EECE5644 Spring 2020 – Take Home Exam 1

Name: Chenan Wang

Email: wang.chena@husky.neu.edu

Submit: Monday, 2020-February-10 before 11:45 ET

Question 1 (60%)

a)

Solution:

We use given mean and sigma to synthesis 10000 data.

According to the likelihood Function, we have:

$$\frac{\mathrm{p}(\mathrm{x}|\mathrm{L}=1)}{\mathrm{p}(\mathrm{x}|\mathrm{L}=0)} > \frac{\lambda_{12}-\lambda 22}{\lambda_{21}-\lambda_{11}}\frac{\mathrm{P}(\omega_2)}{\mathrm{P}(\omega_1)}$$

We first calculate $\frac{p(\mathbf{x}|L=1)}{p(\mathbf{x}|L=0)}$ by:

probability density function with multi-variables:

$$\left(\frac{1}{2\pi}\right)^{k/2}|\Sigma|^{-\frac{1}{2}}e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

Probability of Error is calculated by:

$$(p10\cdot n0\ +\ p01\cdot n1)/N$$

Where p10 is probability of false positive; p01 is probability of false negative; n0 is the number of 0-label examples; n1 is number of the number of 1-label examples; N is the number of both examples.

Then, we calculate the threshold $\gamma = 4$. We verify it by varying the gamma from 0 to infinity, and we draw the probability of error vs. gamma graph:

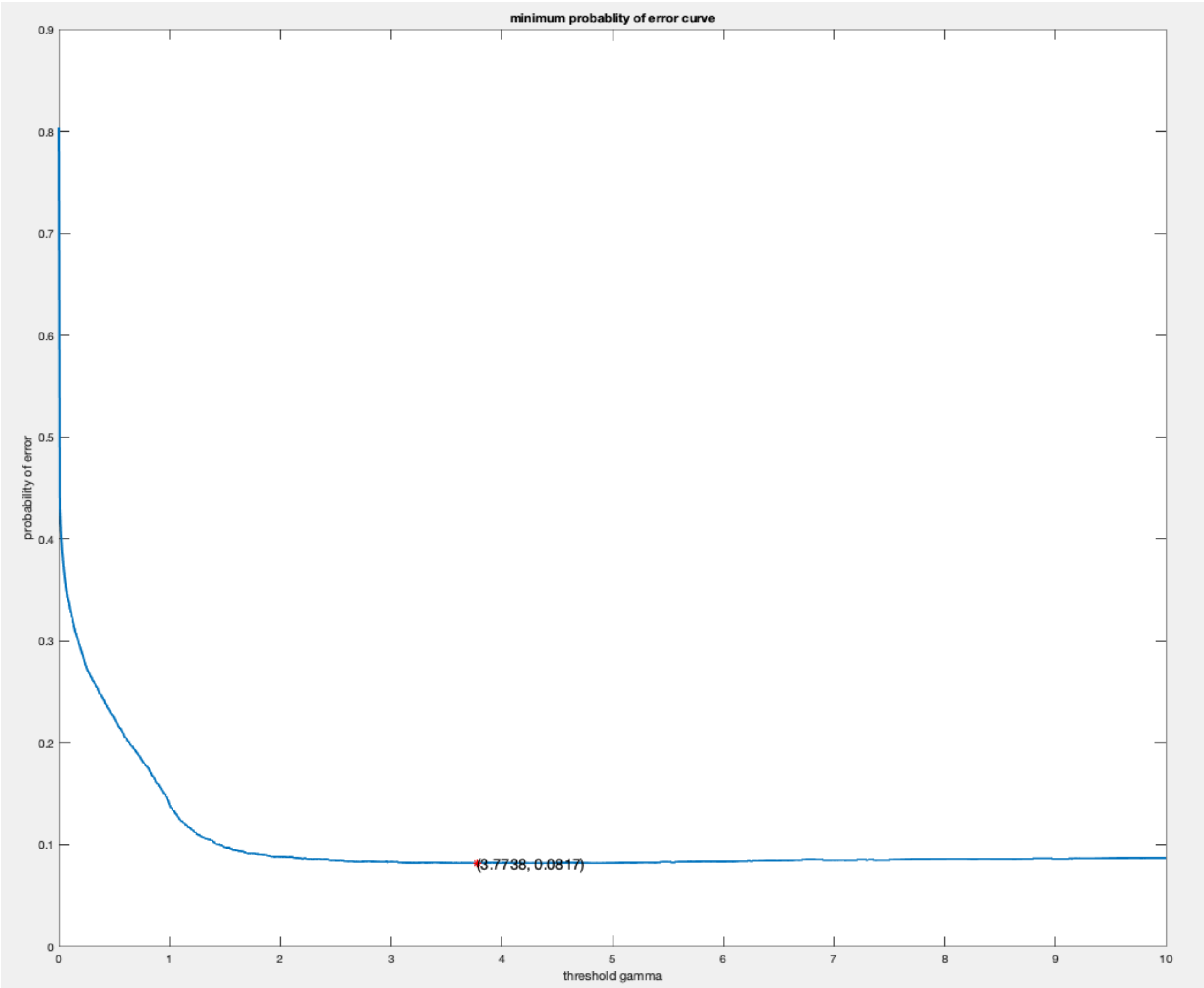

Figure 1.1: probability of error – threshold gamma graph

Then, we compute true positive and false positive to draw the roc curve:

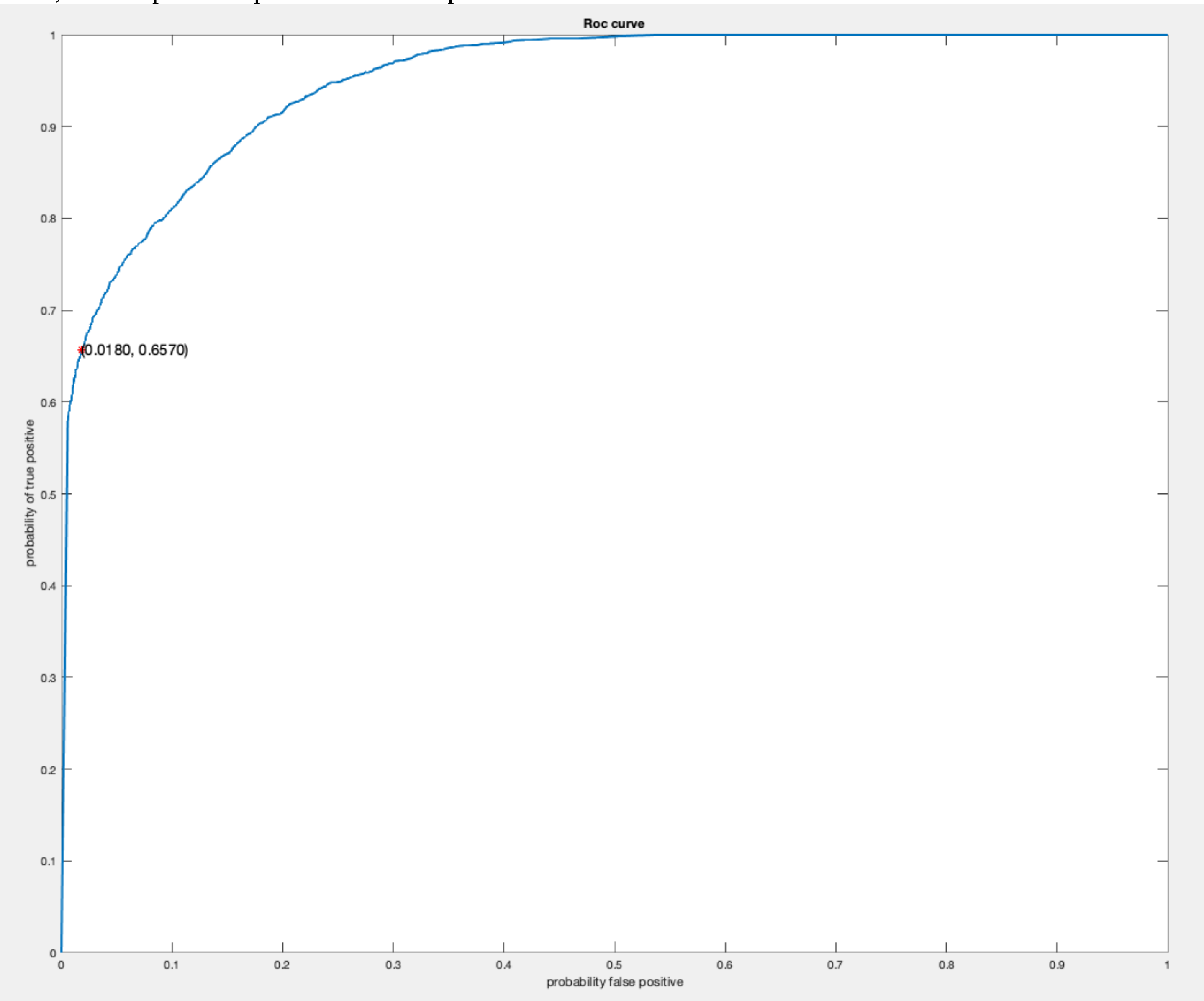

Figure 1.2: the roc curve.

Notice on the roc curve, we superimpose the true positive and false positive values attained by this minimum-P(error) classifier, that is, (0.0180,0.6570).

Conclusion:

Calculated gamma: 4

threshold gamma (from the graph): 3.7738

minimum probability of error (generated by calculated gamma): 0.0823

minimum probability of error (generated by graph observation gamma): 0.0817

b)
Here, we repeat the same steps as in the previous cases, for the Naïve Bayesian approach.
The probability of error vs. gamma graph:

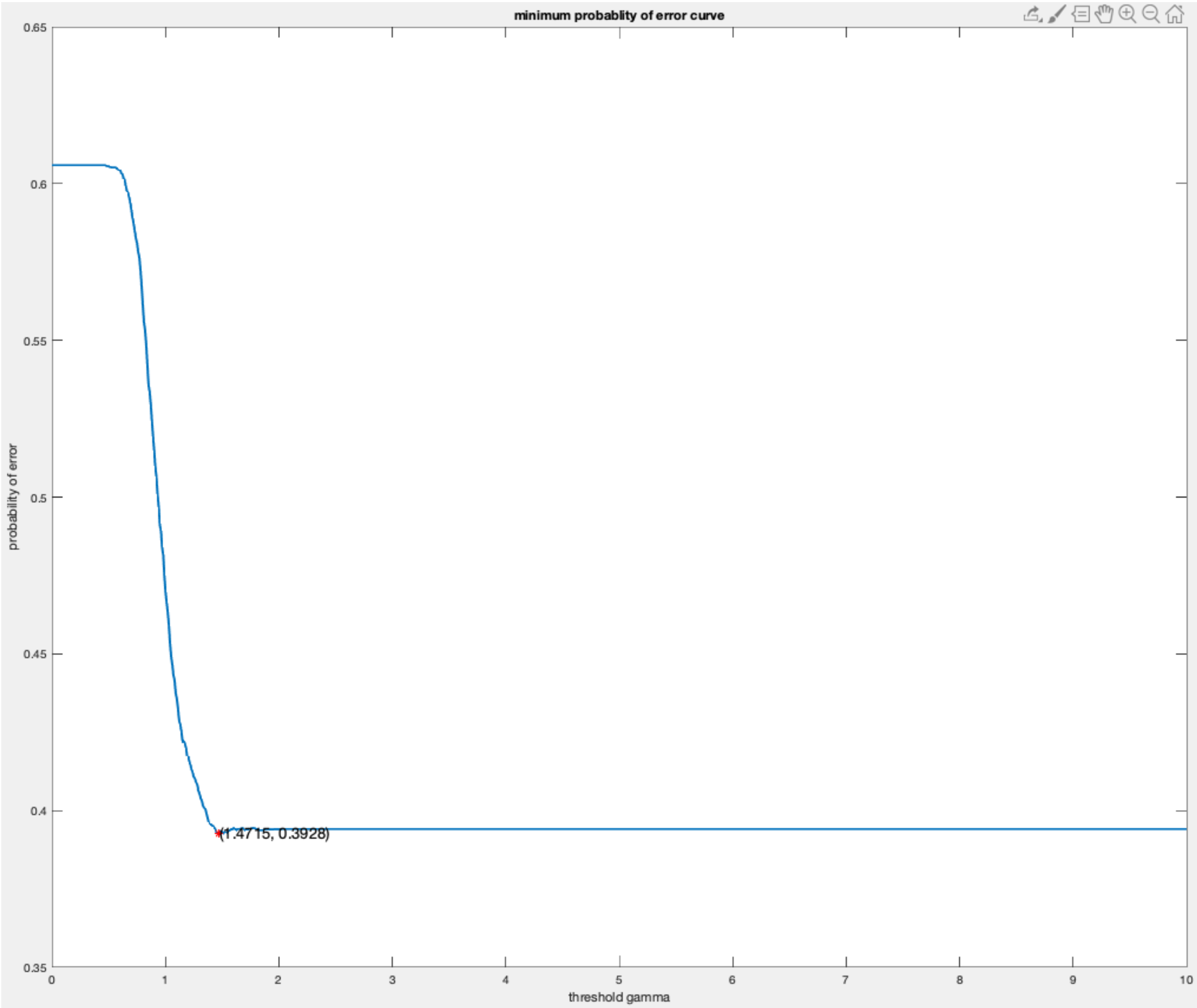

Figure 1.3: probability of error – threshold gamma graph

The ROC curve:

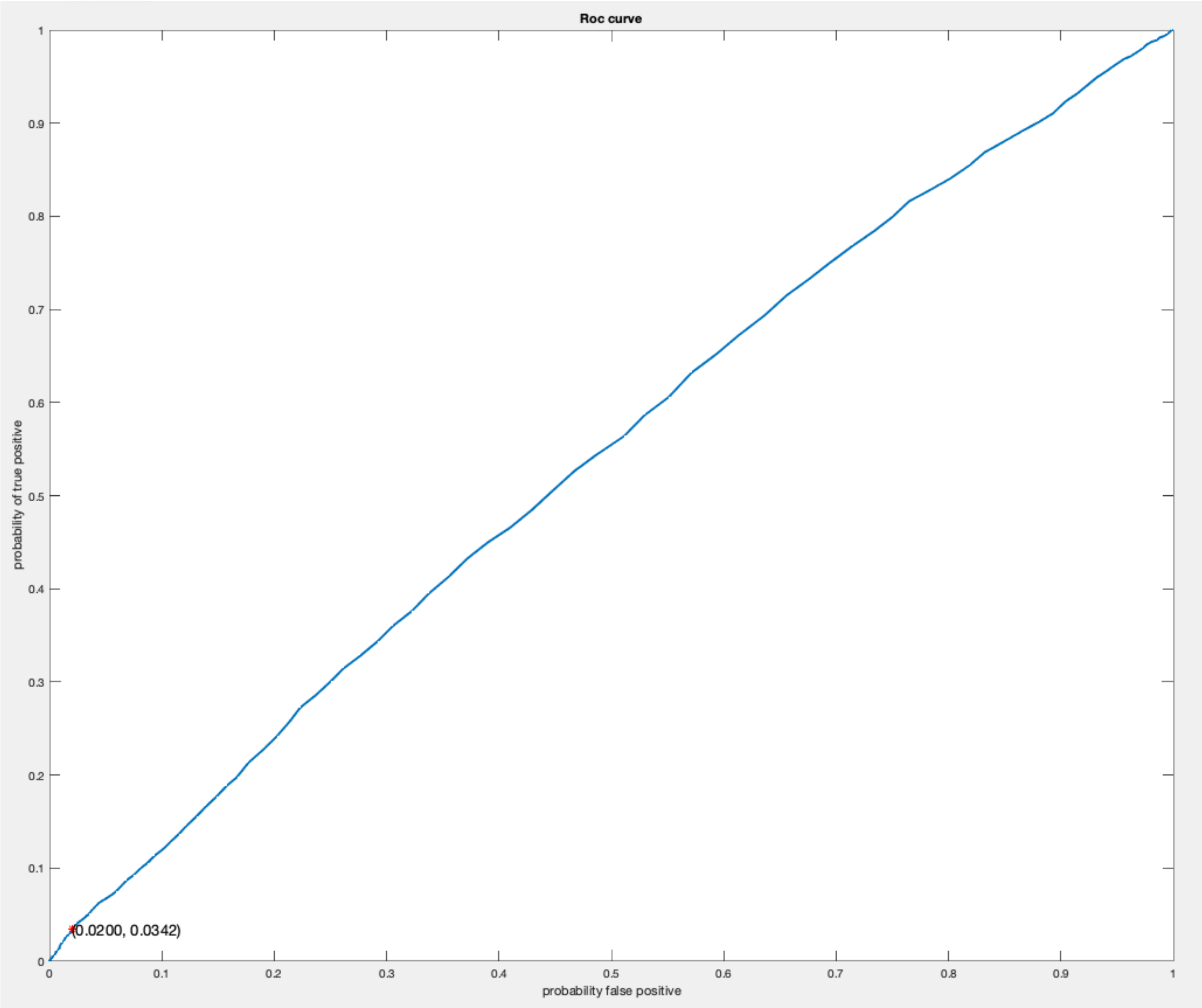

Figure 1.4: the roc curve.

Notice on the roc curve, we superimpose the true positive and false positive values attained by this minimum-P(error) classifier, that is, (0.0200,0.0342).

Conclusion:

Calculated gamma: 1.5

threshold gamma (from the graph): 1.4715

minimum probability of error (generated by calculated gamma): 0.393

minimum probability of error (generated by graph observation gamma): 0.3928

c) Here, we repeat the same steps as in the previous cases, for the LDA approach.
The LDA projection graph:

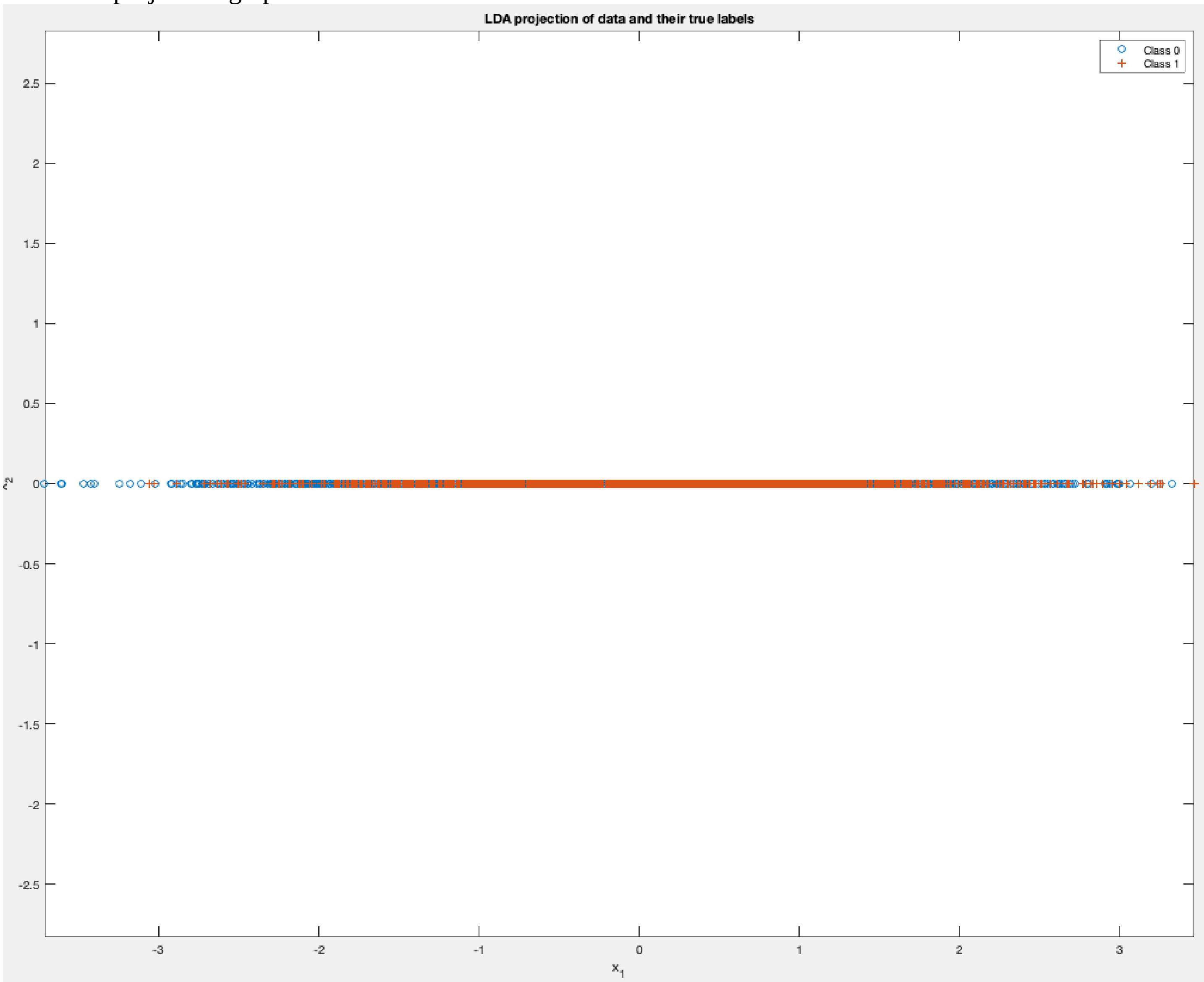

Figure 1.7: LDA projection graph with true labels

The probability of error vs. tau graph:

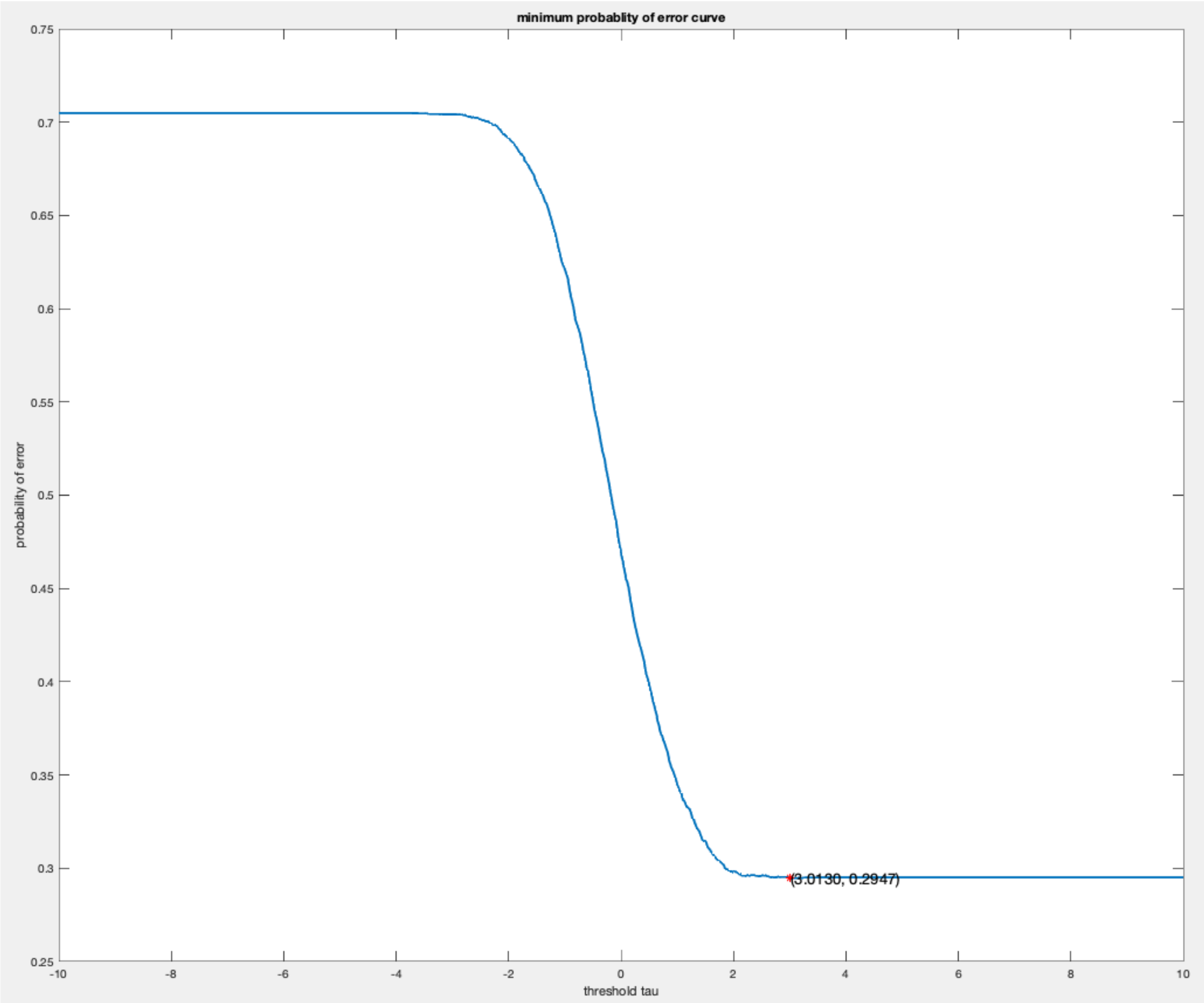

Figure 1.5: probability of error – threshold tau graph

The Roc curve:

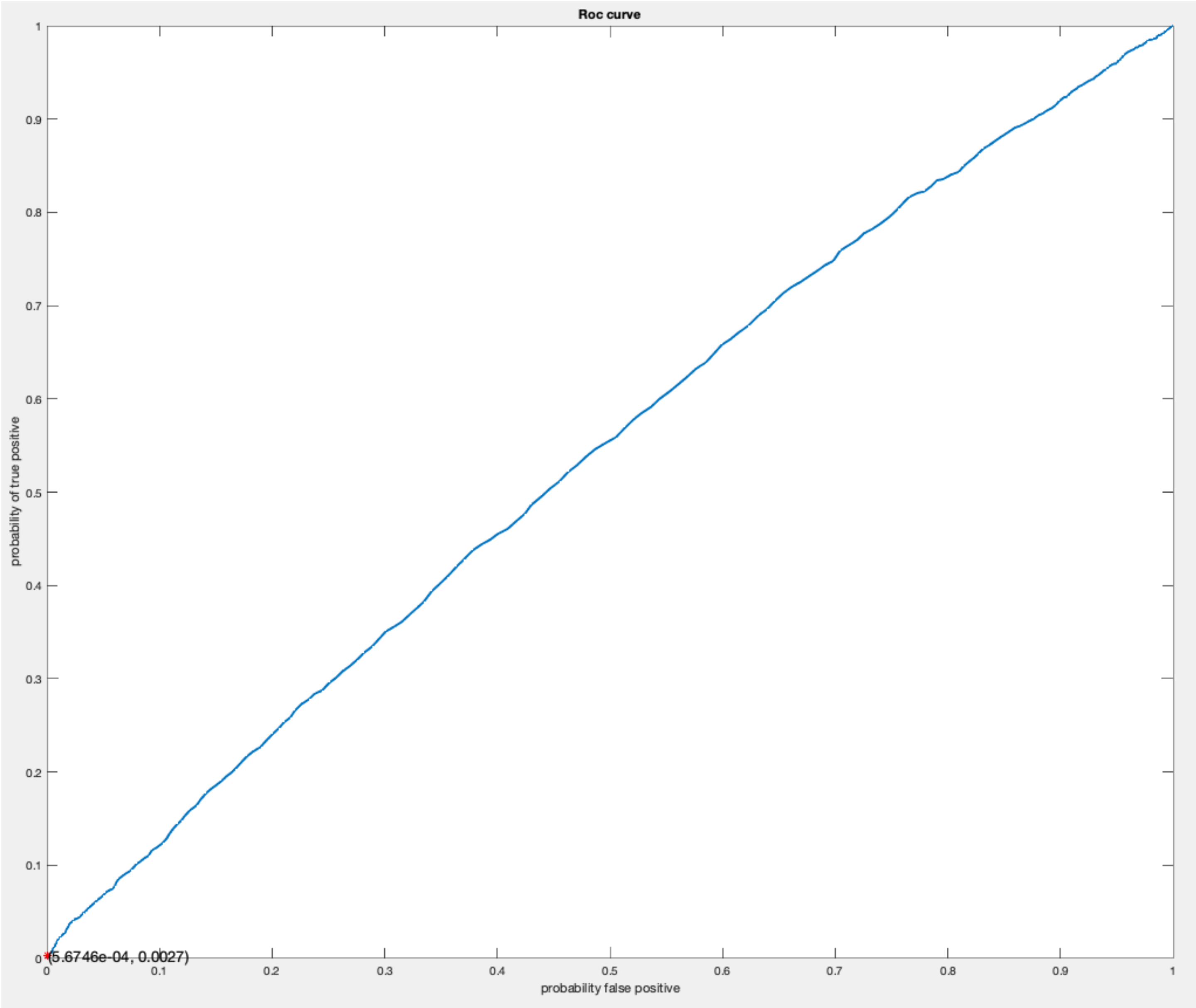

Figure 1.6: the roc curve.

Notice on the roc curve, we superimpose the true positive and false positive values attained by this minimum-P(error) classifier, that is, (5.6746e-04, 0.0027).

Conclusion:

threshold tau (from the graph): 3.0130

minimum probability of error (generated by graph observation tau): 0.2947

Question 2(30%)

1.

We plot 1000 samples for two class-conditional pdfs, both in the form of mixtures of two Gaussians. The scatter plot is below:

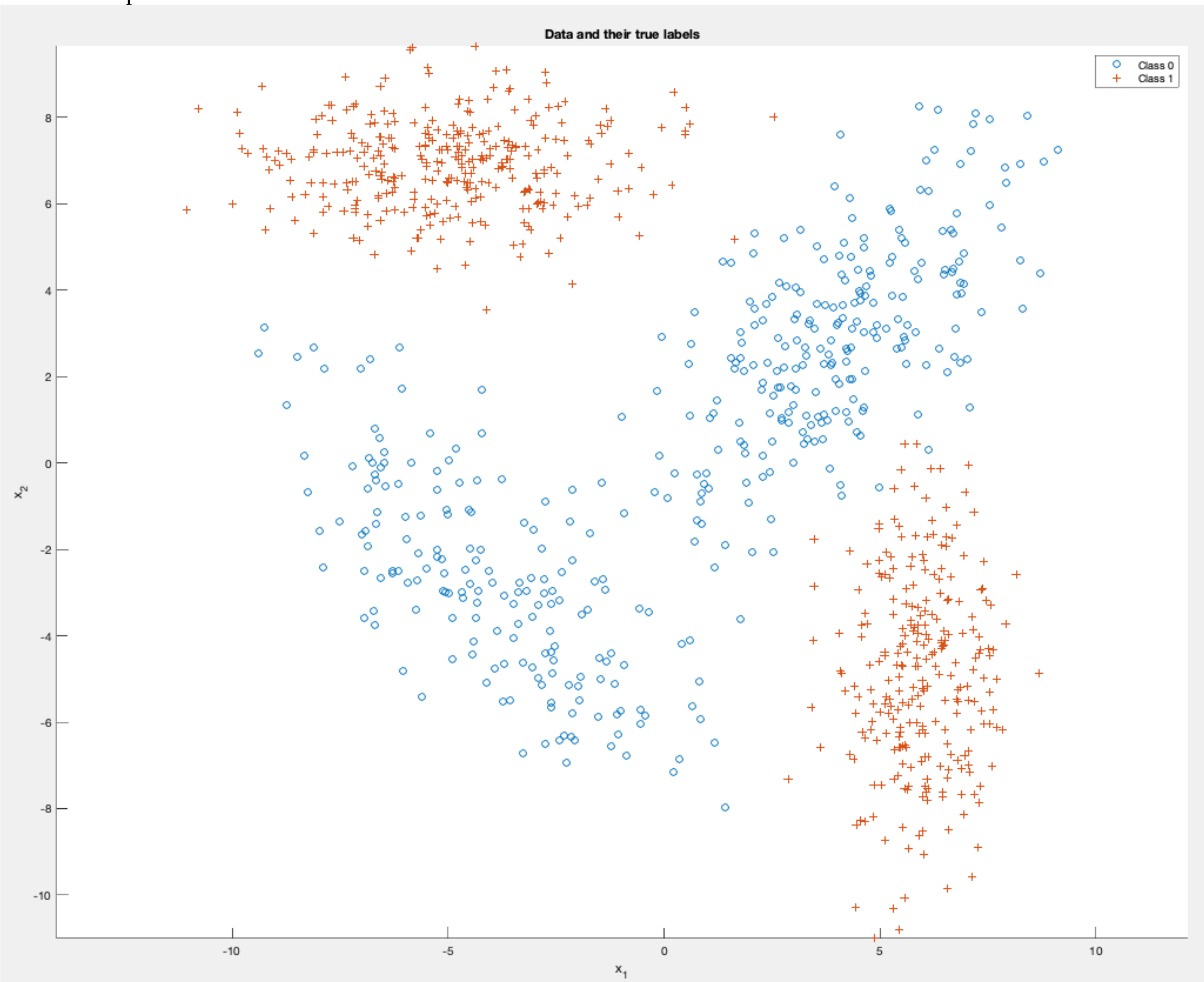

Figure 2.1 the scatter plot.

