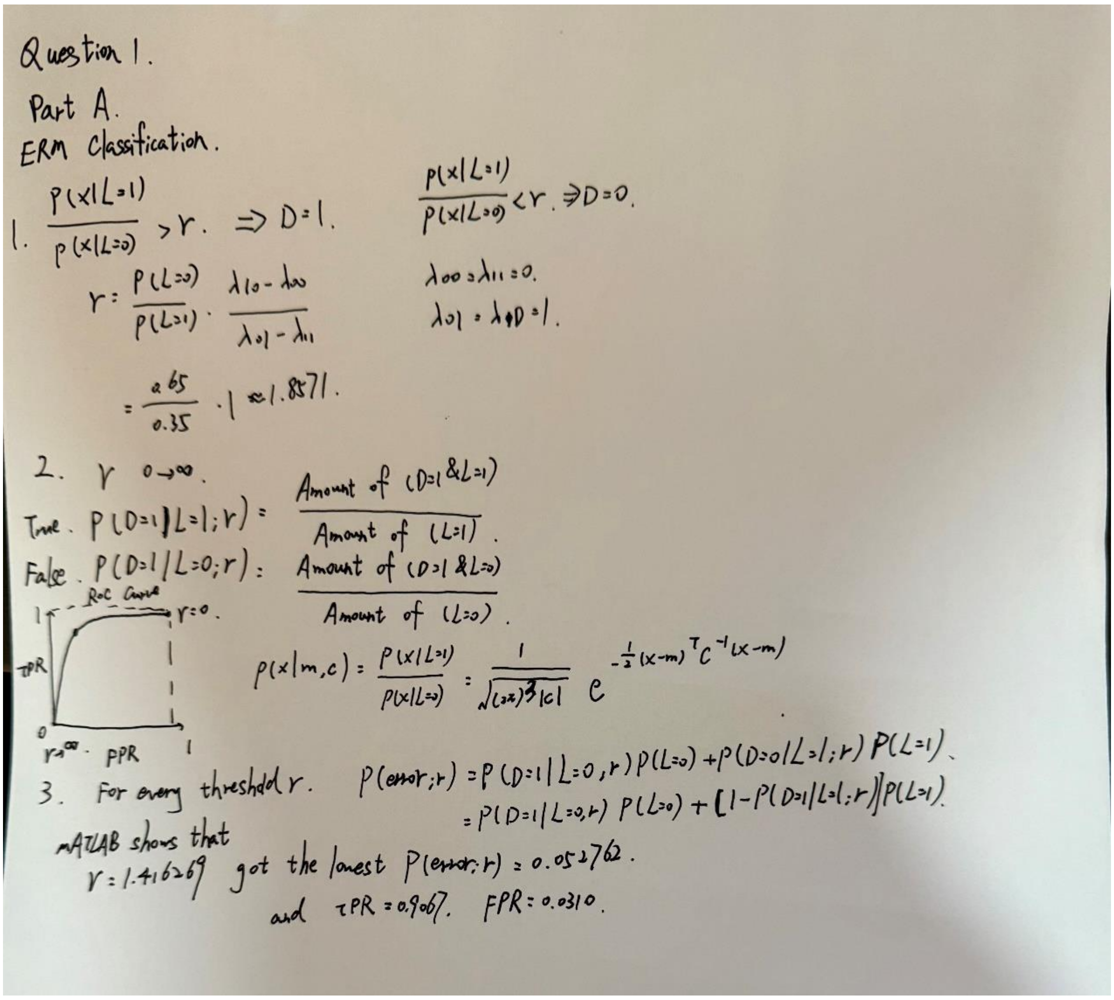
EECE 5644 Assignment 1 Jingcheng Wang

Repository: https://github.com/tomwang777/2025-Fall-EECE-5644-Machine-Learning/tree/main/Assignment%201

Question 1

Generated N0 = 6526, N1 = 3474

Part A



Question 1.

Part A.

ERM Classification.

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

$\gamma = \frac{P(L=0)}{P(L=1)} \cdot \frac{\lambda_{10}-\lambda_{00}}{\lambda_{01}-\lambda_{11}}$ $\lambda_{00} = \lambda_{11} = 0.$

$\lambda_{01} = \lambda_{10} = 1.$

$= \frac{0.65}{0.35} \cdot 1 \approx 1.8571.$

2. γ $0 \to \infty$.

True. $P(D=1|L=1;\gamma) = \frac{\text{Amount of } (D=1 \,\&\, L=1)}{\text{Amount of } (L=1)}.$

False. $P(D=1|L=0;\gamma) = \frac{\text{Amount of } (D=1 \,\&\, L=0)}{\text{Amount of } (L=0)}.$

$p(x|m,c) = \frac{p(x|L=1)}{p(x|L=0)} = \frac{1}{\sqrt{(2\pi)^3|C|}} e^{-\frac{1}{2}(x-m)^T C^{-1}(x-m)}$

3. For every threshold γ. $P(\text{error};\gamma) = P(D=1|L=0,\gamma)P(L=0) + P(D=0|L=1;\gamma)P(L=1).$

$= P(D=1|L=0,\gamma)P(L=0) + [1-P(D=1|L=1;\gamma)]P(L=1).$

MATLAB shows that

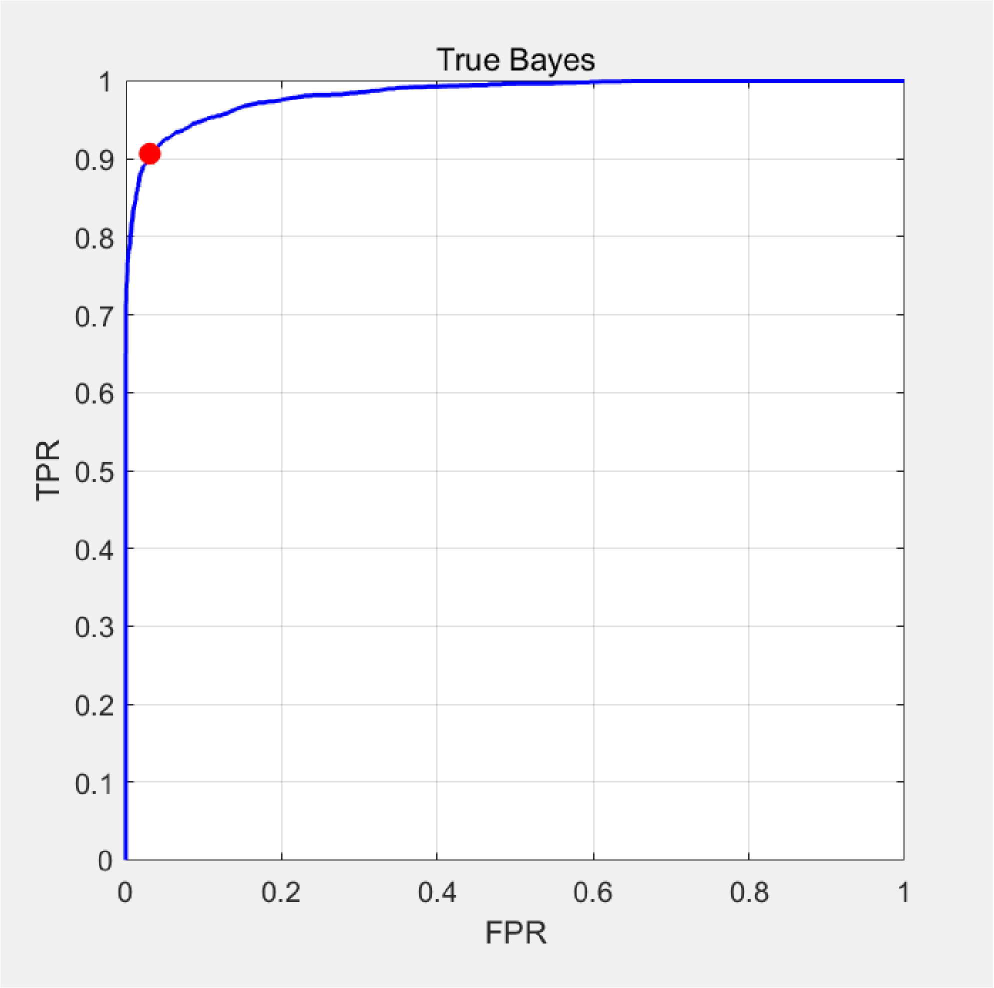
$\gamma = 1.416269$ got the lowest $P(\text{error};\gamma) = 0.052762.$

