EECE 5644 Fall 2022 - Homework 2

Pavan Rathnakar Shetty

Due Date: 10/14/2022, 11:59 pm

Question 1:

x = 2 dimensional random vector
L = class label
p(x) = p(L=0)p(x|L=0)+p(x=1)p(x|L=1)

Priori Probability:
P(L=0)=0.65
P(L=1)=0.35
P(x|L=0)=w1g(x|m01,c01)+w2g(x|m02,c02)
P(x|L=1)=g(x|m1,c1)

Where g is a multivariate gaussian probability density function
m-mean
c-covariance
w1=w2=0.5

N=10000 data samples

m01=[3;0]
m02=[0;3]
cov01=[2 0;0 1]
cov02=[1 0;0 2]
cov1=[1 0;0 1]
m1=[2;2]

Part A:
1)

Part A)

1)

$$\frac{p(x|L=1)}{p(x|L=0)} > \gamma$$

Minimum Risk classification rule:

Since there are 2 class, the following minimum expected risk classification rule in the form of likelihood ratio test applies:

$$D=1 \text{ if } \frac{p(x|L=1)}{p(x|L=0)} > \left(\frac{\lambda_{10}-\lambda_{00}}{\lambda_{01}-\lambda_{11}}\right) * \frac{P(L=0)}{P(L=1)} = \gamma$$

$$D=0 \text{ if } \frac{p(x|L=1)}{p(x|L=0)} < \left(\frac{\lambda_{10}-\lambda_{00}}{\lambda_{01}-\lambda_{11}}\right) * \frac{P(L=0)}{P(L=1)} = \gamma$$

D = decision label

L = ground truth class label

λ_{ij} = represents loss incurred for choosing class 'i' when true class is 'j'

γ = threshold value required to make the decision.

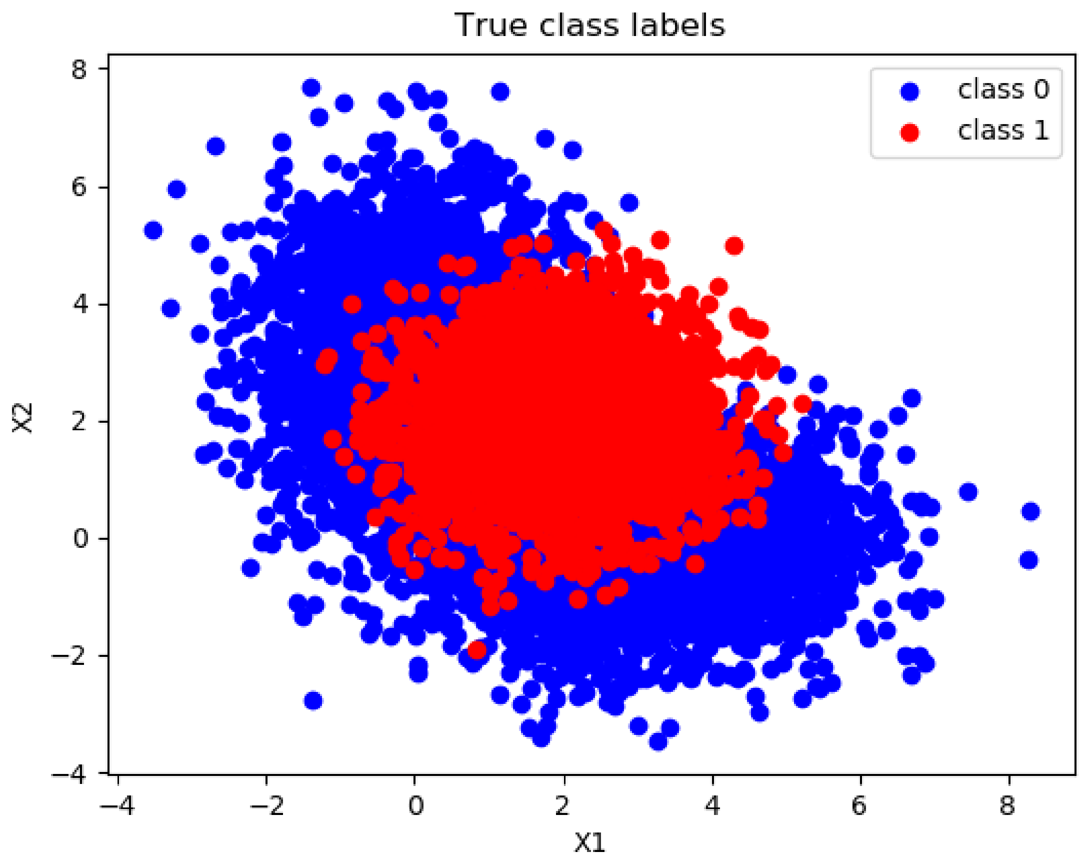
$$\therefore \gamma = \left(\frac{\lambda_{10}-\lambda_{00}}{\lambda_{01}-\lambda_{11}}\right) * \frac{P(L=0)}{P(L=1)}$$

