EECE 5644 Homework 1

Casey Holden

October 20th, 2023

View my GitHub page for my source code:

https://github.com/holdenc20/EECE5644-Homework1

1.

To start, I generated the 4-dimensional sample data using the multivariate Gaussian probability density functions. The two Gaussians are labels $L \in \{0,1\}$ where $P(L=0) = 0.35$ and $P(L=0) = 0.65$. Of the 10,000 samples generated, 3,500 of the samples correspond to the Gaussian $g(x|m_0, C_0)$ were:

$$m_0 = \begin{matrix} -1 \\ -1 \\ -1 \\ -1 \end{matrix}, \quad C_0 = \begin{matrix} 2 & -0.5 & 0.3 & 0 \\ -0.5 & 1 & -0.5 & 0 \\ 0.3 & -0.5 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{matrix}$$

And 6,500 of the samples correspond to $g(x|m_1, C_1)$ were:

$$m_1 = \begin{matrix} 1 \\ 1 \\ 1 \\ 1 \end{matrix}, \quad C_1 = \begin{matrix} 1 & 0.3 & -0.2 & 0 \\ 0.3 & 2 & 0.3 & 0 \\ -0.2 & 0.3 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{matrix}$$

1.A ERM Classification Using the Knowledge of True Data PDF

1.A.1 Minimum Expected Risk Classification Rule

The classification rule we will be using is a likelihood ratio test.

$$\frac{p(x|L=1)}{p(x|L=0)} = \frac{g(x|m_0, C_0)}{g(x|m_1, C_1)} > \gamma = \frac{p(L=0)}{p(L=1)} * \frac{\lambda_{01} - \lambda_{00}}{\lambda_{10} - \lambda_{11}}$$

For loss matrix where D is the decision label, the loss is 1 when $L \neq D$ and 0 when $L = D$.

Loss Matrix Values λ_{ij} where $D = i \mid L = j$	
$\lambda_{00} = 1$	True Negative
$\lambda_{01} = 0$	False Negative
$\lambda_{10} = 0$	False Positive
$\lambda_{11} = 1$	True Positive

Figure 1. The Loss Matrix Costs

From Figure 1, we can see that γ should be as follows:

$$\gamma = \frac{p(L=0)}{p(L=1)} * \frac{\lambda_{01} - \lambda_{00}}{\lambda_{10} - \lambda_{11}} = \frac{0.35}{0.65} * \frac{1 - 0}{1 - 0} = 0.538$$

1.A.2 Implementation of the ERM Classifier

Using the generated samples, we ran the classifier on varying threshold (γ) values. For each of these γ values we calculated the expected predictions for each of the samples. From this set of decisions, we were able to compare them with the label to determine the true positve, false positive, false negative, true negative rates. Using the false positive and false negative rates, we were able to calculate the error of each of the sets of predictions. From the probabilities of false positives and the probabilities of true positives, we were able to generate the following ROC curve.