and TPR = 0.9067. FPR = 0.0310.

The empirical γ value I chose is much smaller than the theoretical optimal threshold, which means that the actual sample situation will deviate greatly from the theoretical result and requires a comprehensive analysis instead of calculating only one of them.

True Bayes:

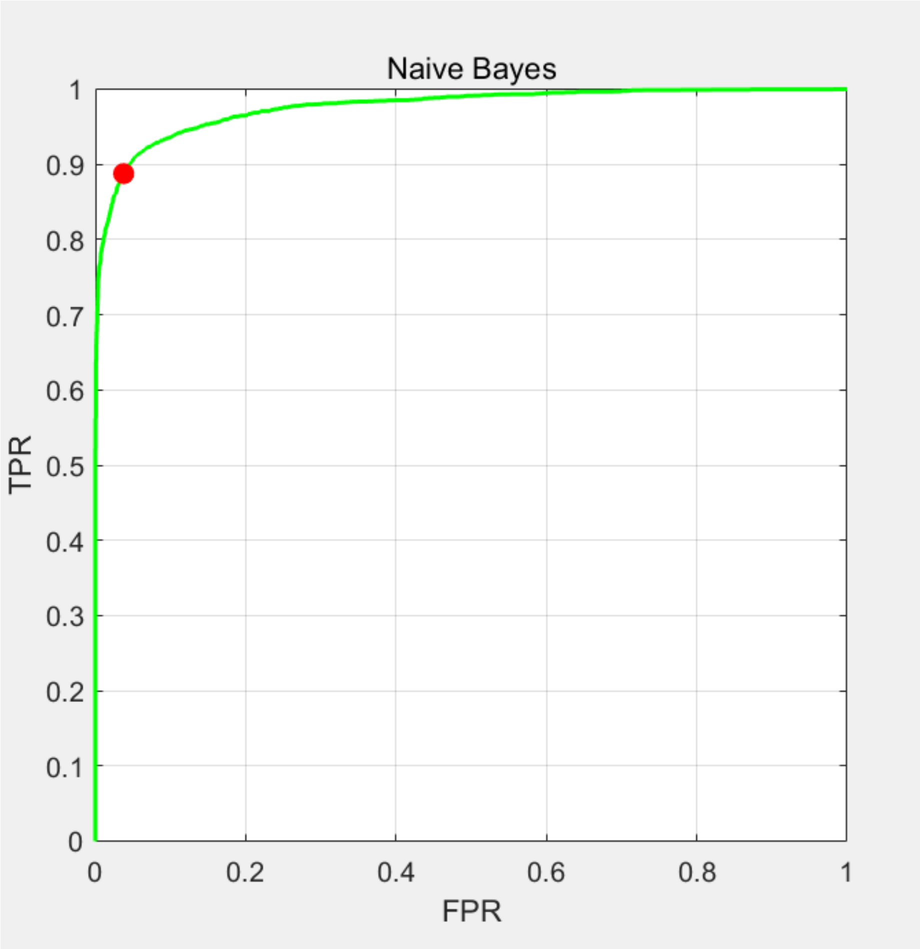
```
    min empirical error = 0.052762, threshold = 1.416269
    TPR = 0.9067, FPR = 0.0310
    theoretical gamma = 1.8571, empirical error at gamma = 0.053613
```



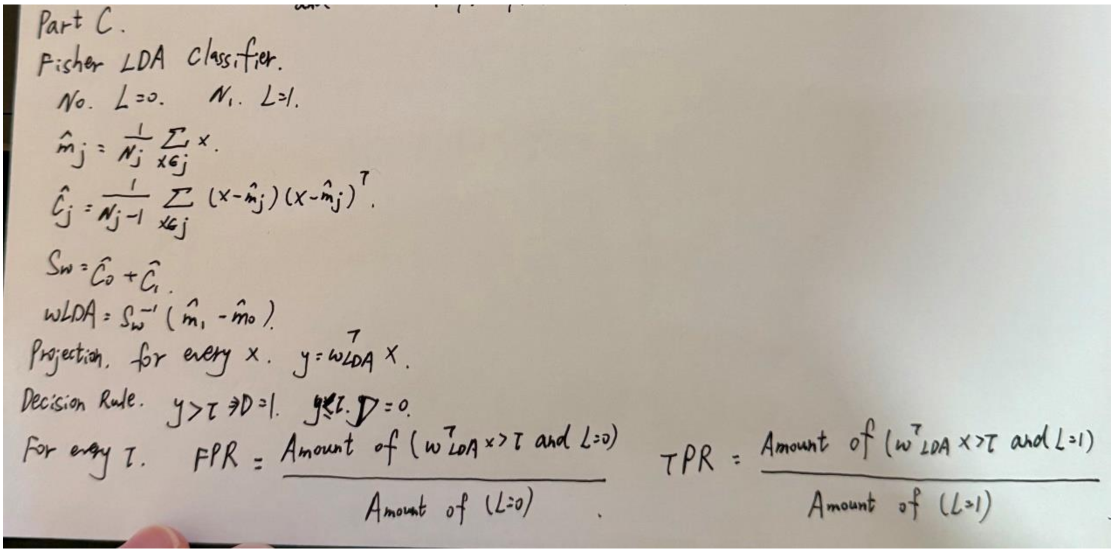
Part B

By incorrectly assuming that the covariance matrix is equal to the identity matrix and recalculating the likelihood ratio corresponding to each threshold, and drawing the ROC curve, we can find that the model mismatch leads to a significant decrease in the threshold results when designing the naive Bayes classifier. Only the values of C0 and C1 are changed, and the others remain unchanged. The minimum error probability of the model increases, which has a negative impact on the model performance. The reason is that the ROC curve moves downward, the minimum error probability becomes larger.

```
Naive Bayes (I):
    min empirical error = 0.063794, threshold = 0.914395
```



Part C



Part C.

Fisher LDA Classifier.

N_0. $L=0$. N_1. $L=1$.

$$\hat{m}_j = \frac{1}{N_j} \sum_{x \in j} x.$$

$$\hat{C}_j = \frac{1}{N_j - 1} \sum_{x \in j} (x-\hat{m}_j)(x-\hat{m}_j)^T.$$

$$S_w = \hat{C}_0 + \hat{C}_1.$$

$$w_{LDA} = S_w^{-1} (\hat{m}_1 - \hat{m}_0).$$

Projection. for every x. $y = w_{LDA}^T x$.