$$= \left(\frac{1-0}{1-0}\right) * \frac{0.65}{0.35}$$

$$\gamma_{\text{theoretical}} = 1.857$$

2)

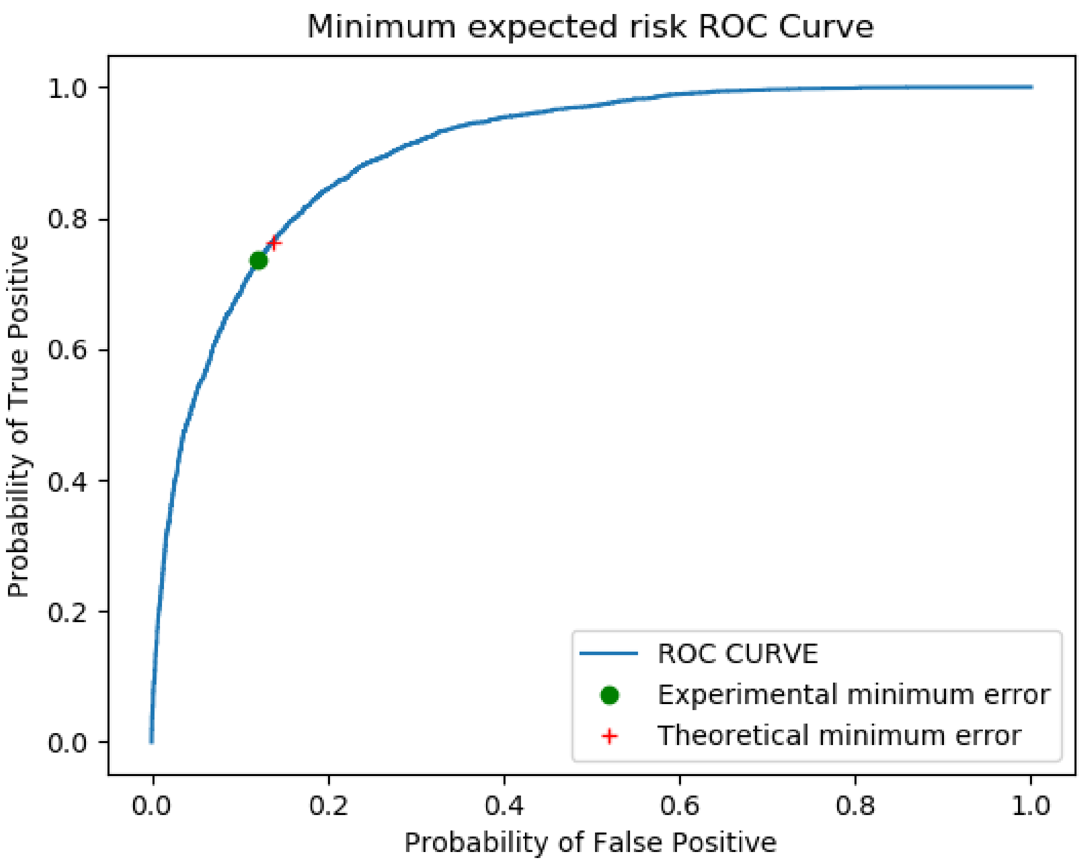
2. N0 = 6584 Number of sample point for class 1
N1 = 3416 Number of sample point for class 2
no_first_distribution = 3346 class 0 sampled from mean m01 and cov 01
no_second_distribution = 3238 class 0 sampled from mean m02 and cov 02