Commented [CH1]: Add chart

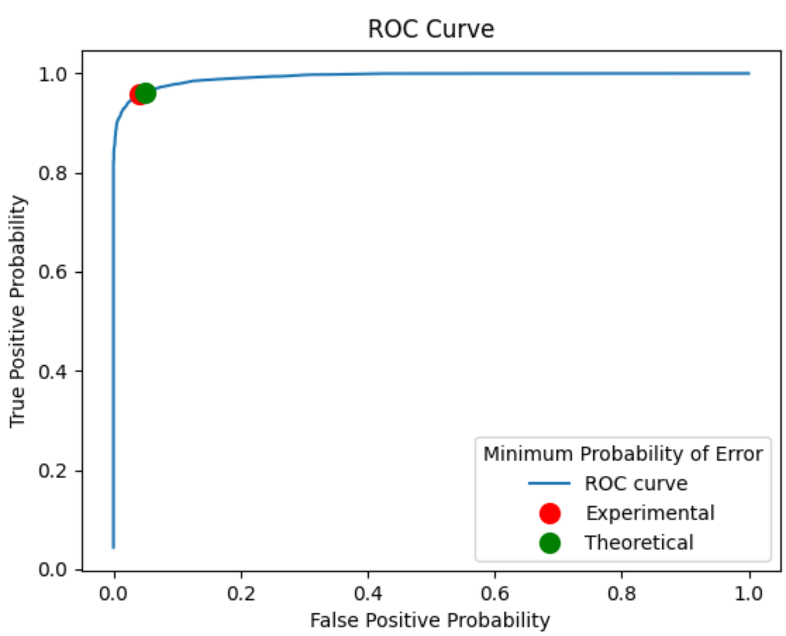

Figure 2. ROC Curve of ERM Classification

1.A.3 Threshold with Minimum Probability of Error

From the errors, calculated using the following equation:

$$P(error;\gamma) = P(D=1|L=0;\gamma)P(L=0) + P(D=0|L=1;\gamma)P(L=1)$$

We were able to find the γ with the minimum error to be

$$\gamma_{experimental} = 0.637$$

$$P(error; \gamma_{experimental}) = 0.045$$

Compared to the $\gamma_{theoretical} = 0.539$, the thresholds are pretty similar considering the wide range of threshold values considered (e^{-30}, e^{30}). For this problem, I varied exponential term form -30 to 30 for 1000 thresholds so there was a lot of values in the (0,1) range.

1.B ERM Classification with Incorrect Knowledge of the Data Distribution

For this problem, I used the same sample data from the $g(x|m_0, C_0)$ and $g(x|m_1, C_1)$ distributions. But when classifying and the data, we used the diagonal matrix of C_0 and C_1.

1.B.1 Minimum Expected Risk Classification Rule

The classification rule for this question is the same as part 1.A.1 results.

1.B.2 Implementation of the ERM Classifier

We used the exact same implementation as the previous section but when classifying the data we used the incorrect data distribution. Using that we generated the following ROC curve.

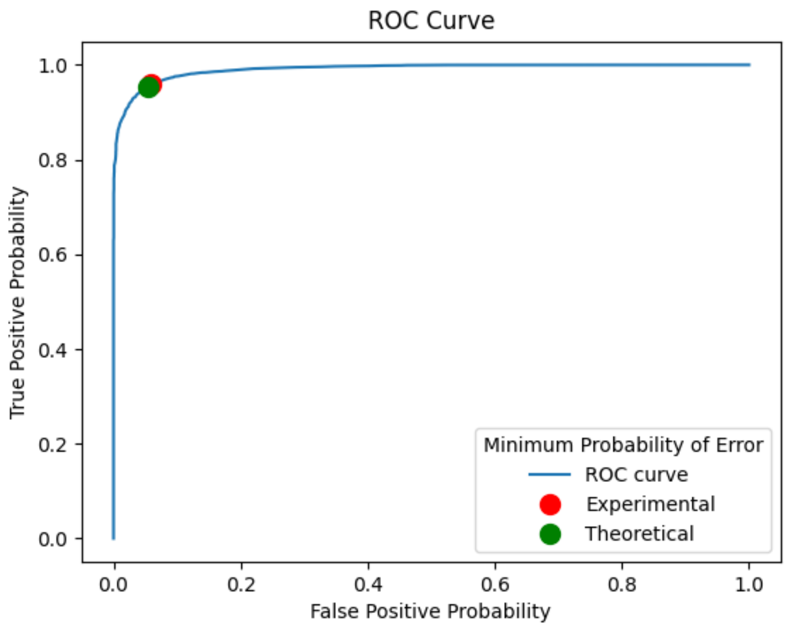

Figure 3. ROC Curve of ERM Classification with Incorrect Knowledge

1.B.3 Threshold with Minimum Probability of Error

Using the same process as 1.A.3, we were able to find the γ with the minimum error to be

$$\gamma_{experimental} \ = \ 0.472$$

$$P(error; \gamma_{experimental}) \ = \ 0.047$$

Although the $\gamma_{experimental}$ is not that similar to the $\gamma_{experimental}$ from part A, the error is almost exactly the same which is surprising. So, a minimal change did occur that increased the minimum probability of error. Although the ROC curve looks almost the same, because of the minimal change in the minimum probability of error, we can assume that there was also a minimal change in the ROC curve.