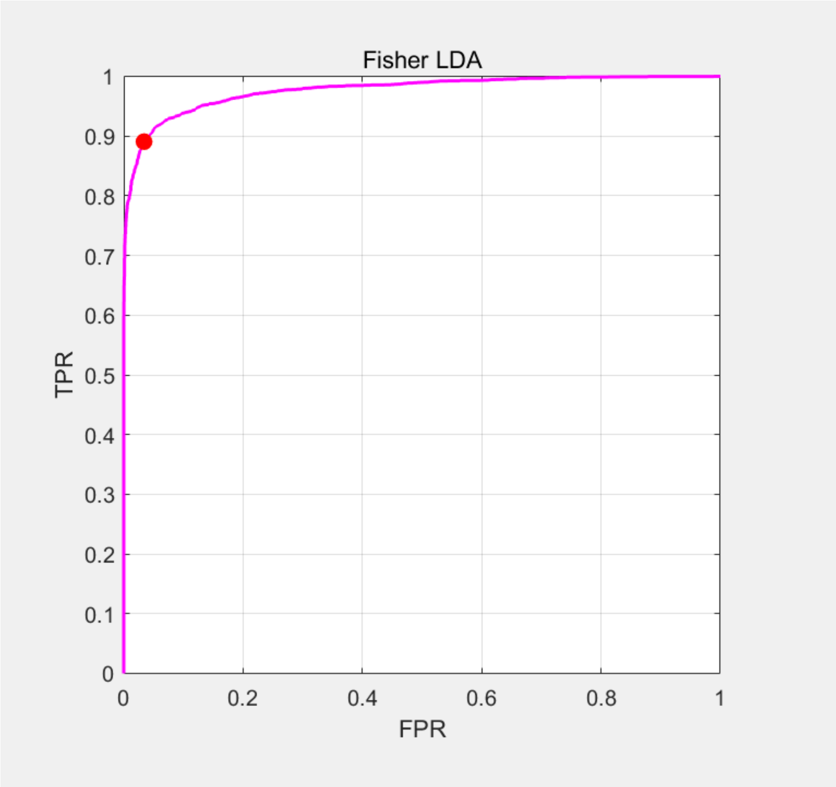
Decision Rule. $y > \tau \Rightarrow D=1$. $y \le \tau$. $D=0$.

For every τ. $FPR = \frac{\text{Amount of } (w_{LDA}^T x > \tau \text{ and } L=0)}{\text{Amount of } (L=0)}$. $TPR = \frac{\text{Amount of } (w_{LDA}^T x > \tau \text{ and } L=1)}{\text{Amount of } (L=1)}$

The performance of Fisher LDA is between True Bayes and Naive Bayes, mainly because it effectively utilizes the correlation information, but the projection causes some loss.

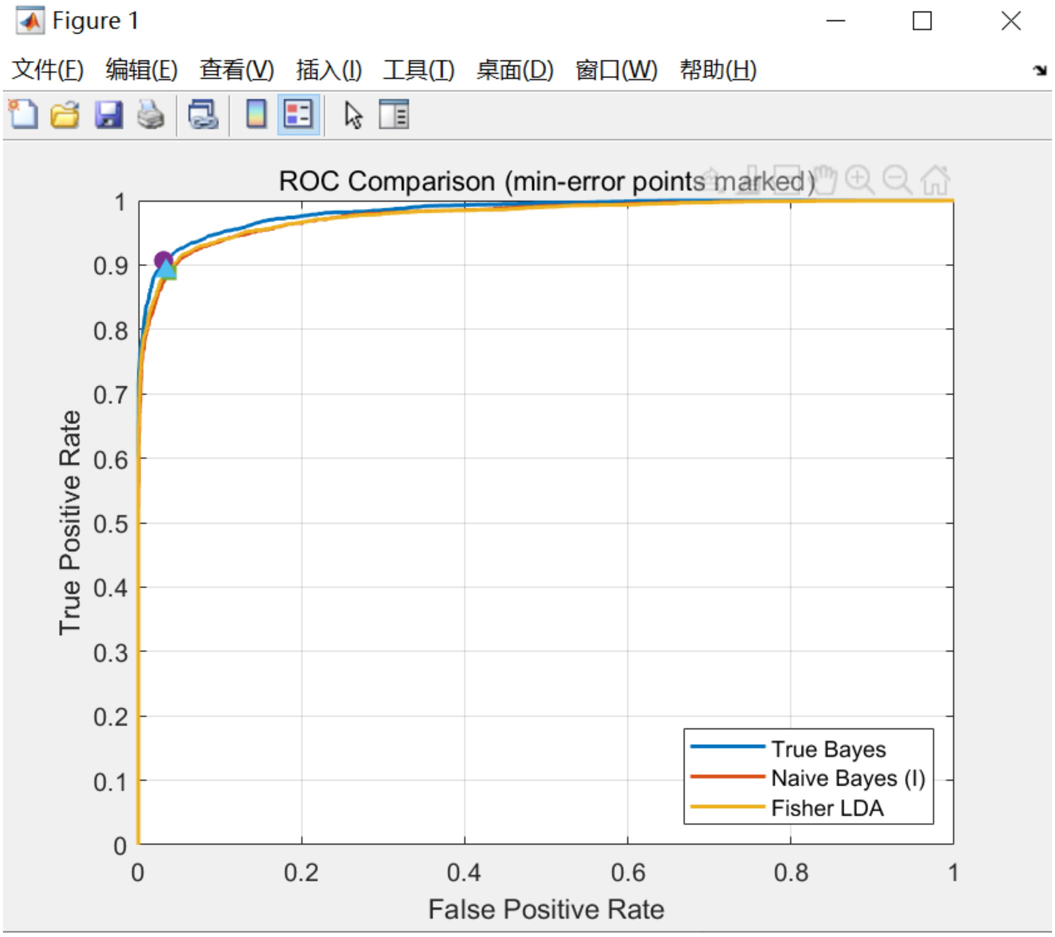
Fisher LDA:

min empirical error = 0.060396, threshold = 0.554445



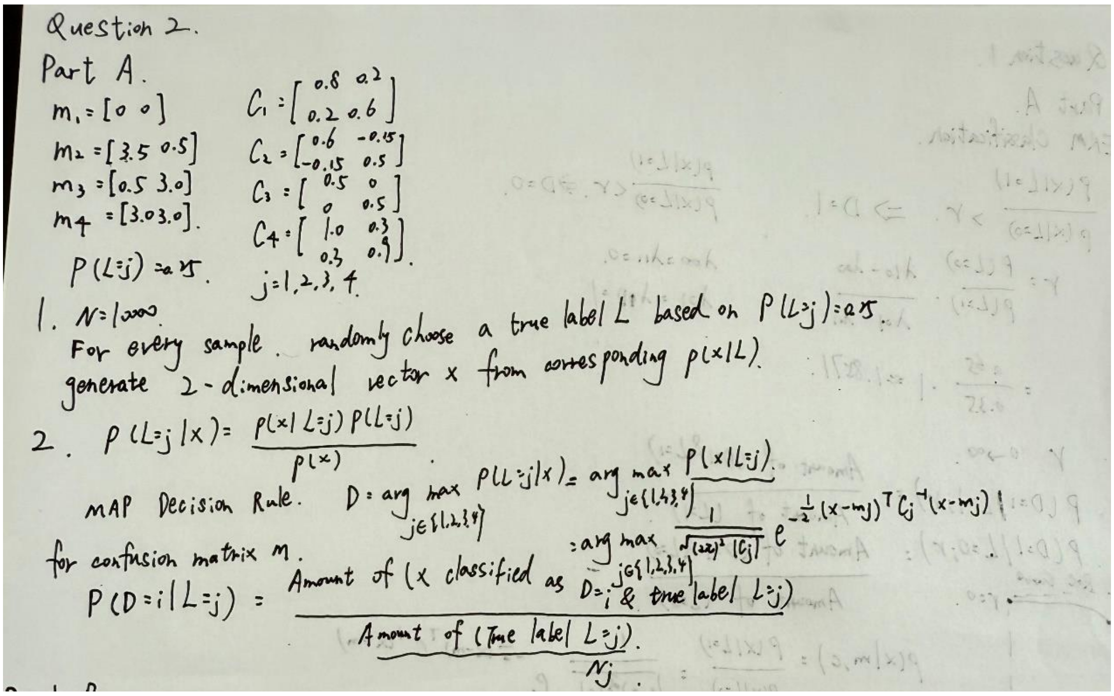
The red dot in each figure indicates the minimum error point.

