EECE 5644 HW1

Owen McElhinney

February 10, 2020

Contents

1 Problem 1 **3**
1.1 Standard Bayesian Classifier 4
1.2 Naive Bayesian Classifier 4
1.3 Linear Discriminant Analysis 5

2 Problem 2 **7**
2.1 Problem Parameters 7
2.2 Decision Rule 8
2.3 Results 8

3 Problem 3 **9**

4 Appendix **10**
4.1 Code Repository: 10

1 Problem 1

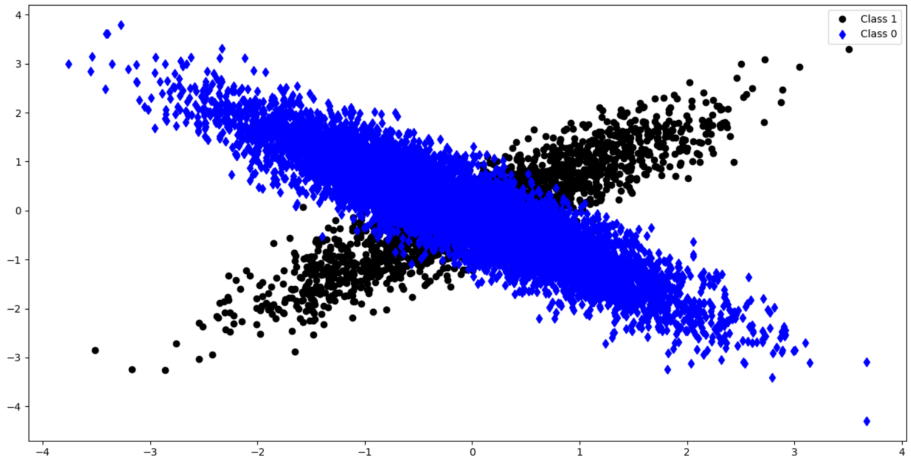

Figure 1: Data samples used for all parts of Problem 1

1.1 Standard Bayesian Classifier

Problem 1.1 uses a standard Bayesian classifier where the class priors and distribution parameters are known. Thus, our classification rule can be written as the standard log-likelihood function for Gaussians:

$$ln(P(x|L=1)) - ln(P(x|L=0)) \gtrless_{D=0}^{D=1} \gamma \tag{1}$$

$$P(x|L=i) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_i|^{\frac{1}{2}}} exp(-\frac{1}{2}(x-\mu_i)^T\Sigma^{-1}(x-\mu_i)) \tag{2}$$

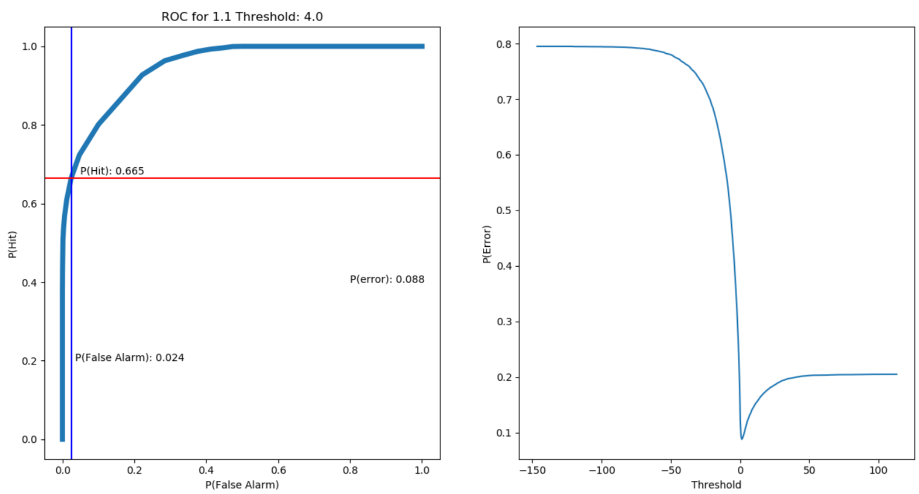

Figure 2: ROC curve and error as a function of threshold for problem 1 part 1

Using equations (1) and (2), the synthetic data was evaluated over a wide range of thresholds γ. Figure 2 shows the ROC curve as well as the P(error) as a function of the threshold values. As can be observed from the graph, the minimum threshold value was found to be at approximately 1.18 as derived from the data. At this threshold, the classifier achieved a P(Error) of 8.81 percent with a P(Hit) of 66.5 percent and a P(False alarm) of 2.4 percent

1.2 Naive Bayesian Classifier

Problem 1 part 2 evaluates the same data set under the Naive Bayesian Classifier. In this case, the covariances of both distributions are (falsely) assumed to be identity matrices. For this classifier, equations (1) and (2) still hold, but now the assumed identity matrix is plugged in to equation (2) instead of the true covariance matrices.