Data Distribution is as shown above (class vs feature vectors)

Implementing ERM classifier:

The classifier was implemented for multiple threshold (γ) values and based on the detection (D=1, L=1) and False Alarm (D=1, L=0) probabilities, the following ROC curve was plotted. ROC curve for minimum expected risk classifier applied on 10,000 samples where gamma is varied from 0 to infinity.



ROC curve for ERM-based classification

3)

Based on the parametric sweep of the threshold values, an estimated minimum probability of error, the value was found, and the theoretical minimum probability of error value was computed using the threshold value obtained from the ratio of the class priors. The true positive rate and false positive rate values corresponding to these threshold values have been superimposed on the ROC curve above.

3) Part A: In this experiment, experimental gama is 1.8458 and experimental minimum probability error is 17.42%.

Theoretical gama is 1.857 and theoretical minimum error is 17.52%.

As observed the $\gamma_{experimental} \to \gamma_{theoretical}$ as ~~[illegible]~~ $n \to \infty$

Part B)

Part B:

Linear Discriminant Analysis based classifier (LDA)

$$D = 1 \text{ if } w_{LDA}^{T} \cancel{*} X > \tau$$

else

$$D = 0$$

w_{LDA} is from generalised S_B, S_w with largest eigen value.

$$S_B = (\mu_0 - \mu_1)(\mu_0 - \mu_1)^T$$

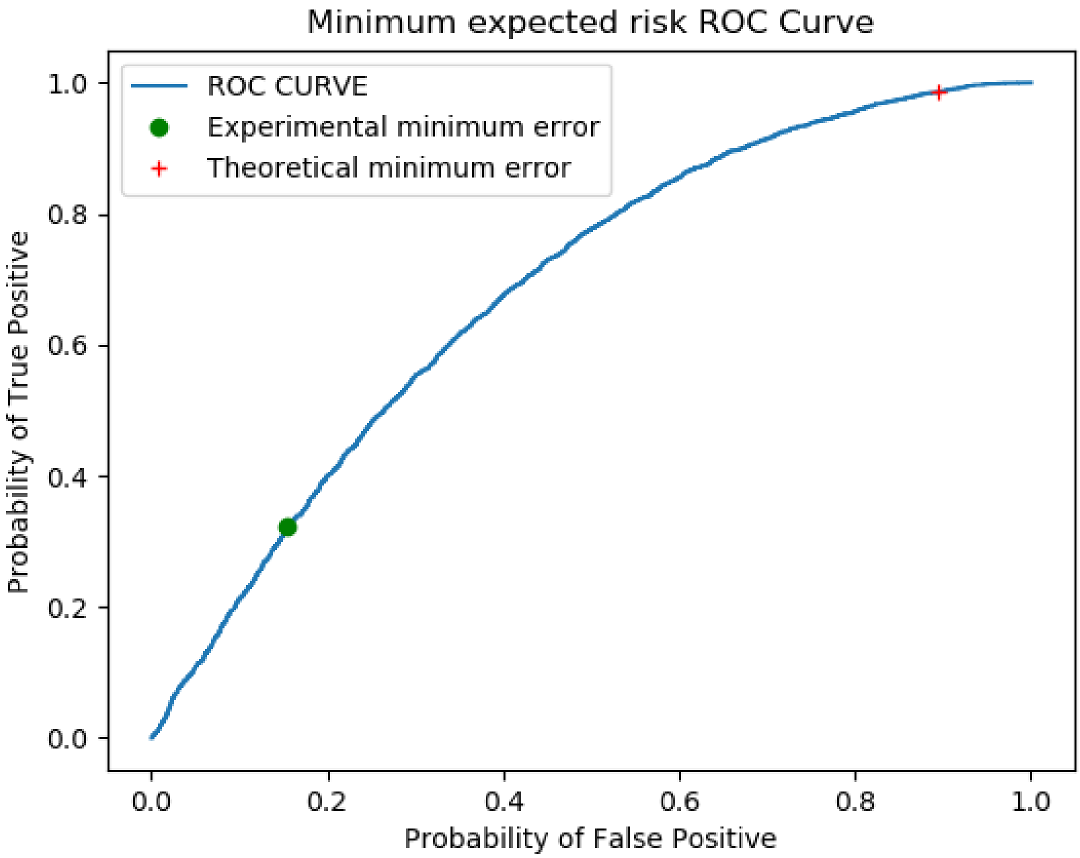
$$S_w = \Sigma_0 + \Sigma_1$$

The LDA ROC curve is relatively poor compared to plot in A as shown. The LDA ROC has higher minimum error both experimental and theoretical.

$$\gamma_{exp} = 3.01890$$

Experimental minimum error = 0.3500 i.e 35%

Theoretical minimum error = 0.3595 i.e 35.95%



ROC curve for LDA-based classification

2)

Question 2

In order to generate samples for the 3-dimensional vector X, 4 Gaussian distributions with the following parameters were considered. The parameters were chosen such that the distance between mean values was twice the average standard deviation.

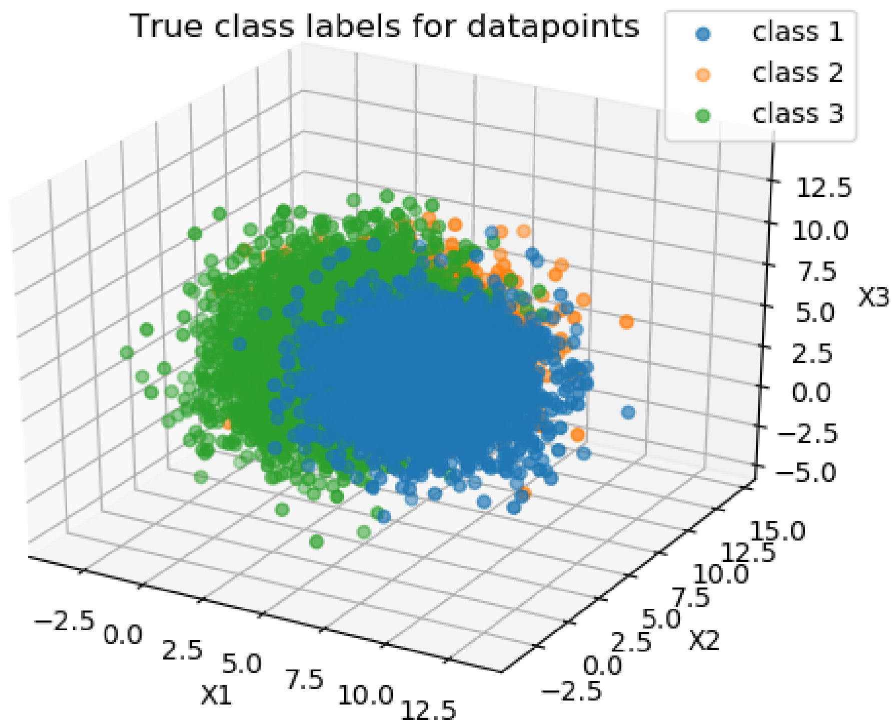
```
mu = [[7, 4, 4], [4, 7, 4], [4, 4, 7], [3, 3, 3]]
C = [
[[4, 0, 0], [0, 4, 0], [0, 0, 4]],
[[4, 0, 0], [0, 4, 0], [0, 0, 4]],
[[4, 0, 0], [0, 4, 0], [0, 0, 4]],
[[4, 0, 0], [0, 4, 0], [0, 0, 4]],
]
```

