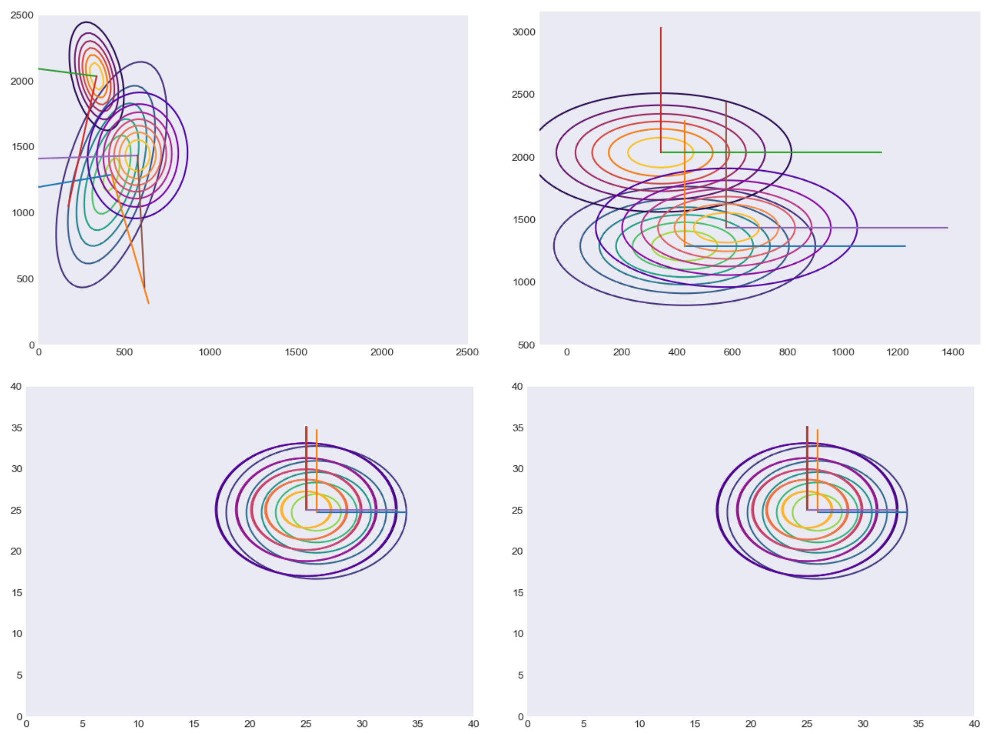

Fig 9:eigen vector and contours for real data up and non linearly seperable data down

The PDF for non linearly seperable data is almost same for three classes and one PDF overlaps the other one and we cannot use linear classifiers to classify non linearly seperable data. Linearly seperable data classifiers gives high percentage success rate so we dot get ant significant change for ROC and DET curves.

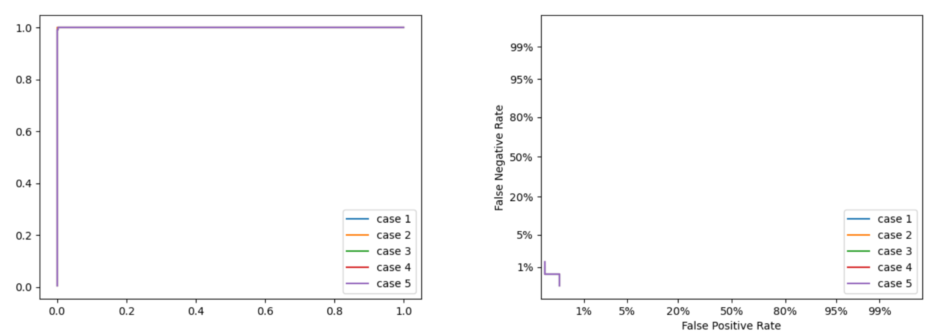

Fig 11:ROc and DET curve for linearly seperable data