Overall



Question 2

Part A



Question 2.

Part A.

$m_1 = [0\ 0]$ $\quad C_1 = \begin{bmatrix} 0.8 & 0.2 \\ 0.2 & 0.6 \end{bmatrix}$

$m_2 = [3.5\ 0.5]$ $\quad C_2 = \begin{bmatrix} 0.6 & -0.15 \\ -0.15 & 0.5 \end{bmatrix}$

$m_3 = [0.5\ 3.0]$ $\quad C_3 = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$

$m_4 = [3.0\ 3.0]$. $\quad C_4 = \begin{bmatrix} 1.0 & 0.3 \\ 0.3 & 0.9 \end{bmatrix}$.

$P(L=j) = 0.25$. $\quad j = 1, 2, 3, 4$.

1. $N = 10000$.
For every sample. randomly choose a true label L' based on $P(L=j) = 0.25$.
generate 2-dimensional vector x from corresponding $p(x|L)$.

2. $p(L=j|x) = \dfrac{p(x|L=j)\,p(L=j)}{p(x)}$.

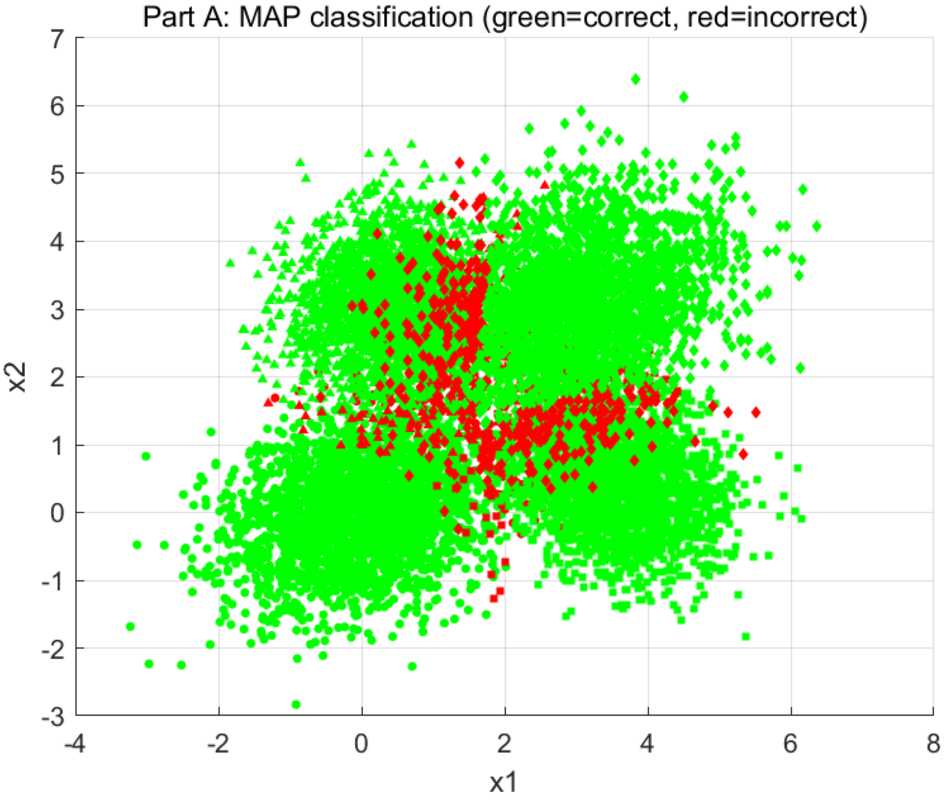
MAP Decision Rule. $D = \arg\max_{j\in\{1,2,3,4\}} p(L=j|x) = \arg\max_{j\in\{1,2,3,4\}} \underline{p(x|L=j)}$.

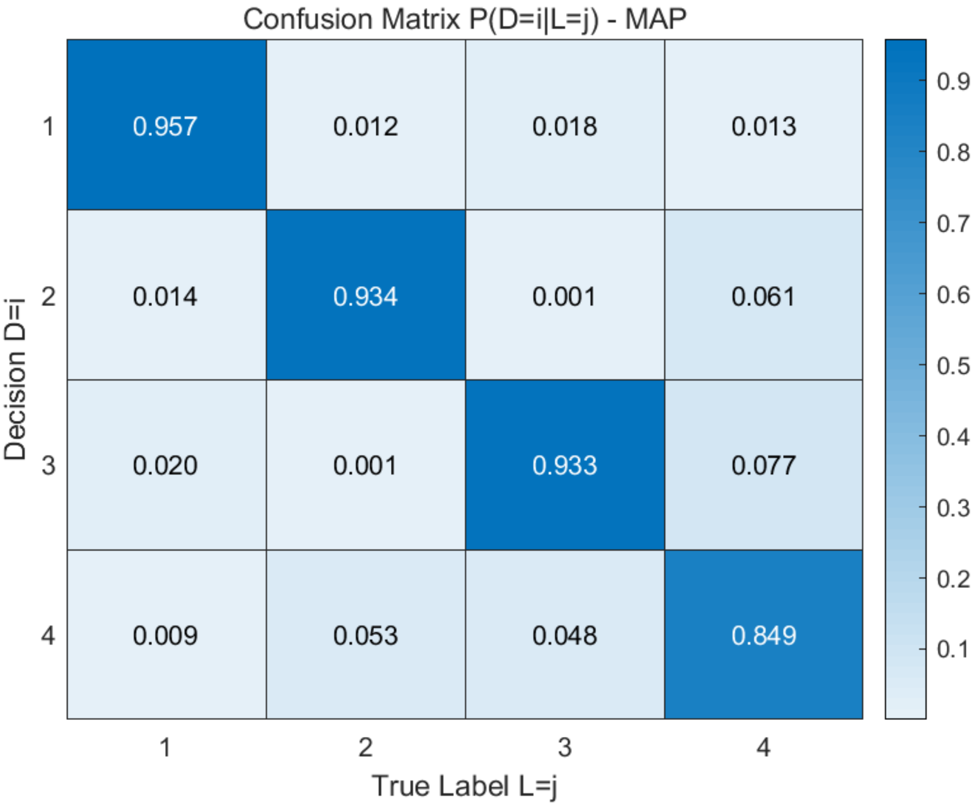
$= \arg\max_{j\in\{1,2,3,4\}} \dfrac{1}{\sqrt{(2\pi)^2\,|C_j|}}\, e^{-\frac{1}{2}(x-m_j)^T C_j^{-1}(x-m_j)}$

for confusion matrix m.