Figure 3 shows the ROC and error curves for this classifier. It can be noticed that the probability of error does not drop below the prior probability for class 1 of 0.2. As such, to achieve minimal P(error), all points are assigned to class 0

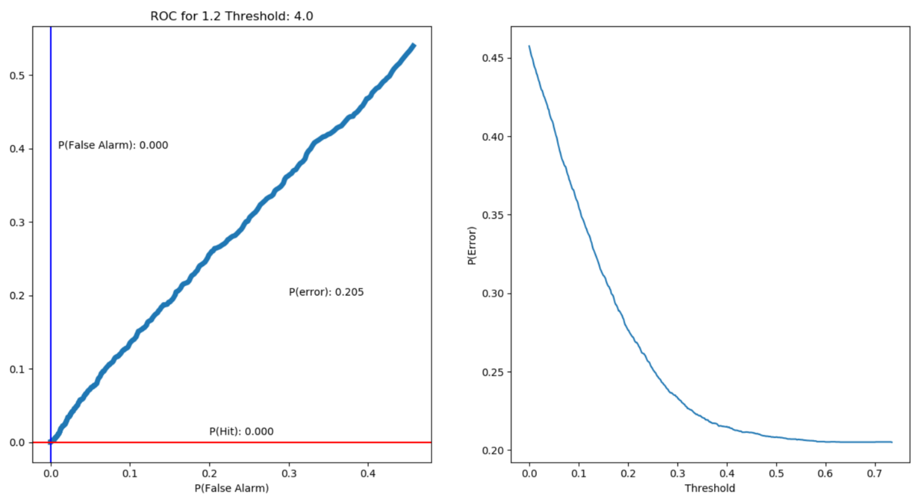

Figure 3: ROC Curve and error as a function of threshold for problem 1 part 2

and the minimum error is 20 percent. Since all points are classified as label 0, our P(hit) is 0 and P(False alarm) is 0.

The large error in this case comes from the false assumption for the covariance matrices. It can be observed from Figure 1 that these distributions have large correlation components on the off-diagonals. By ignoring these, a large amount of error is introduced to the problem

1.3 Linear Discriminant Analysis

The final case considered for this data set was the Linear Discriminant Analysis (LDA) based classifier. This classifier finds a transformation for the data to maximize separation between two classes as the eigen-vector associated with the largest eigen-value of the separability matrix:

$$W = \frac{(\mu_0 - \mu_1)(\mu_0 - \mu_1)^T}{\Sigma_0 + \Sigma_1} \tag{3}$$

The means and covariances for this equation were taken as the sample estimates from the generated data.

With this transformation, the data was evaluated over a wide range of threshold values. As can be seen in Figure 4 this case was similar to part 2 in that the error never dropped below the class prior for label 1. As such, the threshold corresponding to the minimum probability of error simply places all data points in class 0. This means the P(Hit) = 0 and P(False Alarm) = 0 and P(error) 0.2.

In this case, the large error stems from the over-lap of the data. The two distributions share enough overlap, that a classifier based on linearly separating