1.C Fisher Linear Discriminant Analysis Based Classifier

By calculating the mean and covariance of the sample data for both class labels, we can determine the w_{LDA} value for the LDA classifier. To determine the mean and covarience of the sample data, we estimated using the built in sample mean and sample covariance functions. From these estimated mean and covariance values for each of the labels, we can determine the w_{LDA} value.

$$Within\ Class\ Scatter\ Matrix\ =\ SW\ =\ 0.5\ *\ C_{sample,0} + 0.5 * C_{sample,1}$$

$$m_{avg} = 0.5\ *\ m_{sample,0} + 0.5 * m_{sample,1}$$

$$Between\ Class\ Scatter\ Matrix = SB\ =\ 0.5\ *\ ||m_{sample,0}\ -\ m_{avg}||^2 + 0.5 * ||m_{sample,1}\ -\ m_{avg}||^2$$

From that we can get the eigenvalues and eigenvectors of $SW^{-1} \cdot SB$

w_{LDA} is the normalized eigenvector of the maximum eigenvalue.

Using wLDA, we can classify the sample data with $\tau \in (-inf, inf)$.

$$x \cdot w_{LDA} < \tau$$

Using the LDA classifier, we generated the following ROC curve.

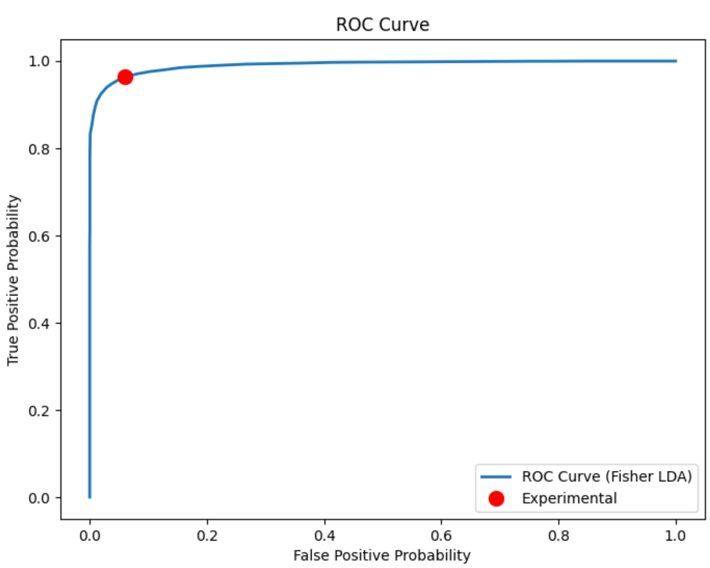

Figure 4. ROC Curve of LDA Classification

The minimum error for the LDA Classification was 0.045 for a threshold of 0.424. This minimum error is less than the minimum error of the ERM Classification with Incorrect Knowledge but more than the minimum error of the ERM Classification.