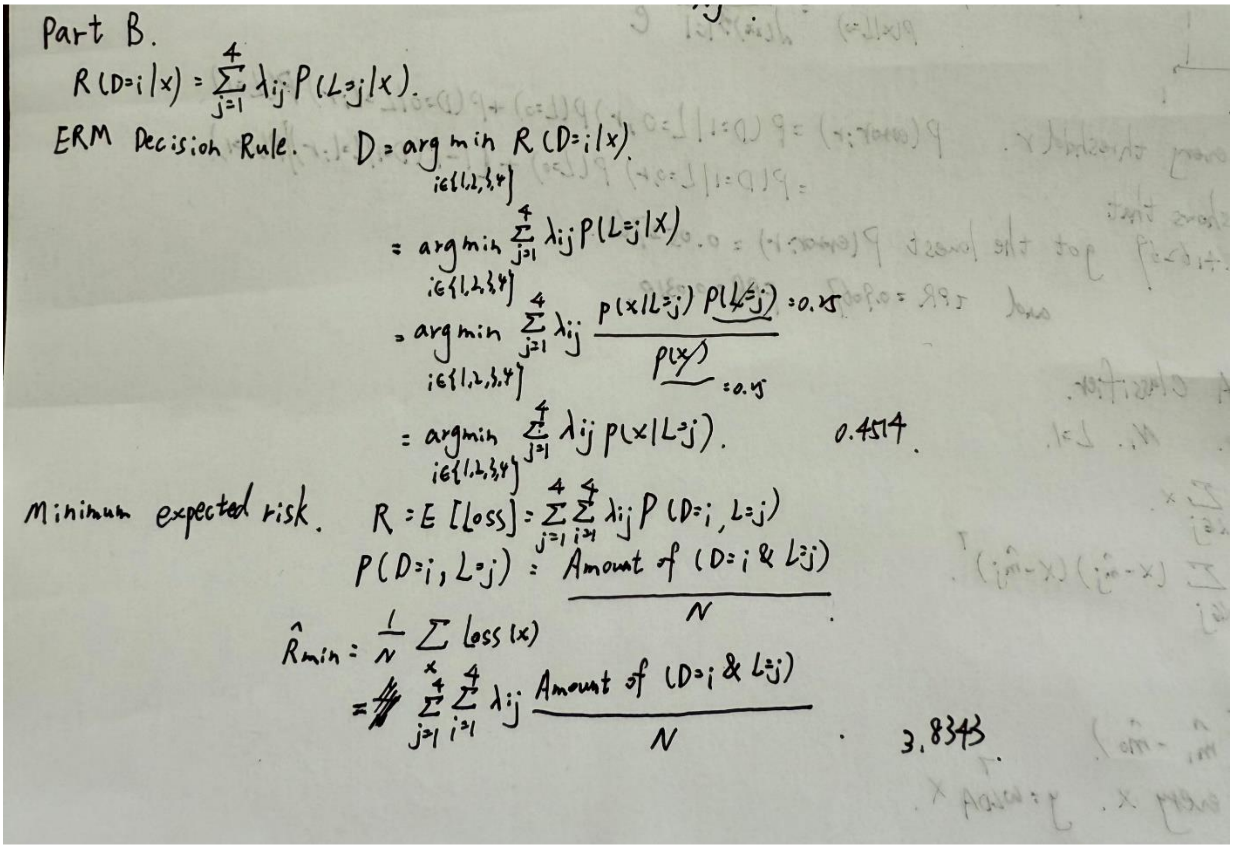
$P(D=i|L=j) = \dfrac{\text{Amount of } (x \text{ classified as } D=i \ \&\ \text{true label } L=j)}{\underbrace{\text{Amount of (True label } L=j)}_{N_j}}$.

Total samples: 10000 | incorrect: 813





Part B



Part B.

$$R(D=i|x) = \sum_{j=1}^{4} \lambda_{ij} P(L=j|x).$$

ERM Decision Rule. $D = \underset{i \in \{1,2,3,4\}}{\arg\min}\, R(D=i|x).$

$$= \underset{i \in \{1,2,3,4\}}{\arg\min} \sum_{j=1}^{4} \lambda_{ij} P(L=j|x)$$

$$= \underset{i \in \{1,2,3,4\}}{\arg\min} \sum_{j=1}^{4} \lambda_{ij} \frac{p(x|L=j)\, \underbrace{P(L=j)}_{=0.25}}{\underbrace{p(x)}_{=0.25}}$$

$$= \underset{i \in \{1,2,3,4\}}{\arg\min} \sum_{j=1}^{4} \lambda_{ij}\, p(x|L=j). \qquad 0.4514.$$

Minimum expected risk. $R = E[Loss] = \sum_{j=1}^{4} \sum_{i=1}^{4} \lambda_{ij} P(D=i, L=j)$

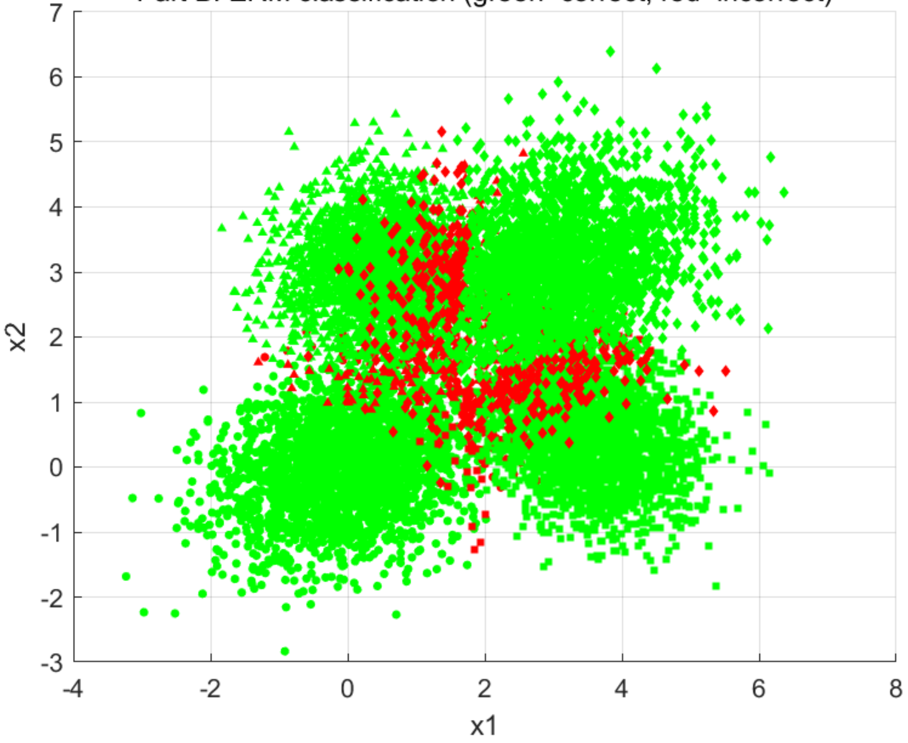
$$P(D=i, L=j) = \frac{\text{Amount of } (D=i \,\&\, L=j)}{N}.$$