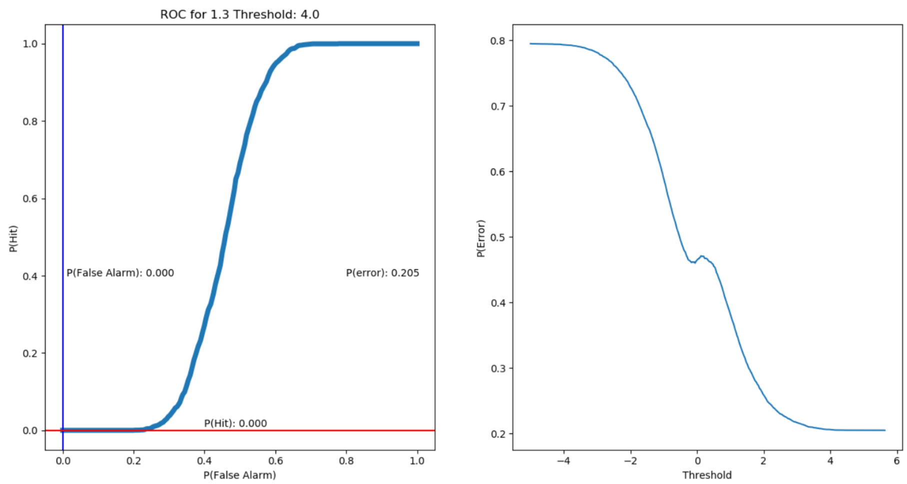

Figure 4: ROC Curve and error as a function of threshold for question 1 part 3

the two distributions is rendered useless.

2 Problem 2

Problem 2 asks us to create our own Bayes Decision problem with two classes, each composed of two Gaussians. The next section will outline the parameters used in the creation of the problem, then go over the decision rule used for this classifier, and finally present a review of the results for the classifier.

2.1 Problem Parameters

Prior probabilities:

Label 0 [P(l = 0) = 0.6]

Label 1 [P(l=1) = 0.4]

Internal parameters for Label 0:

- Label 0 Mixture 0

  Prior Probability: 0.7

  Mean: $\begin{pmatrix} 2 & 1 \end{pmatrix}$

  Covariance: $\begin{pmatrix} 1 & 0.25 \\ 0.25 & 1 \end{pmatrix}$

- Label 0 Mixture 1

  Prior Probability: 0.3

  Mean: $\begin{pmatrix} 1 & -1 \end{pmatrix}$

  Covariance: $\begin{pmatrix} 1 & -0.2 \\ -0.2 & 1 \end{pmatrix}$

Internal parameters for Label 1:

- Label 1 Mixture 0

  Prior Probability: 0.6

  Mean: $\begin{pmatrix} -1 & -1 \end{pmatrix}$

  Covariance: $\begin{pmatrix} 1 & -0.4 \\ -0.4 & 1 \end{pmatrix}$

- Label 1 Mixture 1

  Prior Probability: 0.4

  Mean: $\begin{pmatrix} 0 & -2 \end{pmatrix}$

  Covariance: $\begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix}$

2.2 Decision Rule

The decision rule is similar to the one used in part one of question one, but now each class is composed of two sub-classes. As such, our decision rule becomes:

$$ln(q_{10}P(x|L=1_0)+q_{11}P(x|L=1_1))-ln(q_{00}P(x|L=0_0)+q_{01}P(x|L=0_1)) \gtrless \gamma \tag{4}$$

Where P(x|L=i_j) is evaluated using equation (2) with the specific mean vector and covariance matrix for that subclass.

2.3 Results

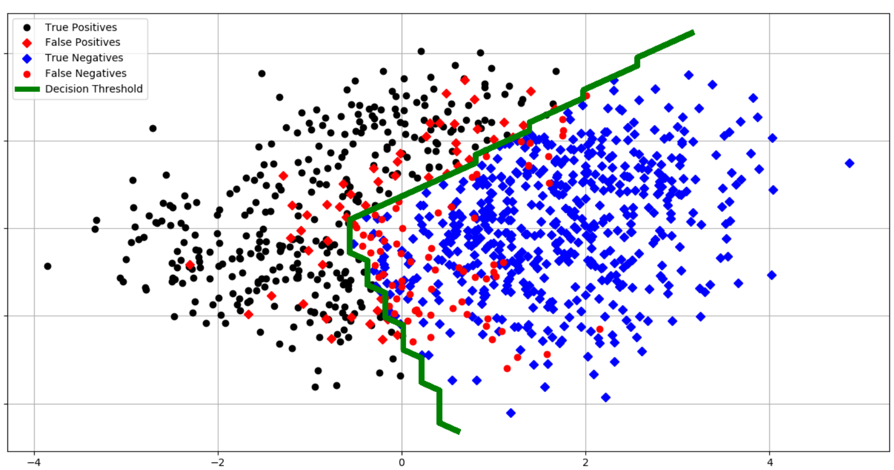

Figure 5: Results for the classifier and decision boundary for the minimum P(error) threshold

Pictured above are the results for each point in the synthetic data. The minimum probability of error achievable for this created system was approximately 12.6 percent. The decision boundary plotted is for a threshold of approximately 0.9 which can be seen in 5 which shows P(error) as a function of threshold. At this minimum threshold, the classifier achieves a P(Hit) of approximately 82.1 percent at a false alarm rate of approximately 9 percent.

