

Homework 1

Github: github.com/

Problem 1

Number of samples from Class 1: 6476, Class 2: 3524

Part A

Gamma MAP (Theoretical): 1.8571428571428574
Probability of Error: 0.0591
Best Gamma (ERM): 1.6561176747588073
Probability of Error(Empirical): 0.05789999999999999

Part B

Gamma MAP (Theoretical): 1.8571428571428574
Probability of Error: 0.0885
Best Gamma (Based on Data): 2.2632109083141407
Probability of Error(Empirical): 0.08779999999999999

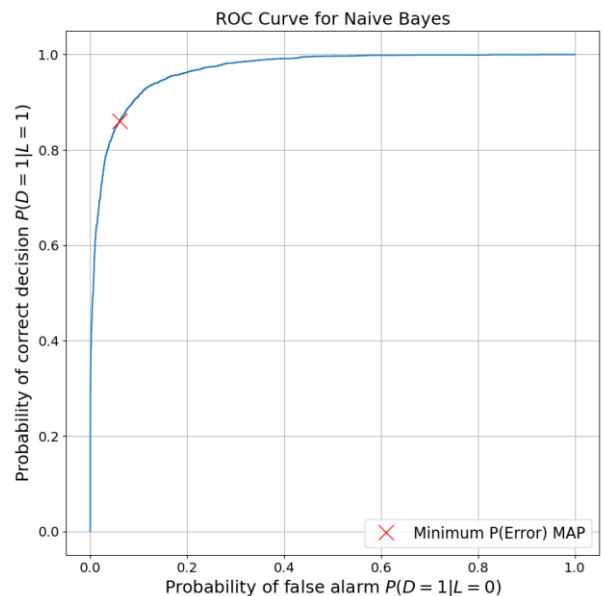
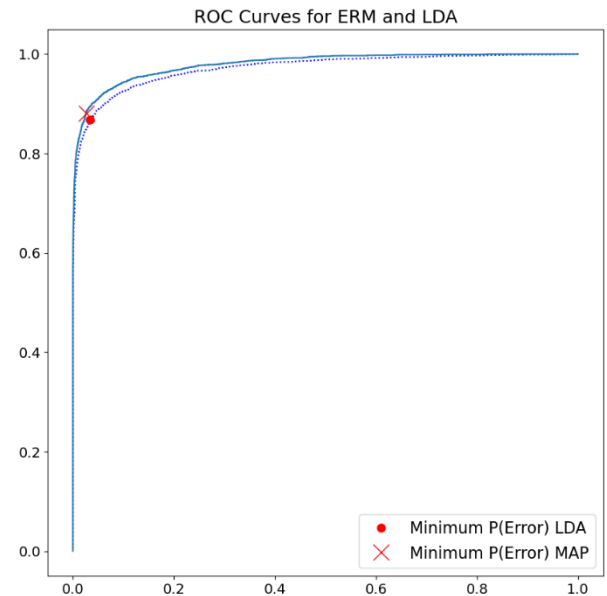
The probability of error using the MAP value for gamma is worse. This is because our sigma values are inaccurate.

Part C

Gamma MAP (Theoretical): 1.8571428571428574
Best Gamma (ERM): 1.6561176747588073
Best Gamma (LDA): 1.5344155791771428
Smallest P(error) for ERM: 0.05789999999999999
Smallest P(error) for LDA: 0.068600000000000001