$$\hat{R}_{min} = \frac{1}{N} \sum_{x} loss(x)$$

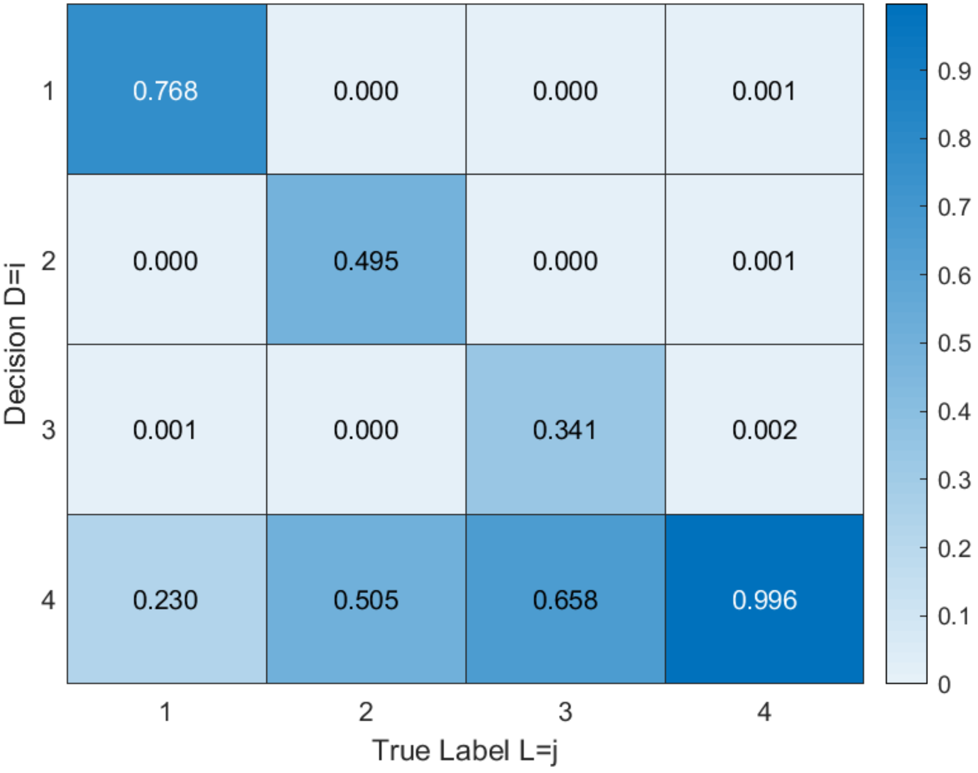
$$= \sum_{j=1}^{4} \sum_{i=1}^{4} \lambda_{ij} \frac{\text{Amount of } (D=i \,\&\, L=j)}{N}. \qquad 3.8343$$

```
Total samples: 10000 | incorrect: 813
MAP empirical risk (Lambda) = 3.8343
ERM empirical risk = 0.4514
```

Part B: ERM classification (green=correct, red=incorrect)

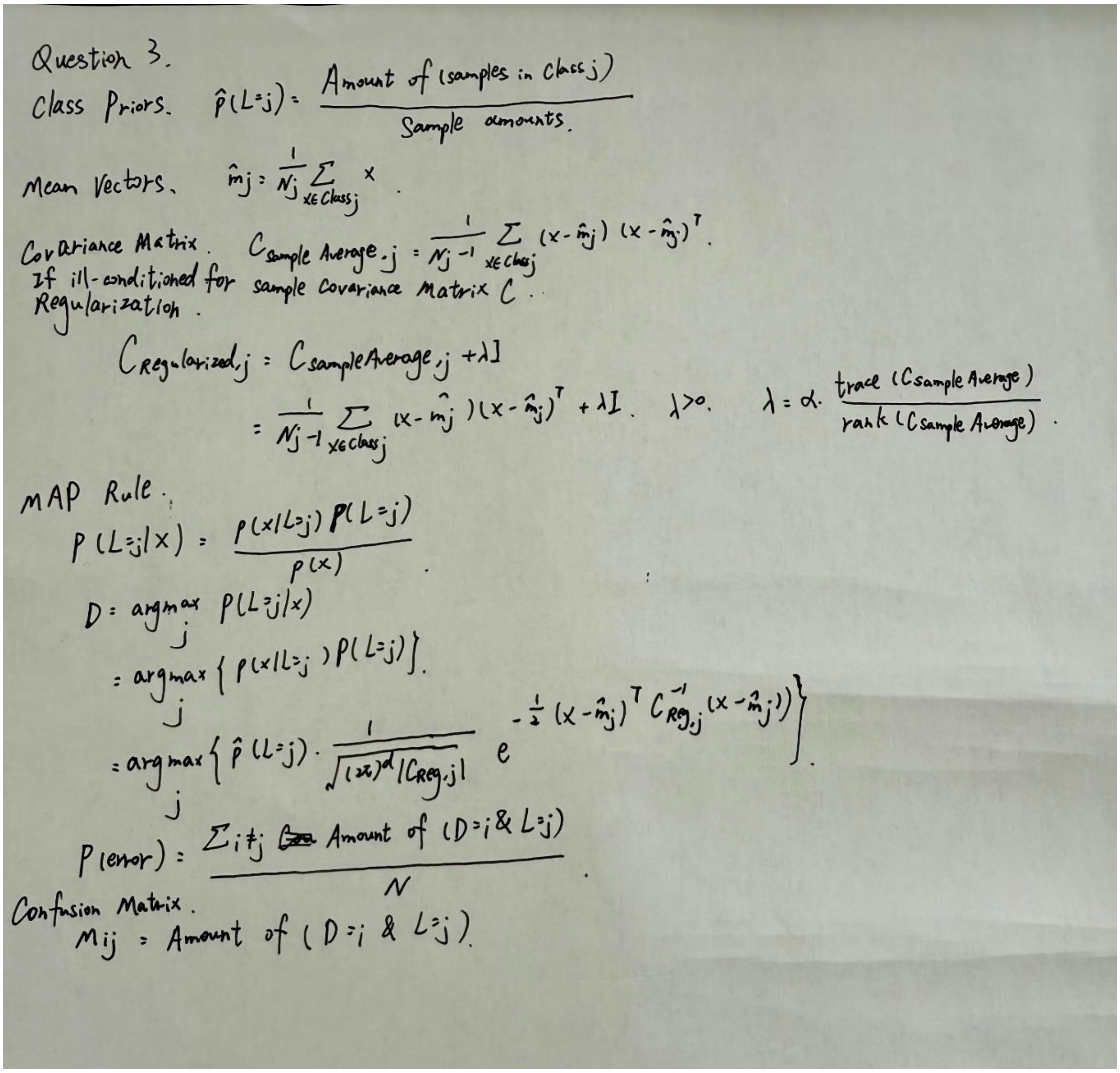


Confusion Matrix P(D=i|L=j) - ERM



	1	2	3	4
1	0.768	0.000	0.000	0.001
2	0.000	0.495	0.000	0.001
3	0.001	0.000	0.341	0.002
4	0.230	0.505	0.658	0.996

Question 3



Question 3.

Class Priors. $\hat{p}(L=j) = \frac{\text{Amount of (samples in Class j)}}{\text{Sample amounts}}$.

Mean Vectors. $\hat{m}_j = \frac{1}{N_j} \sum_{x \in Class_j} x$.

Covariance Matrix. $C_{\text{Sample Average}, j} = \frac{1}{N_j - 1} \sum_{x \in Class_j} (x - \hat{m}_j)(x - \hat{m}_j)^T$.

If ill-conditioned for sample Covariance Matrix C.