2

For this question, we are going to be creating a sample 3-dimensional vectors that are generated by 4 Gaussians. The first Gaussian with p(L=1) = 0.3 will have a mean, and covariance as follows:

$$m_1 = \begin{matrix} -30 \\ -12 \\ -10 \end{matrix}, \quad C_1 = \begin{matrix} 25 & -5 & 3 \\ -5 & 10 & -5 \\ 3 & -5 & 10 \end{matrix}$$

The second Gaussian with p(L=2) = 0.3:

$$m_2 = \begin{matrix} 12 \\ 14 \\ 24 \end{matrix}, \quad C_2 = \begin{matrix} 15 & -5 & 0 \\ -5 & 13 & -5 \\ 3 & -5 & 10 \end{matrix}$$

The third and fourth Gaussians with p(L=3) = 0.4 with equal weights:

$$m_{3a} = \begin{matrix} 45 \\ -34 \\ 7 \end{matrix}, \quad C_{3a} = \begin{matrix} 10 & 3 & -2 \\ 3 & 20 & 3 \\ -2 & 3 & 20 \end{matrix}$$

$$m_{3b} = \begin{matrix} 1 \\ 35 \\ -35 \end{matrix}, \quad C_{3b} = \begin{matrix} 10 & 3 & -2 \\ 3 & 20 & 3 \\ -2 & 3 & 20 \end{matrix}$$

I scaled up the covariance matrices so that there was a significant amount of overlap.

2.A Minimum Probability of Error Classification with MAP classifier

The MAP classifier used a loss matrix of:

$$\Lambda = \begin{matrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{matrix}$$

By finding the minimum probability of error using this 0-1 loss, we were able to classify the sample data resulting in the following confusion matrix. The loss is 1 when $L \neq D$ and 0 when $L = D$.

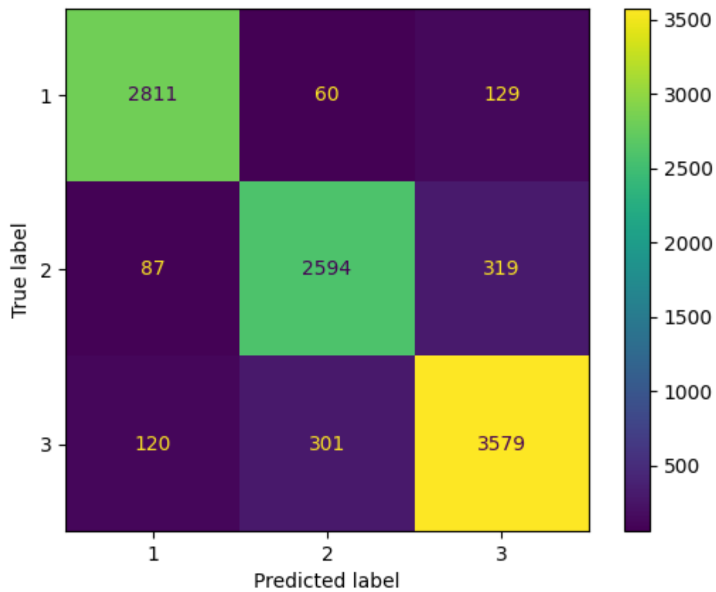

Figure 5. Confusion Matrix of 0-1 MAP Classifier

From this confusion matrix, we can see that each of the true predictions have similarly high prediction accuracy. From this classification there was about a 10% misclassification rate so there is a significant amount of overlap as can be seen in Figure 5.

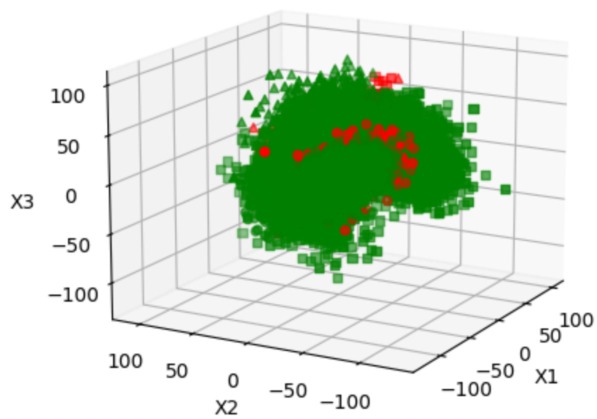

Figure 6. Scatter Plot of the 0-1 MAP Classifier

2.B ERM Classification with Bias Towards Label 3

For this question, we will be using loss matrices that care more about classifying the L = 3 samples correctly. To do this, we will use the following matrices.