Then we draw decision boundaries:

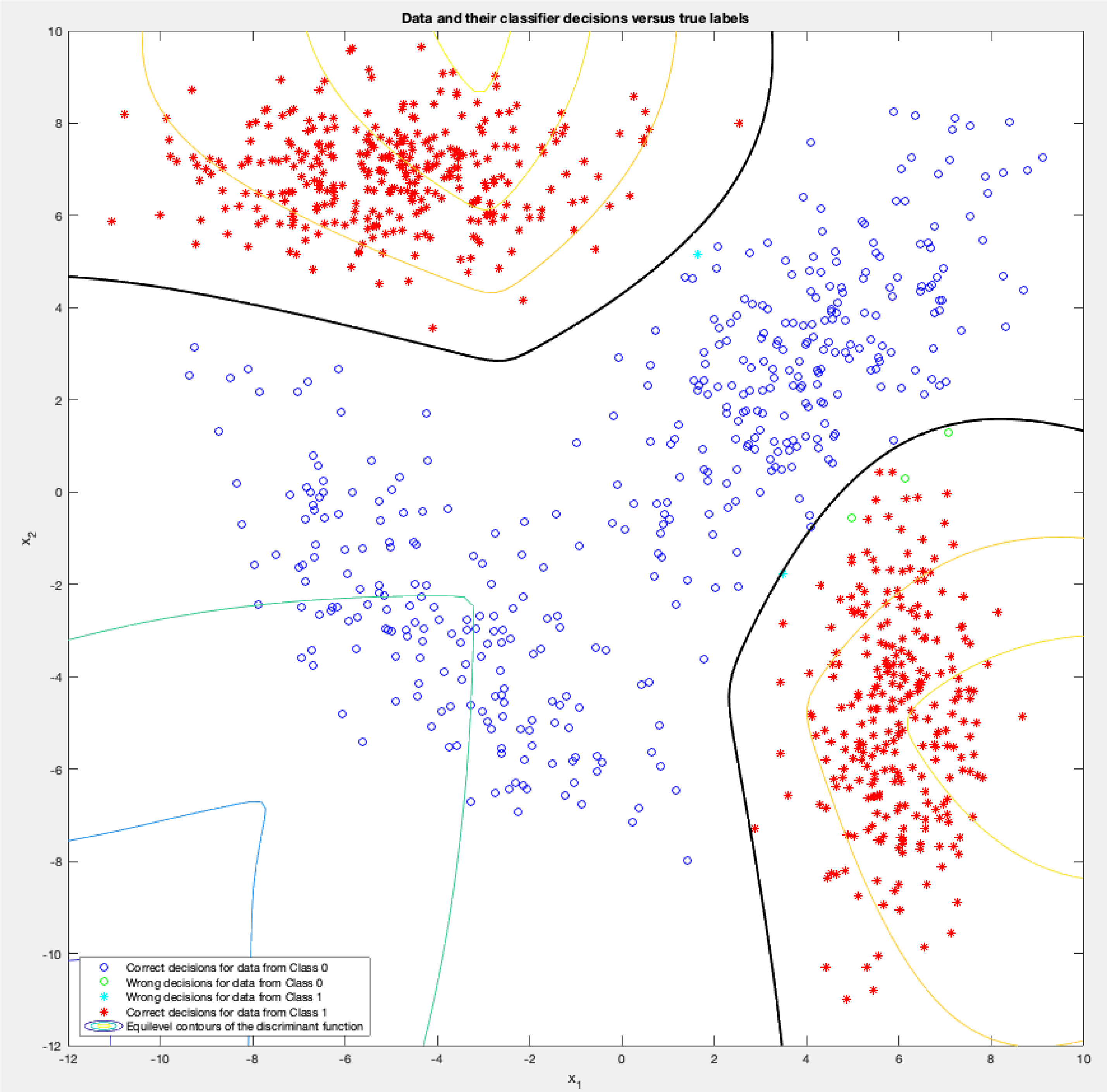

Figure 2.2: scatter plot with decision boundaries. The decision boundaries are the black line in the graph.

As before, we plot the minimum probability of error curve:

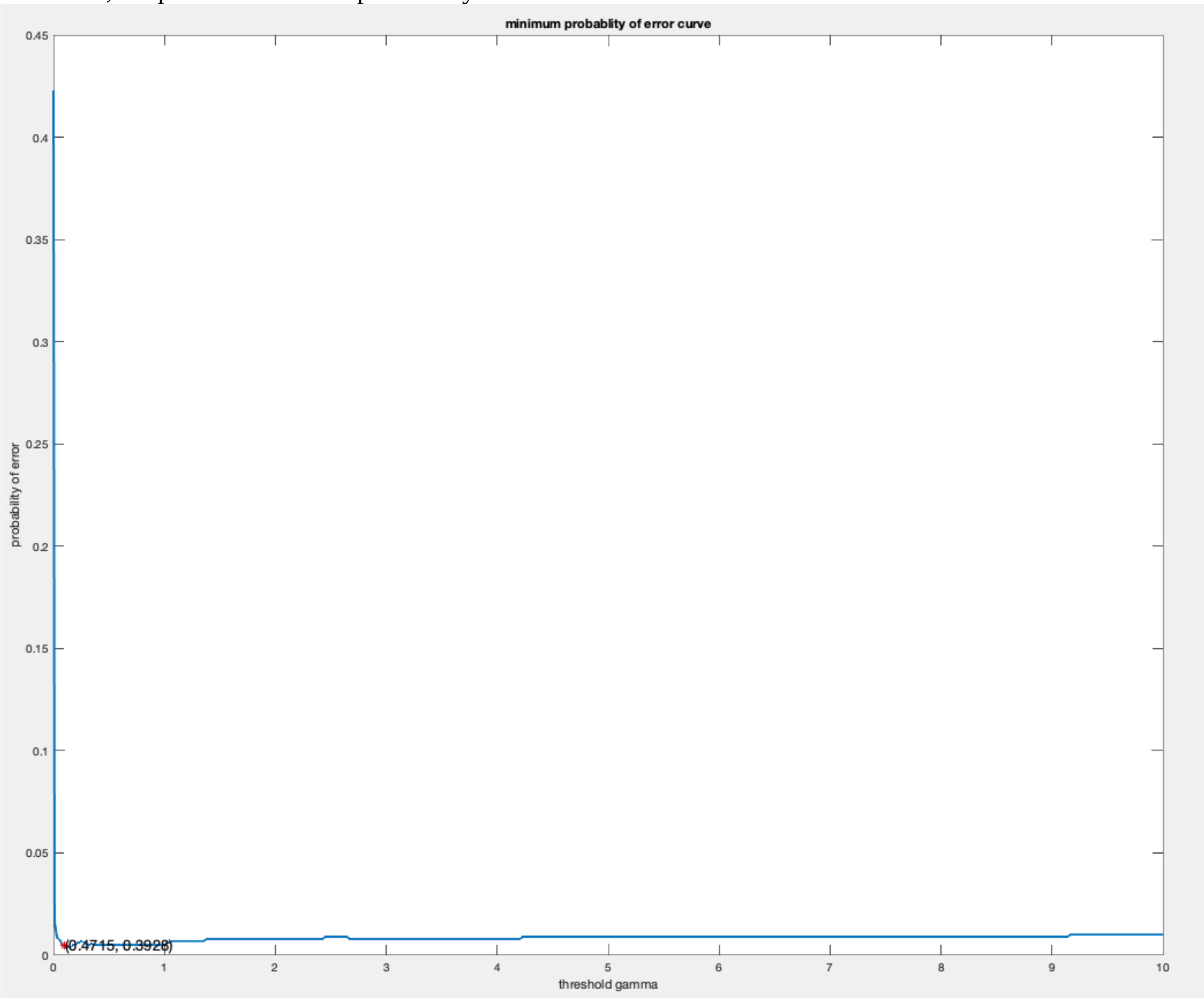

Figure 2.3: minimum probability of error curve

Finally, we use minimum-P(error) classification rule

$$\frac{p(x|L=1)}{p(x|L=0)} > \frac{\lambda_{12} - \lambda 22}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}$$

to calculate threshold gamma, which is 0.6667.

Probability of Error is calculated by:

$$(p10 \cdot n0 \ + \ p01 \cdot n1)/N$$

Where p10 is probability of false positive; p01 is probability of false negative; n0 is the number of 0-label examples; n1 is number of the number of 1-label examples; N is the number of both examples.

Use the gamma, we calculate the minimum probability of error:

With the calculated gamma, it is 0.0050.

With observed gamma from the graph, it is 0.0040.

Question 3
In this question, we have two 1-dimension Gaussians.
We use minimum-P(error) classification rule

$$\frac{p(x|L=1)}{p(x|L=0)} > \frac{\lambda_{12} - \lambda 22}{\lambda_{21} - \lambda_{11}} \frac{P(\omega_2)}{P(\omega_1)}$$

to calculate threshold gamma, which is 1.

Probability of Error is calculated by:

$$(p10 \cdot n0 \ + \ p01 \cdot n1)/N$$

Where p10 is probability of false positive; p01 is probability of false negative; n0 is the number of 0-label examples; n1 is number of the number of 1-label examples; N is the number of both examples.

As before, we plot the minimum probability of error curve:

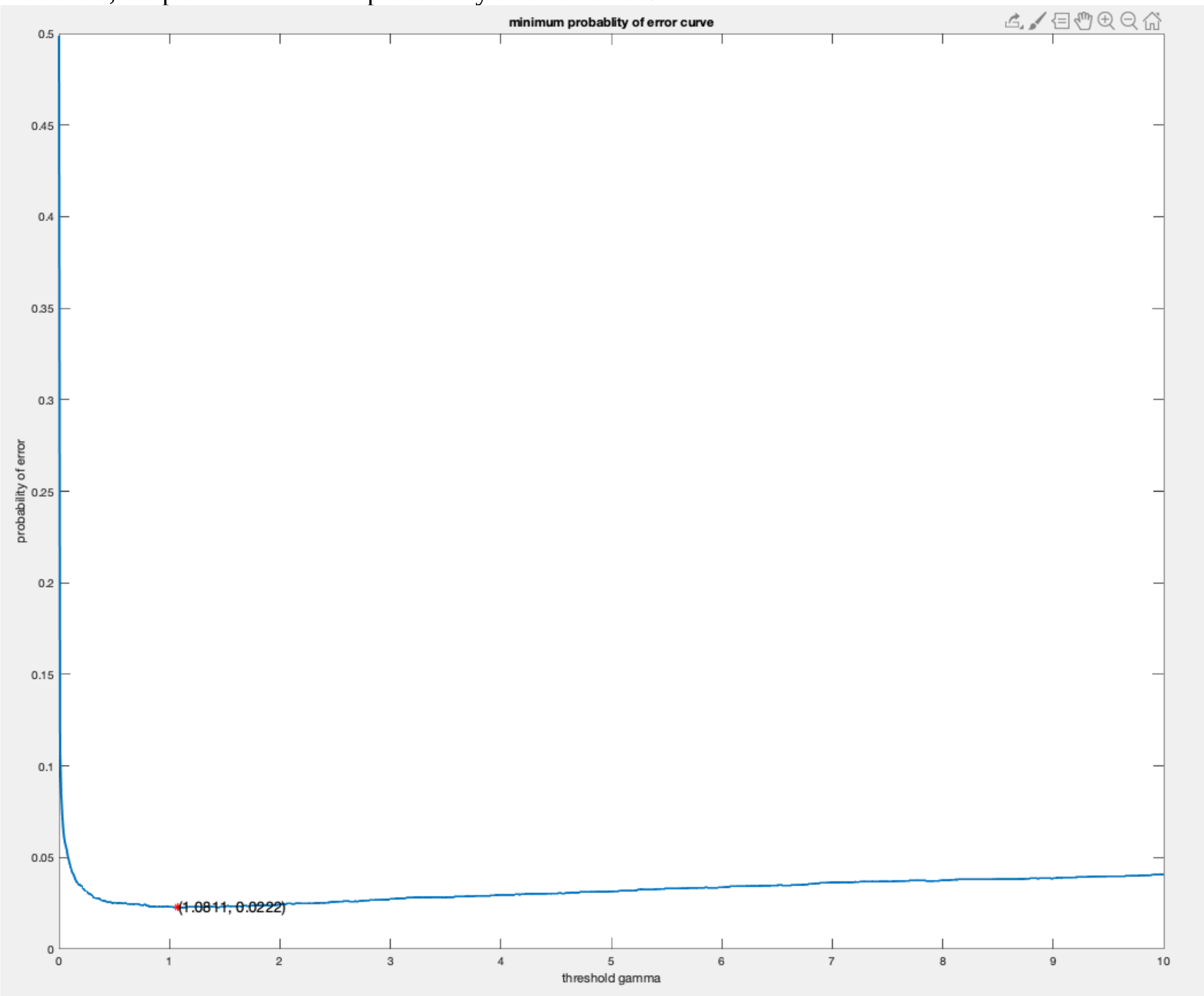

Figure 3.2: minimum probability of error curve. From the graph, when threshold is 1.0811, the probability of error is 0.0222.