Part A

Part 1: Generating sample data

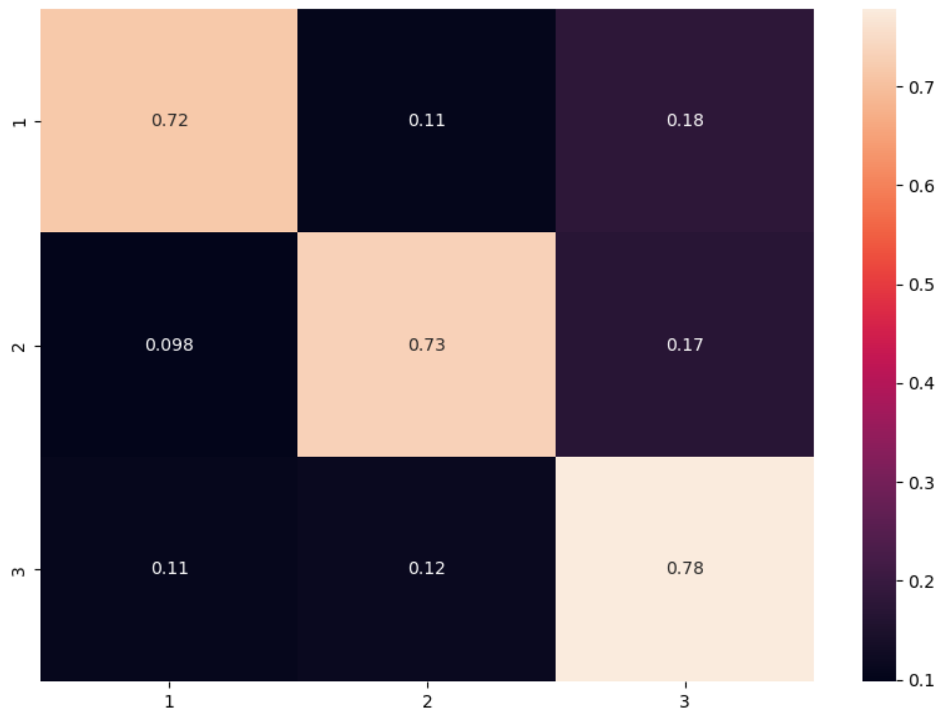
To generate the 10000 samples, the samples belonging to class 1 and 2 were generated from their corresponding gaussian distributions with parameters mu1, C1, and mu2, C2. To generate samples belonging to class 3, each data point was taken from either Gaussian 3 (mu3, C3) or Gaussian 4 (mu4, C4) with a 50% probability, thus satisfying the condition of the class 3 data being generated from a mixture of 2 Gaussians with equal weights. Following is a plot of the samples for each class