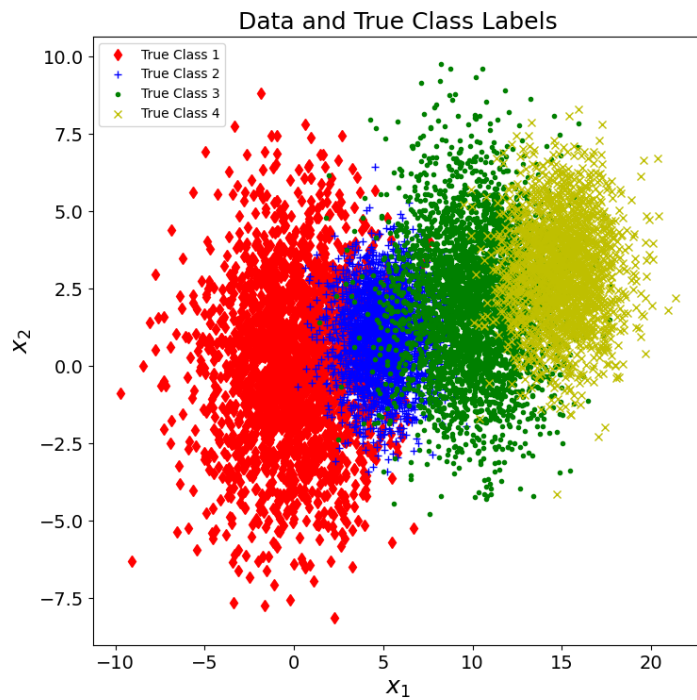
The gamma value for ERM is closer to the theoretical value than LDA. Also, the probability of error for the best empirically calculated gamma value is better when using ERM than LDA.

Since $P(\text{error}|\text{LDA}) > P(\text{error}|\text{ERM})$, we can say that ERM performs better than LDA.



Problem 2

Part A



Empirically Estimated Probability of Error: 0.1298

Confusion Matrix (rows: Predicted class, columns: True class):

[[1786, 100, 9, 0]

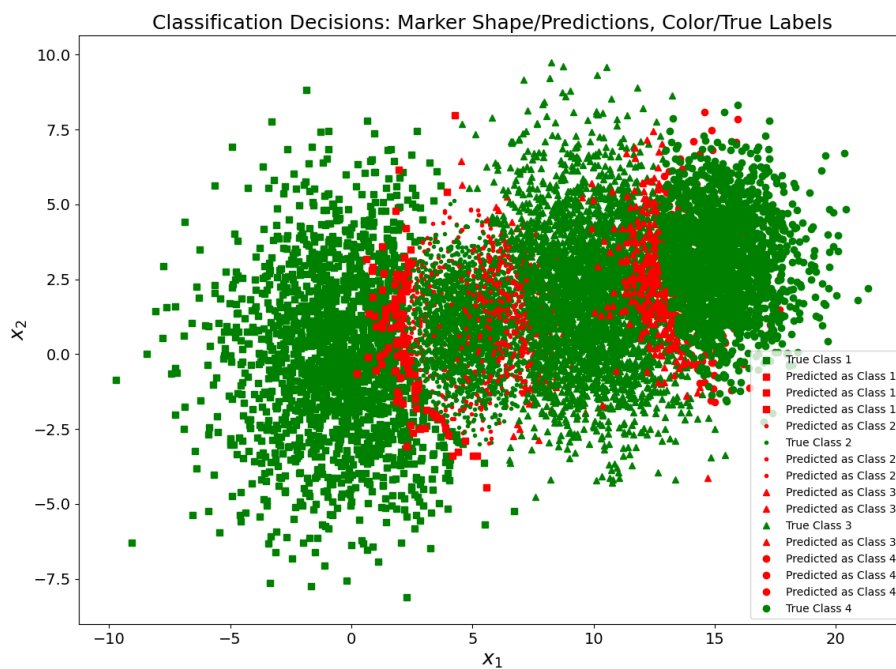
[240, 2263, 261, 0]

[5, 181, 2894, 256]

[0, 0, 246, 1759]]