$$\Lambda_{10} = \begin{matrix} 0 & 10 & 10 \\ 1 & 0 & 10 \\ 1 & 1 & 0 \end{matrix}$$

$$\Lambda_{100} = \begin{matrix} 0 & 100 & 100 \\ 1 & 0 & 100 \\ 1 & 1 & 0 \end{matrix}$$

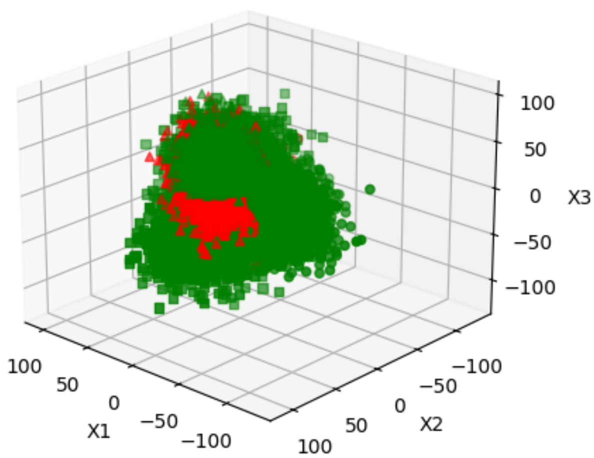

Figure 7. Scatter Plot of the ERM Classification Using the Λ_{10} loss matrix.

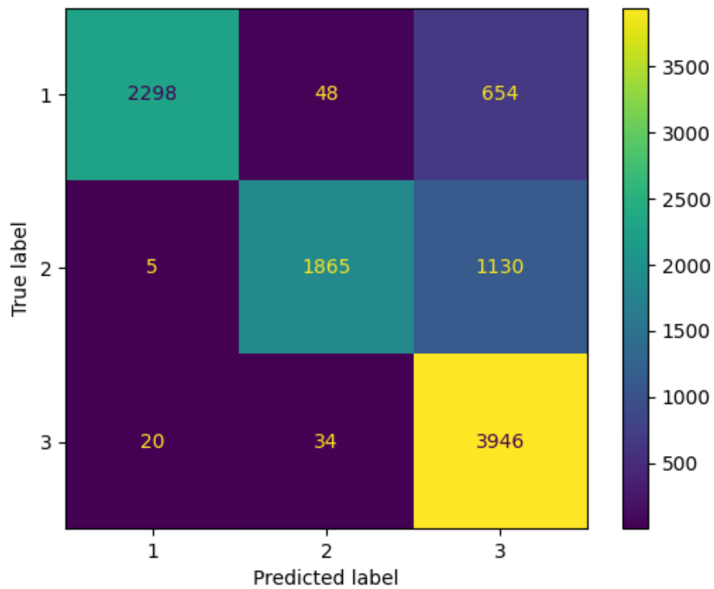

Figure 8. Confusion Matrix of the ERM Classification using the loss matrix Λ_{10}

As we can see in the confusion matrix from Λ_{10}, we can see that most of the samples with L=3 are classified correctly. We can also see that Labels 1 and 2 are often incorrectly classified as L=3 which makes sense. Compared to Figure 5, we can see that the correct predictions for labels 1 and 2 have decreased but the correct predictions for label 3 have increased as expected. From this loss matrix, about 18% of the samples are misclassified. This is an increase from the 10% in part 1 which is expected when prioritizing a class.

For the ERM classification with the Λ_{100} loss matrix, we can see similar changes as Λ_{10} but to a greater extent.