The two labels were fairly clustered together, so significant error has to be expected in this scenario. The means of the two labels were separated by just two standard deviations in some locations which accounts for a lot of the error.

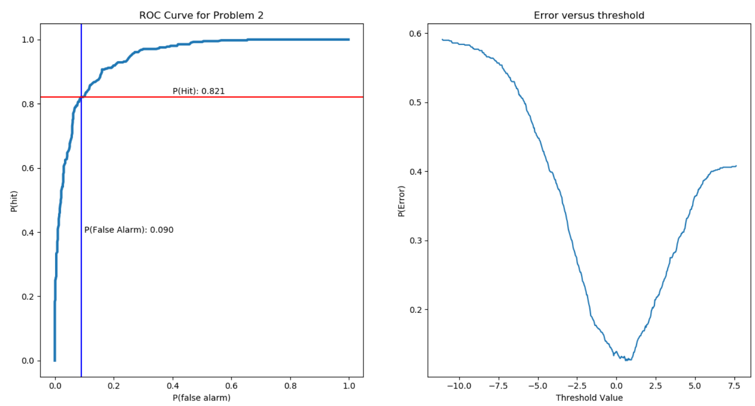

Figure 6: ROC Curve and P(Error) as a function of threshold for this classifier

3 Problem 3

Problem 3 investigates a minimum error classification rule for two classes with equal class prior probabilities. With these parameters, the natural fit for this problem was the Maximum-Likelihood classifier. This decision rule is given by:

$$\frac{P(x|L=1)}{P(x|L=0)} \gtrless 1_{D=0}^{D=1} \tag{5}$$

Figure 7 demonstrates the idea graphically: for any incoming x, the likelihood of that x is evaluated for each of the distributions, and the higher value is chosen. With the given parameters this gives the simple rule that if x>0 decide 1, else decide 0. For x=0 it is equally likely for both classes and further information is required for an informed decision.

As for error, the minimum error rule is to set the threshold at their intersection of x=0. For this rule, the actual error in the model can be found by figuring out the likelihood that distribution 0 is greater than 0 and distribution 1 is less than zero. Formally, this error is given by:

$$P_{01} + P_{10} \tag{6}$$

$$0.5\int_{0}^{\infty} N(-2,1) + 0.5\int_{-\infty}^{0} N(2,1) \tag{7}$$

Which, according to principles of the ideal Gaussian is approximately 2.2 percent.

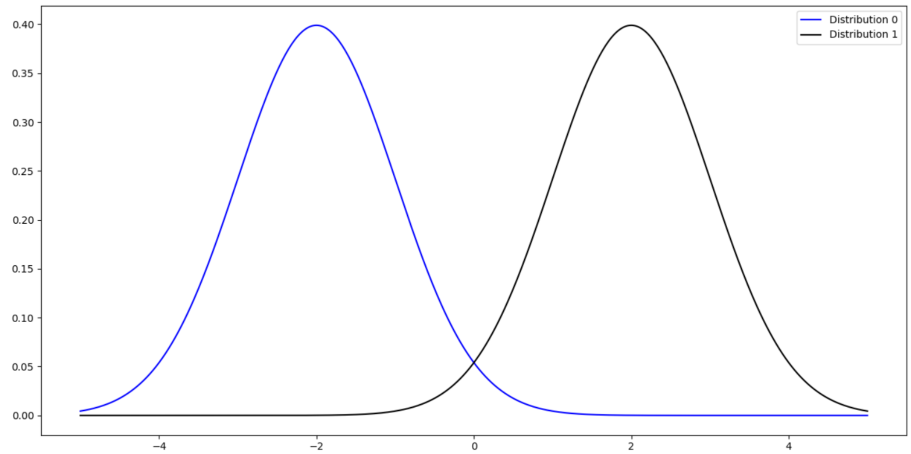

Figure 7: Graphical demonstration of the distributions for problem 3

4 Appendix

4.1 Code Repository:

https://github.com/o-mcelhin/eece-5644/tree/master/Homework_1