Use the gamma, we calculate the minimum probability of error:
With the calculated gamma, it is 0.023.
With observed gamma from the graph, it is 0.0222.

We have the pdf graph like in figure 3.1:

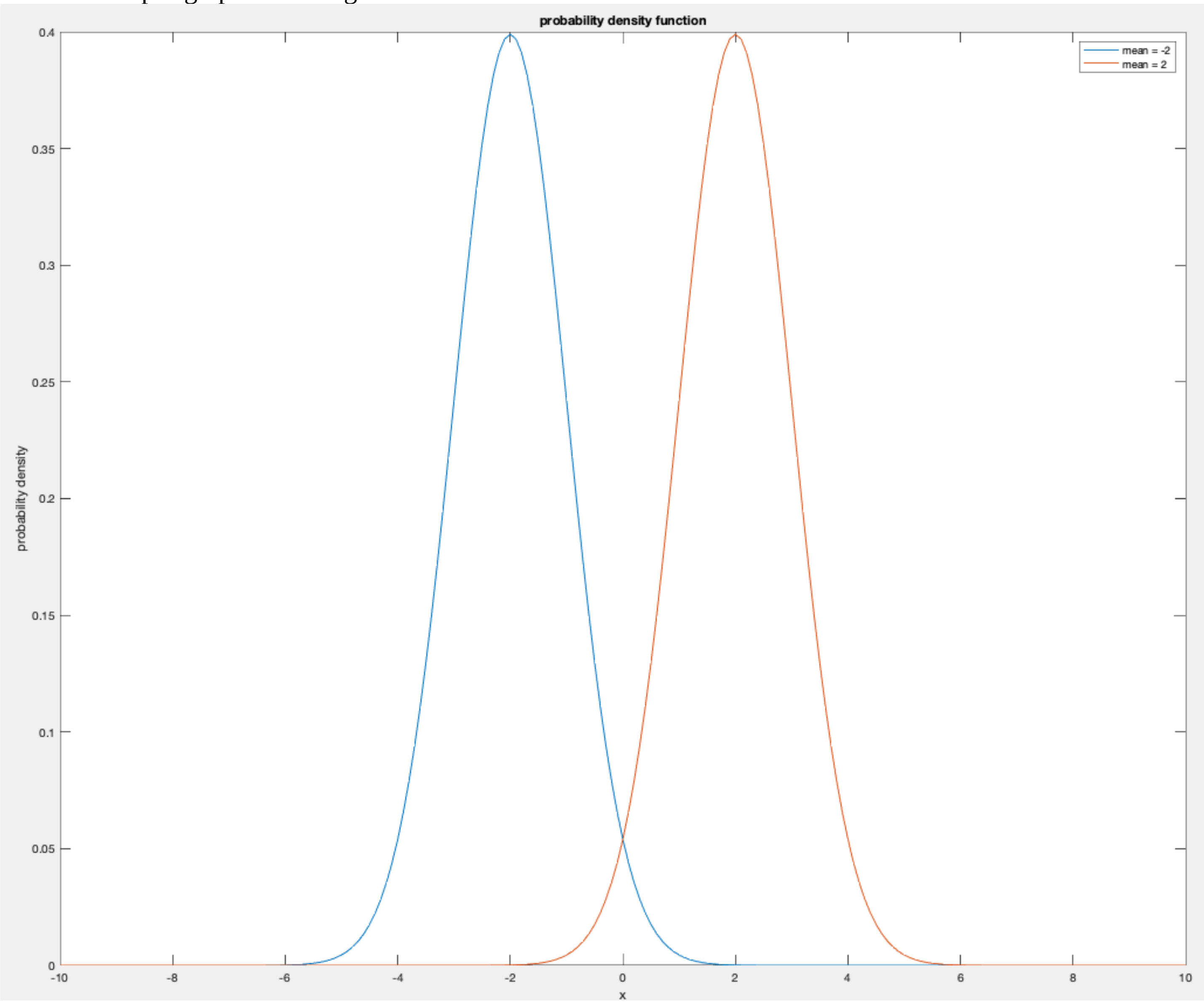

Figure 3.1: pdf graph for two unidimensional gaussian.

Then we need to calculate the integral in order to verify the minimum probability of error, with the equation from stackexange.com[1], we can calculate the area that is overlapped by two curves.

Let c denote the point of intersection where the pdf's meet in the green zone of your plot Then, the area of your green intersection zone is simply:

$$P(X_1 > c) + P(X_2 < c) = 1 - F_1(c) + F_2(c) = 1 - \frac{1}{2}\text{erf}\left(\frac{c-\mu_1}{\sqrt{2}\sigma_1}\right) + \frac{1}{2}\text{erf}\left(\frac{c-\mu_2}{\sqrt{2}\sigma_2}\right)$$

where erf(.) is the error function.

Point c is the solution to $f_1(x) = f_2(x)$ within the green zone, which yields:

$$c = \frac{\mu_2\sigma_1^2 - \sigma_2\left(\mu_1\sigma_2 + \sigma_1\sqrt{(\mu_1-\mu_2)^2 + 2\left(\sigma_1^2-\sigma_2^2\right)\log\left(\frac{\sigma_1}{\sigma_2}\right)}\right)}{\sigma_1^2 - \sigma_2^2}$$

Figure 3.2: the function used to calculate area overlapped.

The overlapping area is 0.0455, and the minimum probability of error is half of it, which is 0.0222.
Thus, the smallest probability of error is half of the overlapping area, or integration of overlapping curves, which is 0.0222, verifying the result we calculated from measured or calculated gamma.

Work Cited

[1] wolfieswolfies 6. "Calculate Probability (Area) under the Overlapping Area of Two Normal Distributions." *Cross Validated*, 1 June 1964, stats.stackexchange.com/questions/103800/calculate-probability-area-under-the-overlapping-area-of-two-normal-distributi.

Appendix

All the code and original word document (contains all the images in it) are available in GitHub: https://github.com/steven202/EECE5644TakeHomeExam1.git