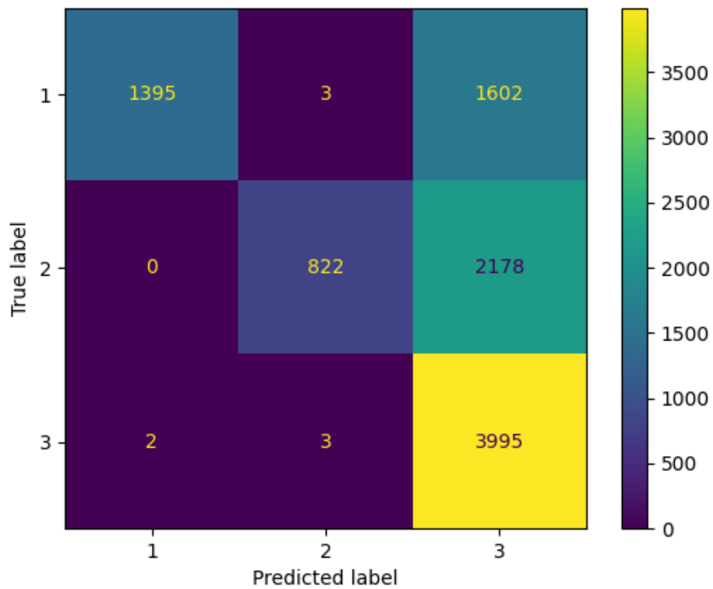

Figure 9. Confusion Matrix of the ERM Classification using the loss matrix Λ_{100}

As we can see in the confusion matrix from Λ_{100}, we can see that almost all of the samples with L=3 are classified correctly. We can also see that Labels 1 and 2 are incorrectly classified as L=3 at a high percentage. Compared to Figure 5, we can see that the correct predictions for labels 1 and 2 have decreased but the correct predictions for label 3 have increased as expected. From this loss matrix, about 38% of the samples are misclassified. This is an increase from the 10% in part 1 which is expected when prioritizing a class.

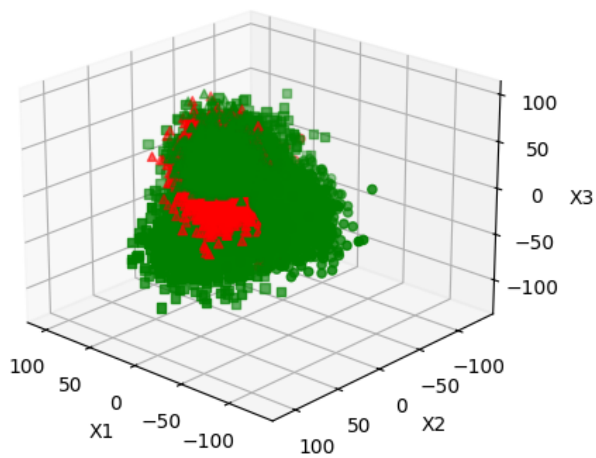

Figure 10. Confusion Matrix of the ERM Classification using the loss matrix Λ_{100}

3

For this question, we will be analyzing the UCI Wine Quality dataset and the Human Activity Recognition dataset. We will be calculating the Gaussian class conditional pdfs for each of the labels. Using these class conditional pdfs, we are going to generate decisions using the minimum-probability of error classifier. For the loss matrix, we are going to be using the following formula to generate the loss matrix:

$$\lambda = \frac{\alpha \mathrm{trace}(C_{SampleAverage})}{rank(C_{SampleAverage})}$$

From this loss matrix, we can calculate the label with the minimum probability of error for each of the samples. From these decisions, we can calculate the percentage of the samples that are correctly labeled.