Total Number of Misclassified Samples:
1298

Empirically Estimated Probability of Error: 0.1298



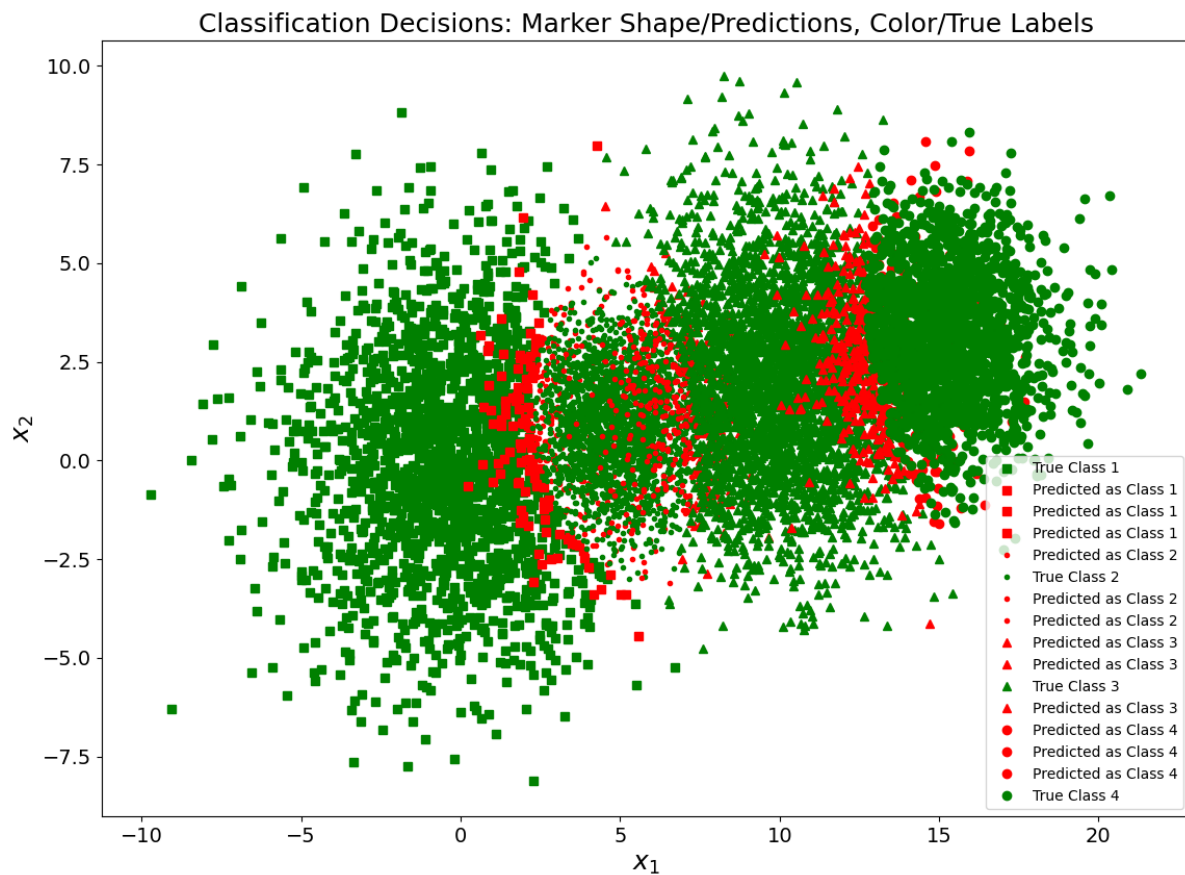
Part B

Confusion Matrix (rows: Predicted class, columns: True class):

```
[[1784, 100, 8, 0]
 [243, 2269, 265, 0]
 [4, 175, 2891, 256]
 [0, 0, 246, 1759]]
```

Total Number of Misclassified Samples: 1297

Empirically Estimated Probability of Error: 0.1297



The two results from part A and B look very similar. In fact, they have about the same number of misclassified samples and probability of error. There is not much difference between the two despite the new loss matrix.

Problem 3

WINE

Number of samples from Quality Rating 0: 0

..... Quality Rating 1: 0

..... Quality Rating 2: 0

..... Quality Rating 3: 20

..... Quality Rating 4: 159

..... Quality Rating 5: 1511

..... Quality Rating 6: 2167

..... Quality Rating 7: 865

..... Quality Rating 8: 172

..... Quality Rating 9: 4

..... Quality Rating 10: 0

Confusion Matrix (rows: Predicted class, columns:
True class):

```
[[ 4  2  1  0  0  0  0]
```

```
[ 4 23 11 0 19  6  0]
```

```
[ 4 39 699 262 191  8  0]
```

```
[ 0  7 560 1851 107  1  0]
```

```
[ 4 69 229 54 525 60  0]
```

```
[ 3 18 11  0 23 97  0]
```

```
[ 1  1  0  0  0  0  4]]
```

Total Number of Misclassified Samples: 1695

Empirically Estimated Probability of Error: 0.3461