Regularization.

$$C_{\text{Regularized}, j} = C_{\text{sampleAverage}, j} + \lambda]$$

$$= \frac{1}{N_j - 1} \sum_{x \in Class_j} (x - \hat{m}_j)(x - \hat{m}_j)^T + \lambda I. \quad \lambda > 0. \quad \lambda = \alpha \cdot \frac{\text{trace}(C_{\text{sample Average}})}{\text{rank}(C_{\text{sample Average}})}.$$

MAP Rule.

$$P(L=j|x) = \frac{p(x|L=j)\,P(L=j)}{p(x)}.$$

$$D = \underset{j}{\text{argmax}}\; P(L=j|x)$$

$$= \underset{j}{\text{argmax}} \{ p(x|L=j) P(L=j) \}.$$

$$= \underset{j}{\text{argmax}} \left\{ \hat{p}(L=j) \cdot \frac{1}{\sqrt{(2\pi)^d |C_{Reg,j}|}}\, e^{-\frac{1}{2}(x - \hat{m}_j)^T C_{Reg,j}^{-1} (x - \hat{m}_j)} \right\}.$$

$$P(\text{error}) = \frac{\sum_{i \neq j} \text{Amount of } (D=i \,\&\, L=j)}{N}.$$

Confusion Matrix.

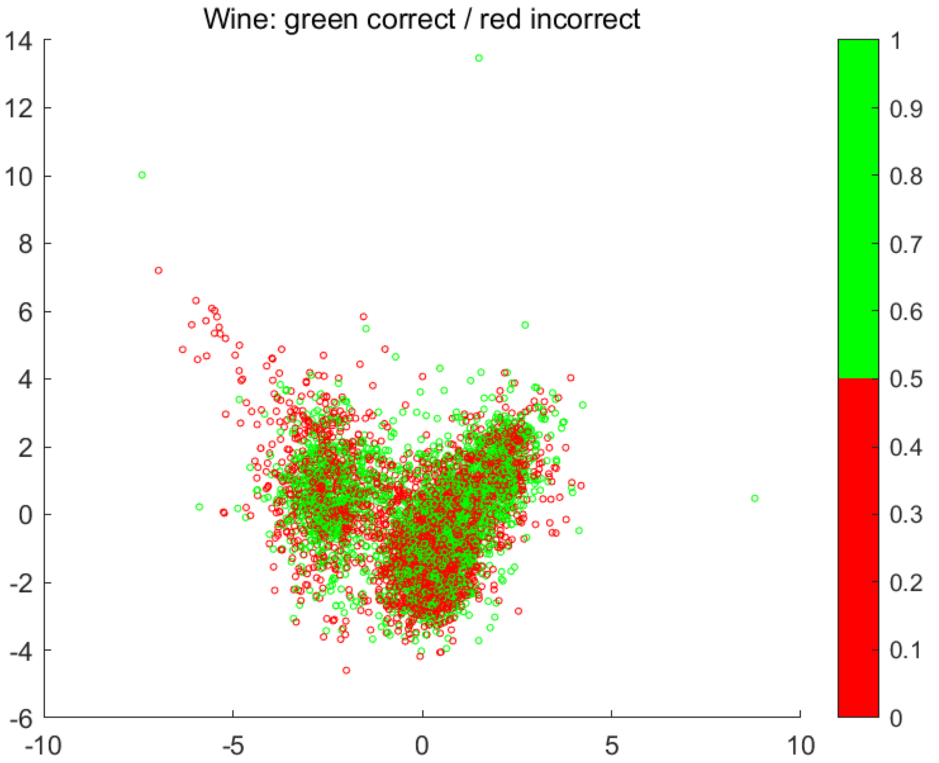
M_{ij} = Amount of $(D=i \,\&\, L=j)$.

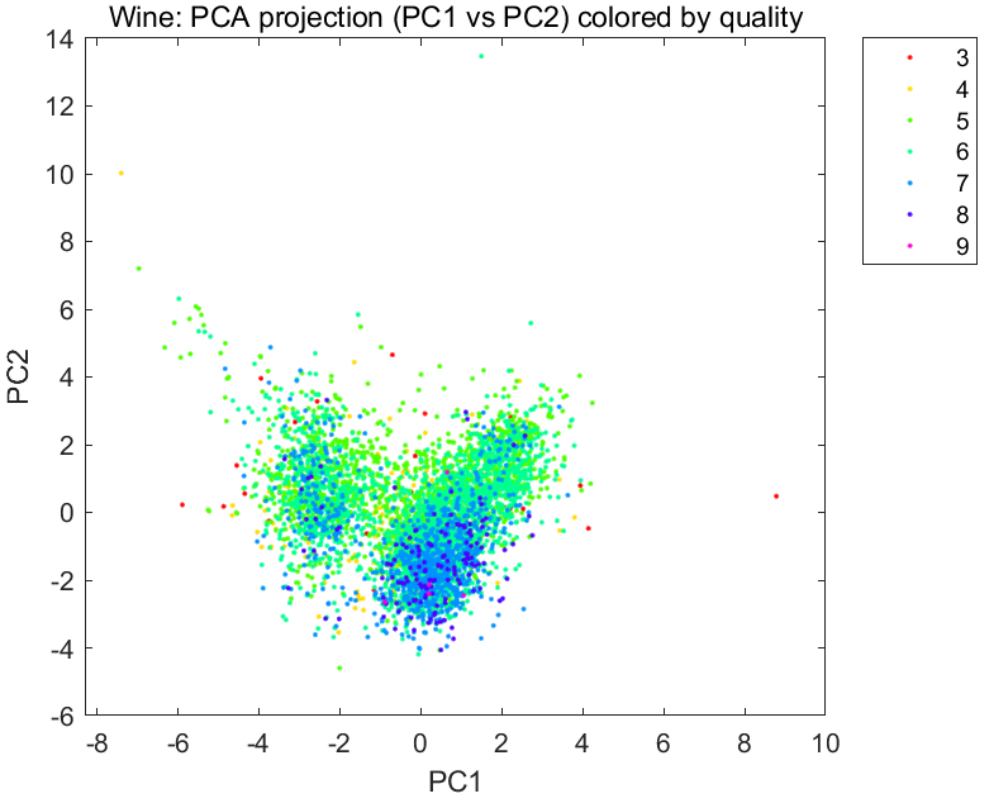
Wine dataset (all data) results: N = 6497