3.1 Wine Quality Dataset

The class priors for this dataset are as follows:

P(L = 3)	0.0041
P(L = 4)	0.0333
P(L = 5)	0.2975
P(L = 6)	0.4487
P(L = 7)	0.1797
P(L = 8)	0.0357
P(L = 9)	0.0010

Figure 11. Class Priors of the Wine Quality Dataset

The sample mean and covariance of the dataset are calculated using the following formulas:

$$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

$$q_{jk} = \frac{1}{N - 1}\sum_{i=1}^{N}(X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k)$$

After generating the loss matrix of:

0	8.95	8.53	8.38	8.06	8.11	7.75
8.95	0	2.51	2.36	2.04	2.09	1.73
8.53	2.51	0	1.95	1.62	1.67	1.31
8.38	2.36	1.95	0	1.48	1.52	1.17
8.06	2.04	1.62	1.48	0	1.20	0.84
8.11	2.09	1.67	1.52	1.20	0	0.89
7.75	1.73	1.31	1.17	0.84	0.89	0

Figure 12. Loss Matrix for Wine Quality Dataset

Using the class priors, loss matrix, the sample mean, and sample covariance, we can now calculate the minimum probability of error for each of the samples. The decisions will be label with the minimum probability of error. From these decisions and correct labels, we can determine the percentage of correctly classified samples, which I found to be 50.02%. We can also find the confusion matrix(Figure 13). From Figure 12 and Figure 13, we can see that a large percentage of the class samples are labeled 5, 6, or 7.

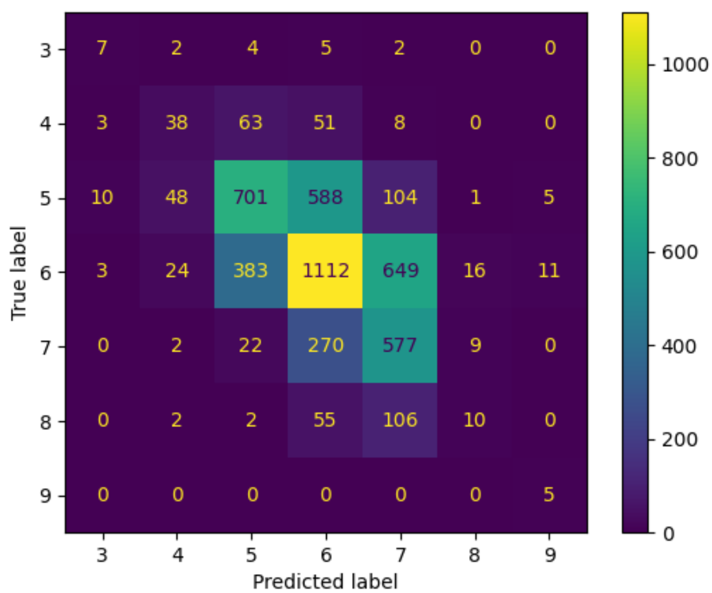

Figure 13. Confusion Matrix of True Labels and the Predicted Labels

3.2 Human Activity Dataset

For this dataset, we pretty much follow the exact same process as part 1. After generating the Gaussian distributions for each of the class labels and the loss matrix, we are able to classify the data. I tried using the training data in the classifier and I was getting 100% correct classification rate. So instead I trained the model with the training dataset then classified the testing dataset. From this testing dataset, I achieved a 94.2% classification accuracy which can be seen in the decision matrix in Figure 14.