Class priors are: 0.3, 0.3, 0.4

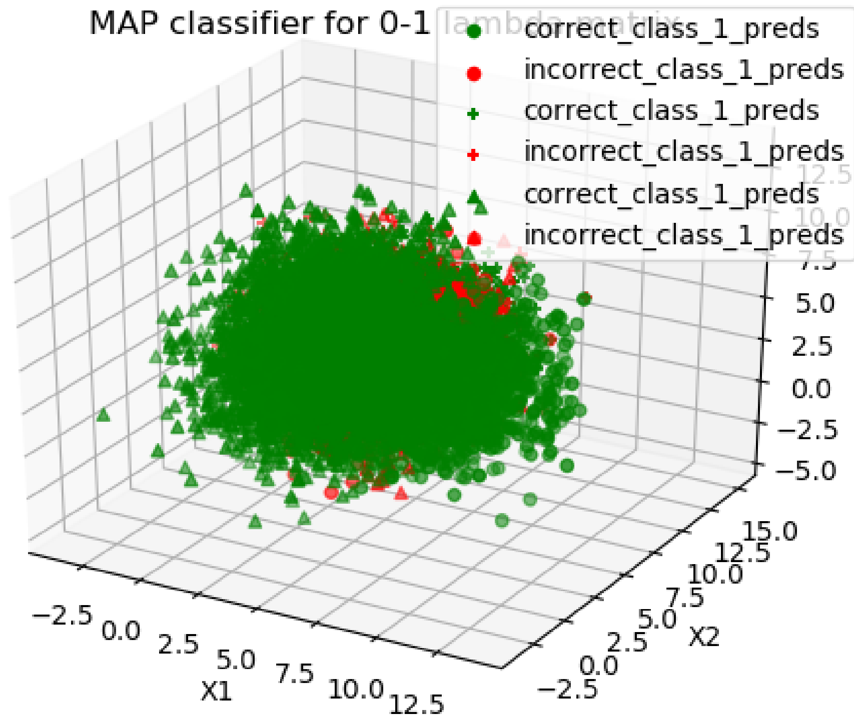


a plot of sample data

Based on this classification rule, the classifier was implemented to classify the 10K samples and the following confusion matrices were obtained:



confusion matrix



Data Visualization

Part A

2) Landa matrix 0-1 : $\Lambda_1 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$

~~Rule~~ Decision rule to achieve minimum probability of error for the sample data is specified below:

$$D(x) = \arg\min R(\alpha_i | x)$$

$$= \arg\min_{D \in \{1,2,3\}} \sum_{l=1}^{c} \lambda_{Dl}\, p(x | L = l)\, P(l = l)$$

C - number of classes

~~λ_{Dl} represents~~

Based on 0-1 loss, accuracy of classification was found to be

$$\text{Accuracy} = \sum_{i=1}^{3} \frac{(D = i | L = i)}{N} * 100 = 25.48\%$$

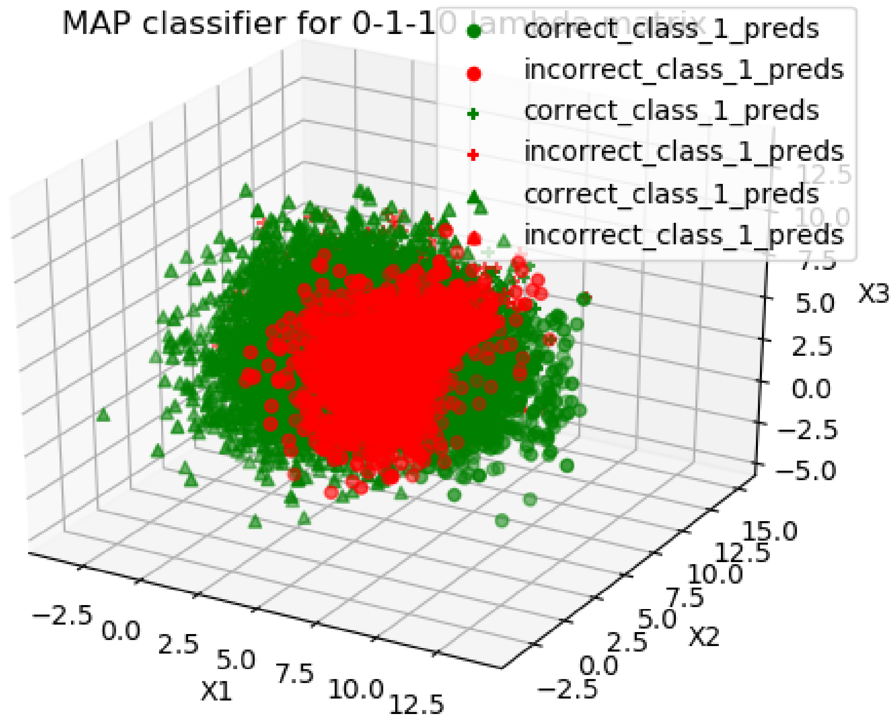
PART B

Part B:

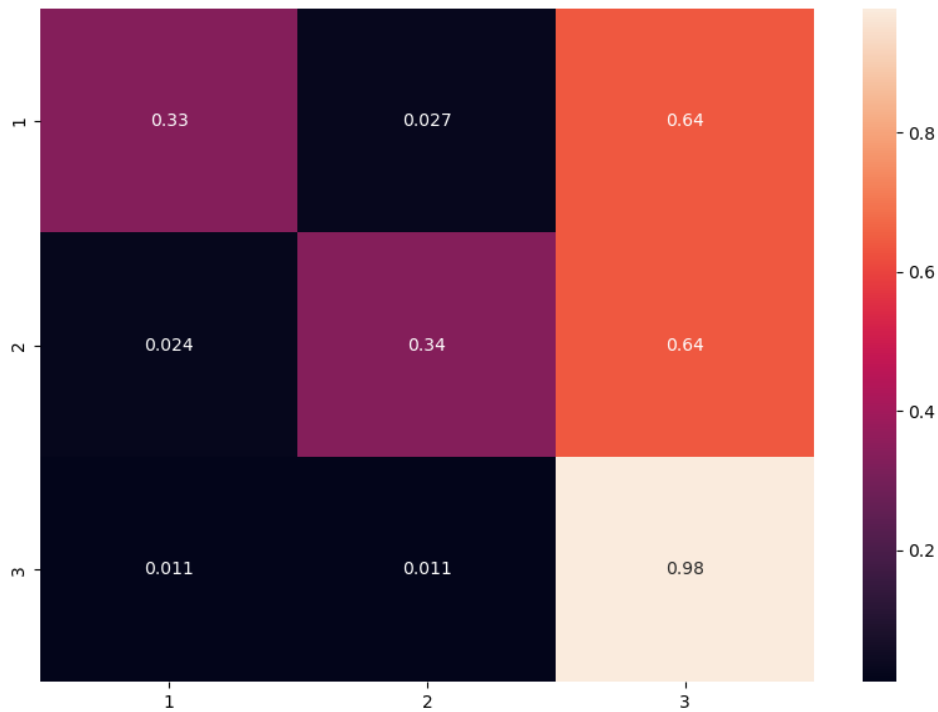
$\Lambda_2 = \begin{bmatrix} 0 & 1 & 10 \\ 1 & 0 & 10 \\ 1 & 1 & 0 \end{bmatrix}$ $\quad \Lambda_3 = \begin{bmatrix} 0 & 1 & 100 \\ 1 & 0 & 100 \\ 1 & 1 & 0 \end{bmatrix}$

we do the same as Part A with given ~~B~~ landa matrix.

For 0-1-10 lambda matrix, we have the following data visualization after the classification and confusion matrix as shown below:

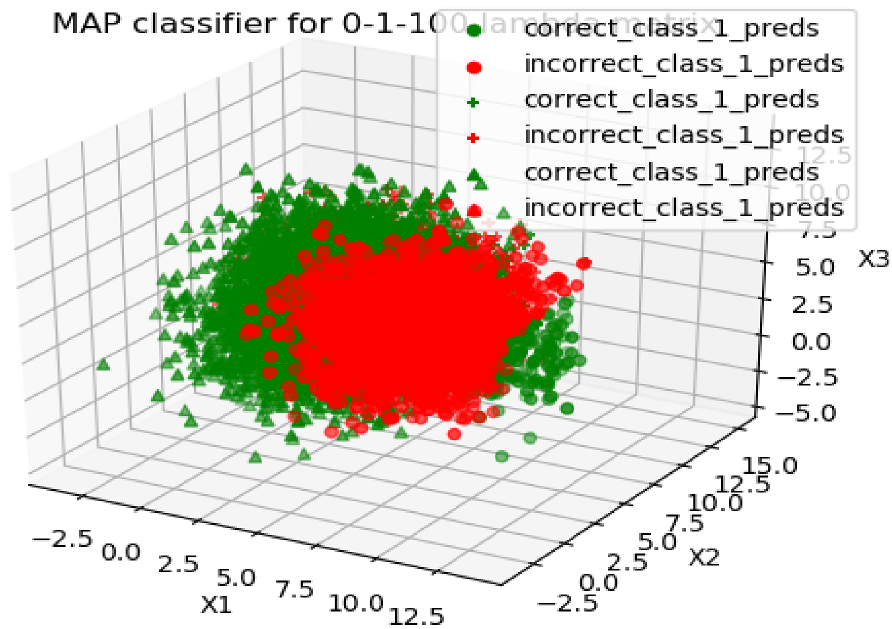


Data Visualization after classification

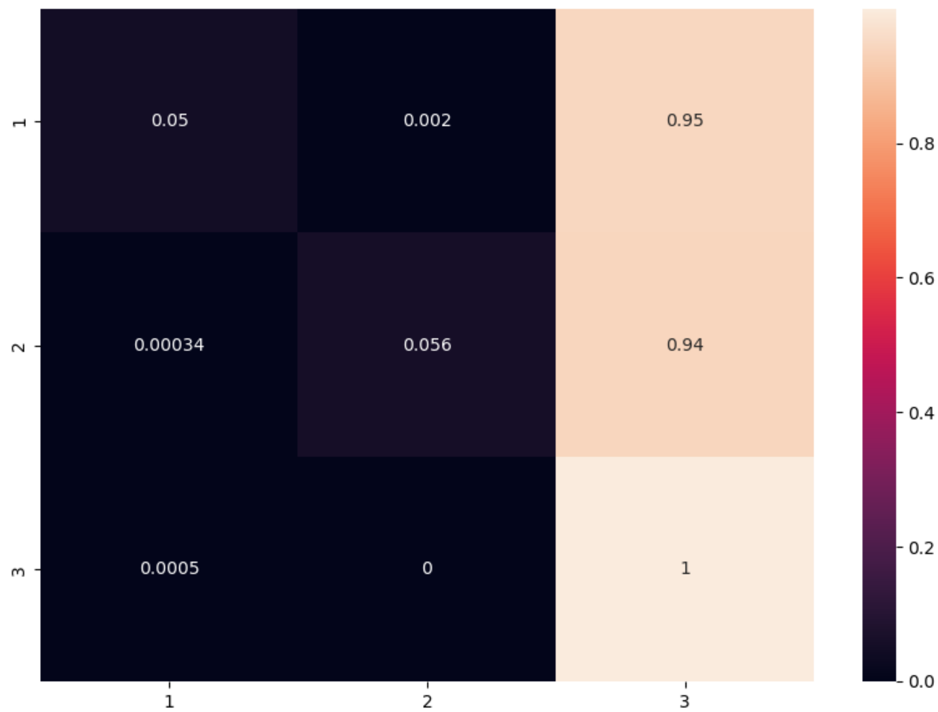


confusion matrix

For 0-1-100 lambda matrix, we have the following data visualization after the classification and confusion matrix as shown below:



Data Visualization after classification



confusion matrix

Part B:

$$\Lambda_2 = \begin{bmatrix} 0 & 1 & 10 \\ 1 & 0 & 10 \\ 1 & 1 & 0 \end{bmatrix} \quad \Lambda_3 = \begin{bmatrix} 0 & 1 & 100 \\ 1 & 0 & 100 \\ 1 & 1 & 0 \end{bmatrix}$$

we do the same as Part A with given ~~B~~ landa matrix.

Total loss for Λ_1 = 25.48 %

Total loss ~~of~~ for Λ_2 = 40.74 %

Total loss for Λ_3 = 56.78 %

Observation:

- The true positive for class 3 increases with the increase in error penalty in class 3.
- As ~~seen~~ inferred in the confusion matrix, if the error penalty is significant, we notice $P\left(\frac{\text{predicted class 3}}{\text{total class 3}}\right) = 1$
- There is also increase in misclassification due to the bias present in the classifier tending towards class 3 due to high penalty for misprediction of class 3.
- The overall loss as described above ~~worsens~~ worsens for balanced dataset.

$$\frac{\text{No. of correct predictions}}{\text{No. of data}}$$

GitHub Repository for Code:
https://github.com/Pavan-r-shetty/5644.git