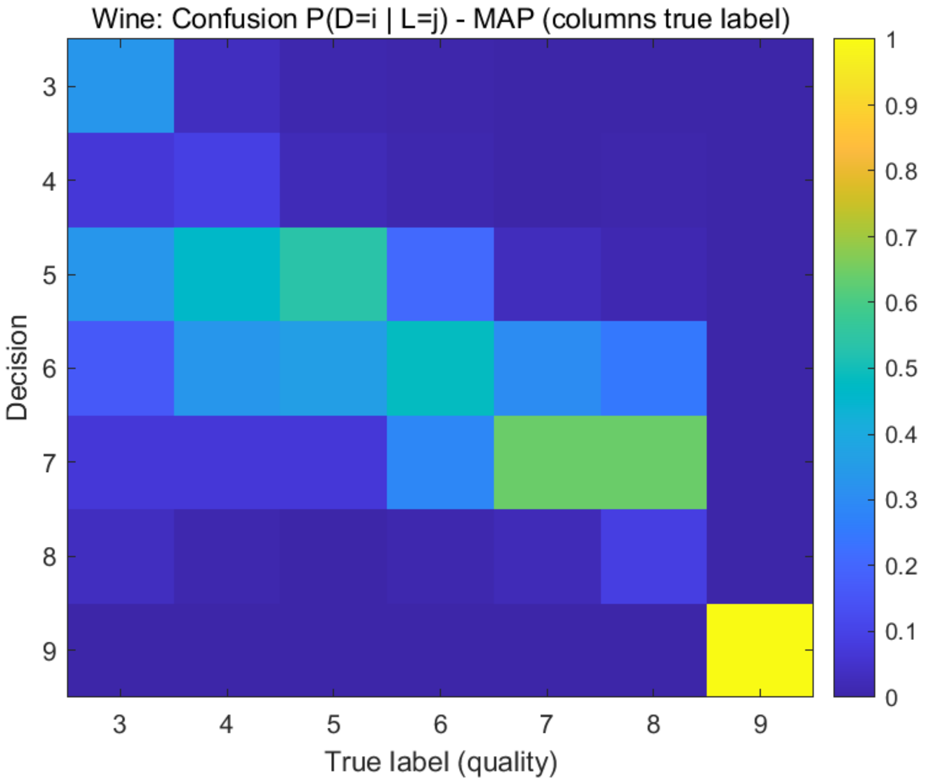
d = 11

classes = 7

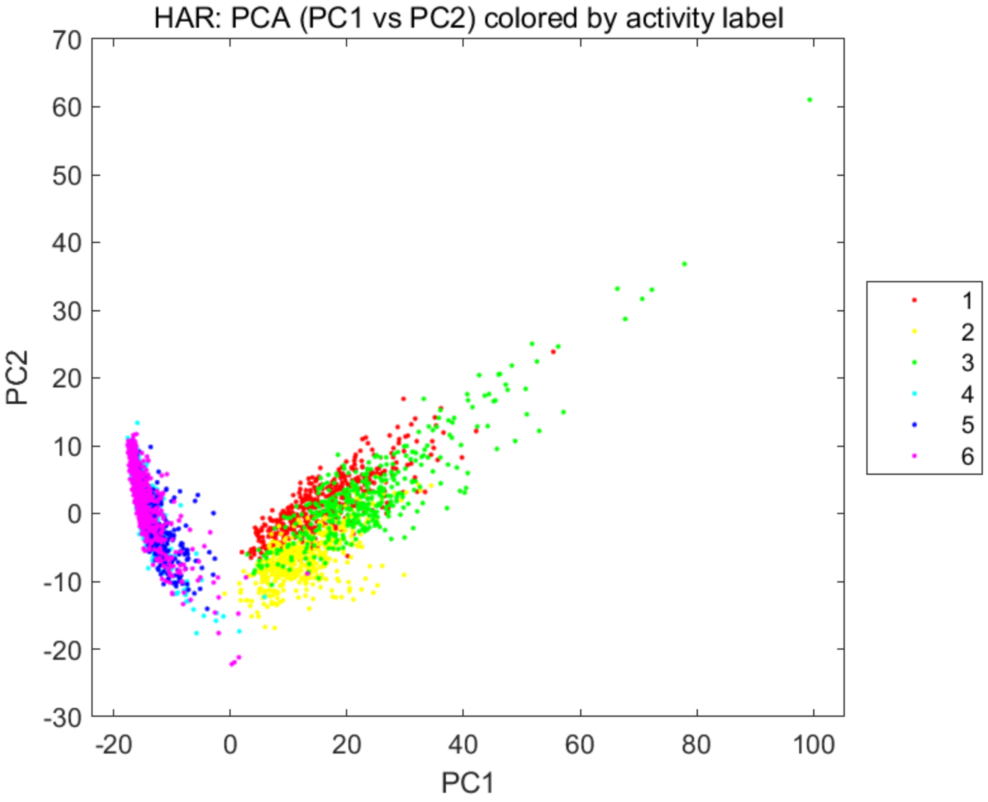
Training MAP error = 0.4982

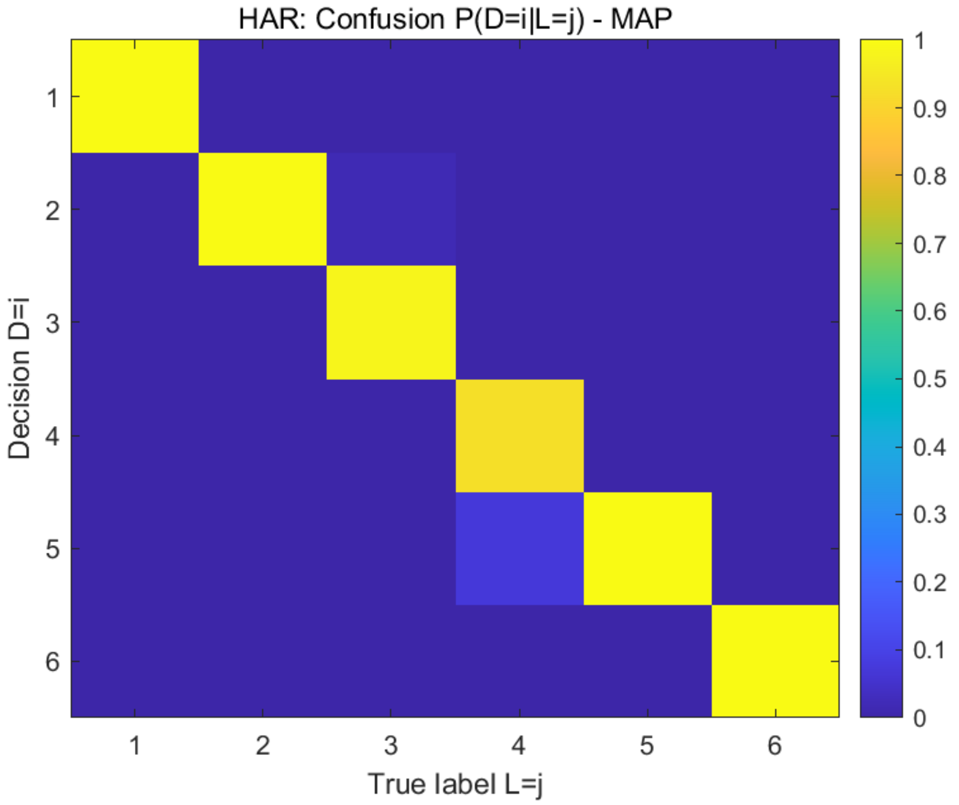

Wine: green correct / red incorrect
14
12
10
8
6
4
2
0
-2
-4
-6
-10
-5
0
5
10
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0


Wine: PCA projection (PC1 vs PC2) colored by quality
PC2
PC1
14
12
10
8
6
4
2
0
-2
-4
-6
-8
-6
-4
-2
0
2
4
6
8
10
3
4
5
6
7
8
9



```
HAR dataset results: N = 10299
d = 561
classes = 6
Training MAP error = 0.0149
```





The Gaussian class conditional model may not be appropriate for these datasets, as the PCA projections show significant aliasing, suggesting an oversimplified assumption of a Gaussian distribution for the same class. The Gaussian model choice may be too simplistic for this dataset, leading to model bias, higher training errors, and increased confusion. GMMs and mixture models may be needed to improve training quality and discriminative performance.

Citation

1. Course recording
2. Course notes
3. Course codes provided on Canvas
4. Discussion with classmates
5. Generative AI models