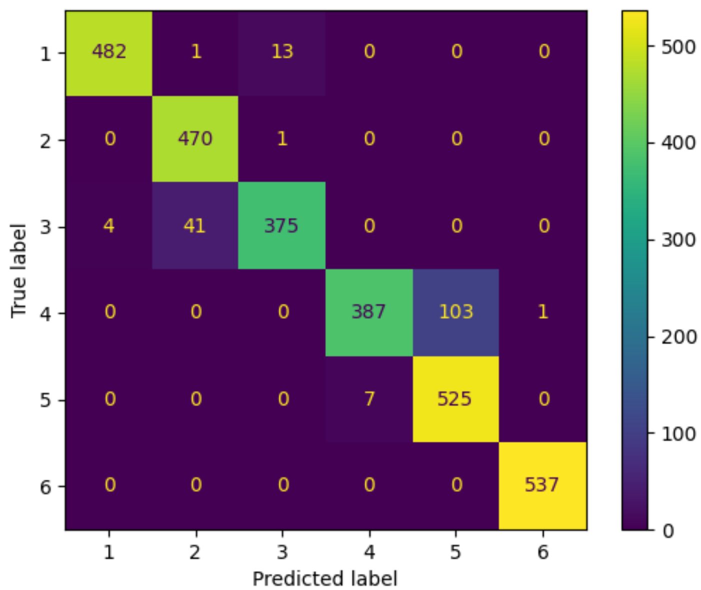

Figure 14. Confusion Matrix for the Human Activity Dataset

3.3 Discussion

As we can clearly see from the decision matrices for both datasets, the classifier does a much better job classifying the Human Activity Dataset than the Wine Quality Dataset. Gaussian class conditional models assume that the data within each class is normally distributed. These models are better suited for datasets that have certain characteristics. The characteristic is Continuous Data – data that follows a normal distribution.

For the Human Activity Dataset, all the data was continuous – all the data was from -1 to 1. So, the GMM was well suited for this data. For the Wine Quality dataset, the data was also continuous. But I think the issue with the dataset is that because all of the data is concentrated around a quality of 5,6,7 and the quality rankings of the wine is an ambiguous process in the first place, there is probably a lot of overlap in the decisions. Whereas the classes for the Human Activity Dataset are very distinct options. You can see this in the confusion matrix where there is a lot of overlap in the decisions for 5,6,7.

Appendix:

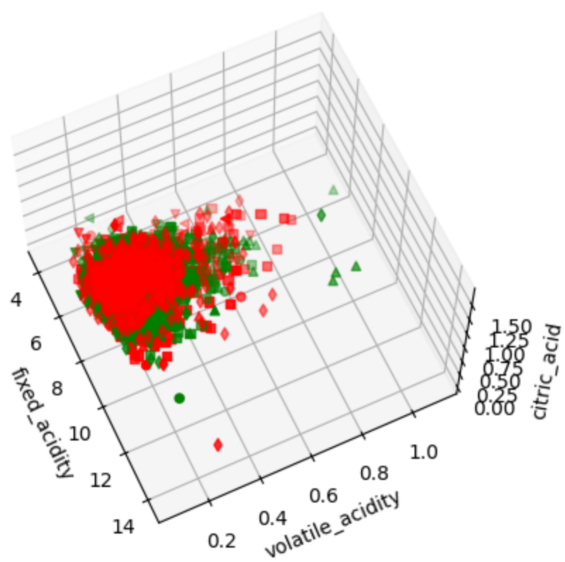

Figure 15. A Scatter Plot of the Subset of Fixed Acidity, Volatile Acidity, and Citric Acid.

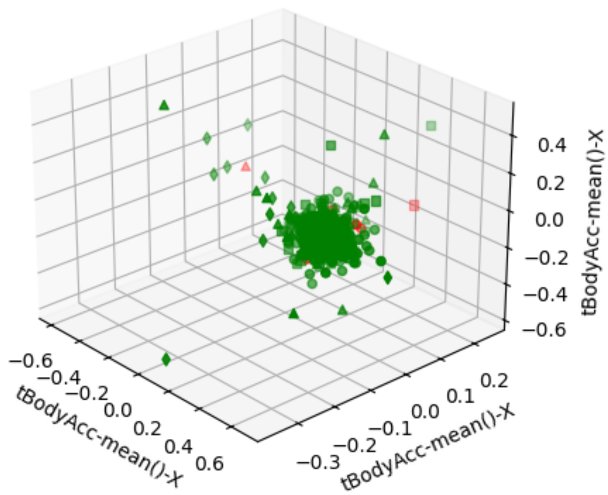

Figure 16. A Scatter Plot of the Subset of tBodyAcc-mean()-X, tBodyAcc-mean()-Y, and tBodyAcc-mean()-Z