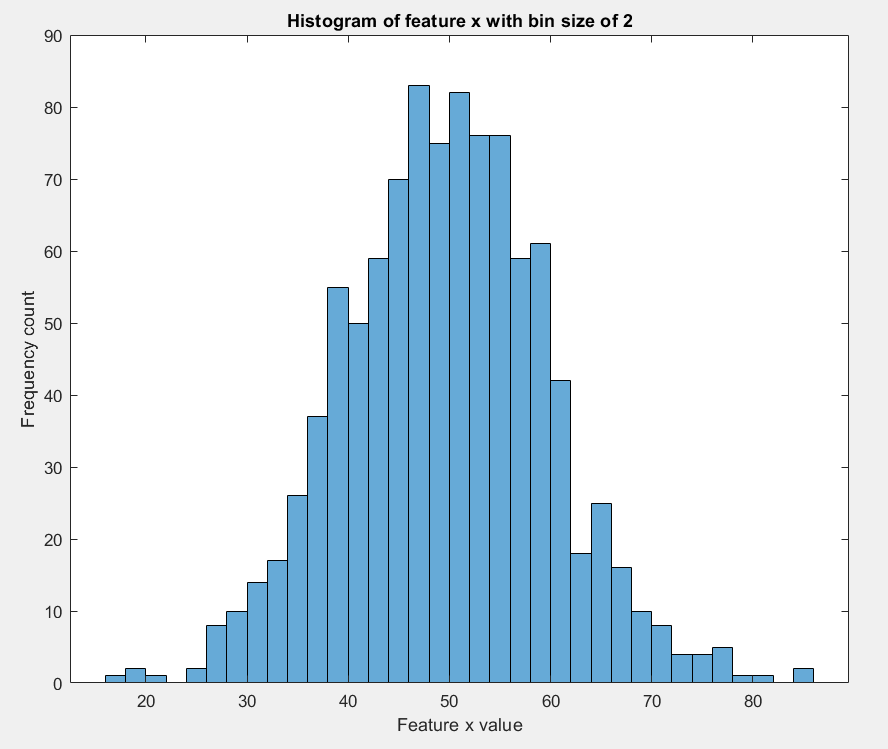
CSE 802 Homework 2

By Abhiram Durgaraju

Problem 1:

a)

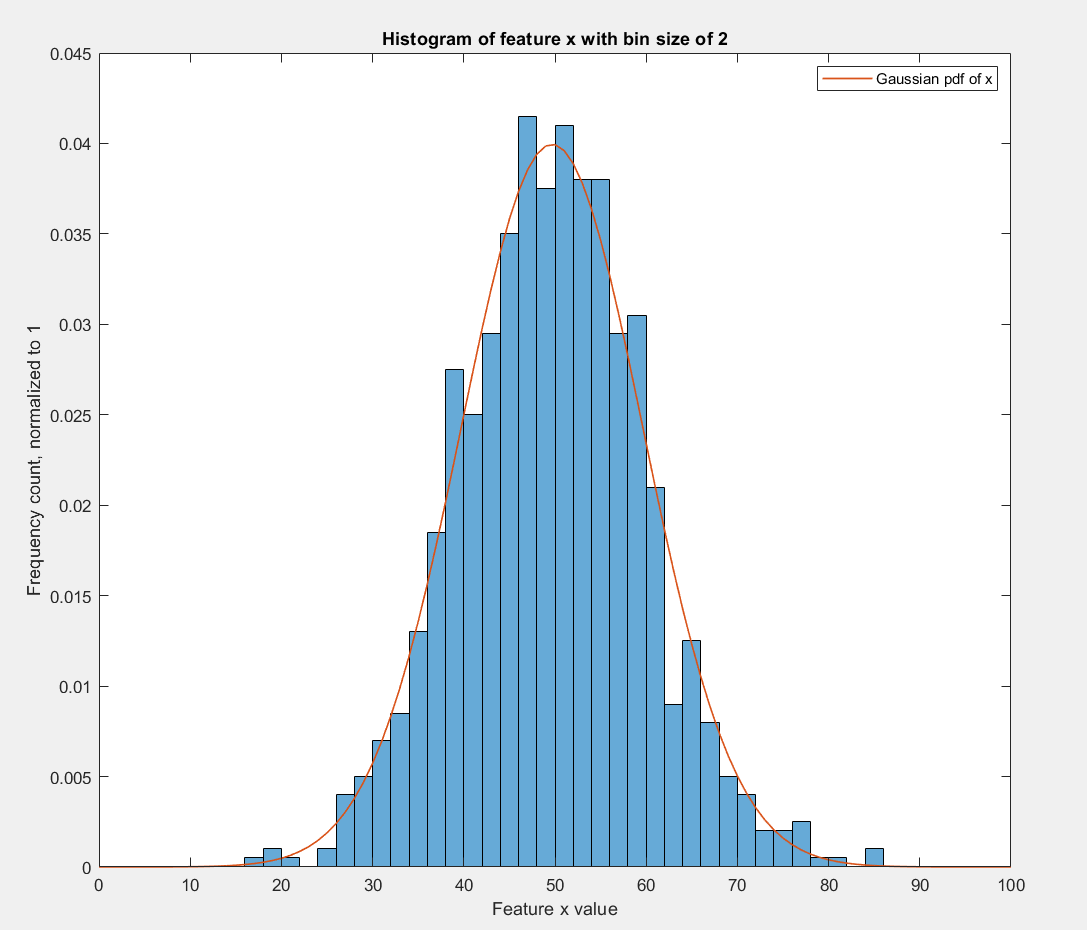


b) Used MATLAB to find mean and biased variance:

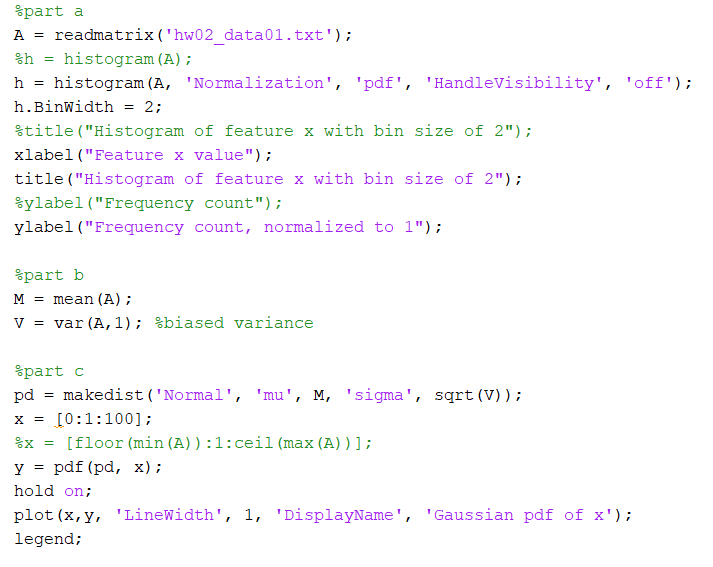
mean = 49.6737

biased variance = 99.6936

c)



Code used for problem 1:



Problem 2:

All computations were done in MATLAB.

1. Determinant of the covariance matrix = 21
2. Inverse of the covariance matrix:
3. Eigenvector of covariance matrix:

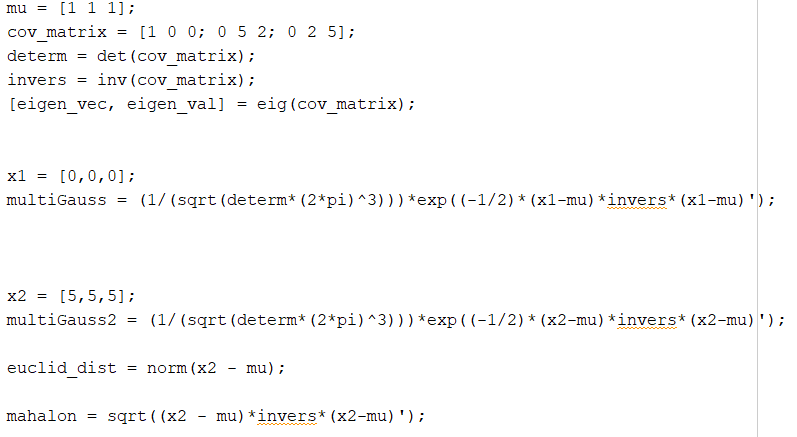
Eigenvalues of the covariance matrix:

1. Probability density at (0, 0, 0)t = 0.0073

Probability density at (5, 5, 5)t = 4.7271e-07

1. Euclidean distance between the mean and (5,5,5)t is 6.9282
2. Mahalonobis distance between the mean and (5,5,5)t is 4.5356

Code for problem 2:



Problem 3:

We have a problem where the covariance matrix of the two classes are the same but have two different means. This is exactly the same as “Case 2” in the textbook where

The decision boundary between two distributions can be given by where is the discriminant function associated with class 1 and is the discriminant function associated with class 2 .

For case 2, this can be simplified to the following expression (as shown in textbook, equation (61-63) on page 23): where is the decision boundary. The full expression for is shown below:

For expression simplicity let us write the above in terms of , which is a scalar:

Therefore, can be expressed as follows:

Notice that when the priors are equal:

Therefore,

It is clear to see that in this scenario, the boundary lies exactly halfway in between the two means.

For sufficiently disparate priors, the boundary will NOT lie in between the two means. This is seen in two situations: when and

Assume :

Situation 1:

When

Situation 2:

When

When we assume the opposite (), the same result occurs:

We substitute the original expression for to get the following expression in terms of the priors:

Problem 4:

To find the minimum overall risk of a given problem, Bayes tells us that we need to compute the conditional risk and select the action that gives the lowest risk. For a two-class problem, the decision boundary therefore is given by

Using equation (14) from the textbook on page 8, we know that:

We have assumed from the problem that . Therefore, the expression can be reduced to:

We also know from Bayes formula that:

This can be substituted back into the original expression:

We can cancel on both sides:

We are given that and . There we can substitute the Gaussian forms into the equation as follows:

We can cancel the term

We can take the natural logarithm of both sides to simplify the expression:

Therefore, the decision boundary (or )is given by:

Problem 5:

1. As before, the Bayes decision boundary can be calculated when the posterior probability of the two classes are equal:

Using Bayes formula:

Since the priors are equal and is on both sides, the expression can be simplified to:

The conditional density conforms to a Cauchy distribution, and so it follows that:

We can cancel the term and then simplify to get the following:

The decision boundary is therefore computed as

b)

The probability of misclassification can be expressed according to the textbook, page 6, as follows:

Given that we calculated the decision boundary to be , we can write our conditional error based on the decision policy as follows, according to textbook page 5:

We need to integrate over the region overlapped by each density. For that we need to make an assumption about the relativity of and

Let us assume of for this case that . Therefore we can describe as a sum of two regions with bounds from to and to as shown below:

Using Bayes formula we can subsite the posterior probability to cancel out the term:

Assuming that , we place the them outside of the integral along with

We can use u-substitution to simplify the integral, but this will change the bounds on the integrals

Let us and

Then,

The new bounds on the integrals are defined as:

So, we can represent the original integral as follows:

We can cancel the constant outside the integral and replace the integral with arctangent:

Since

Let

Notice that, the above equation is in the form:

Recall the following trigonometric relation:

We can simplify our expression as:

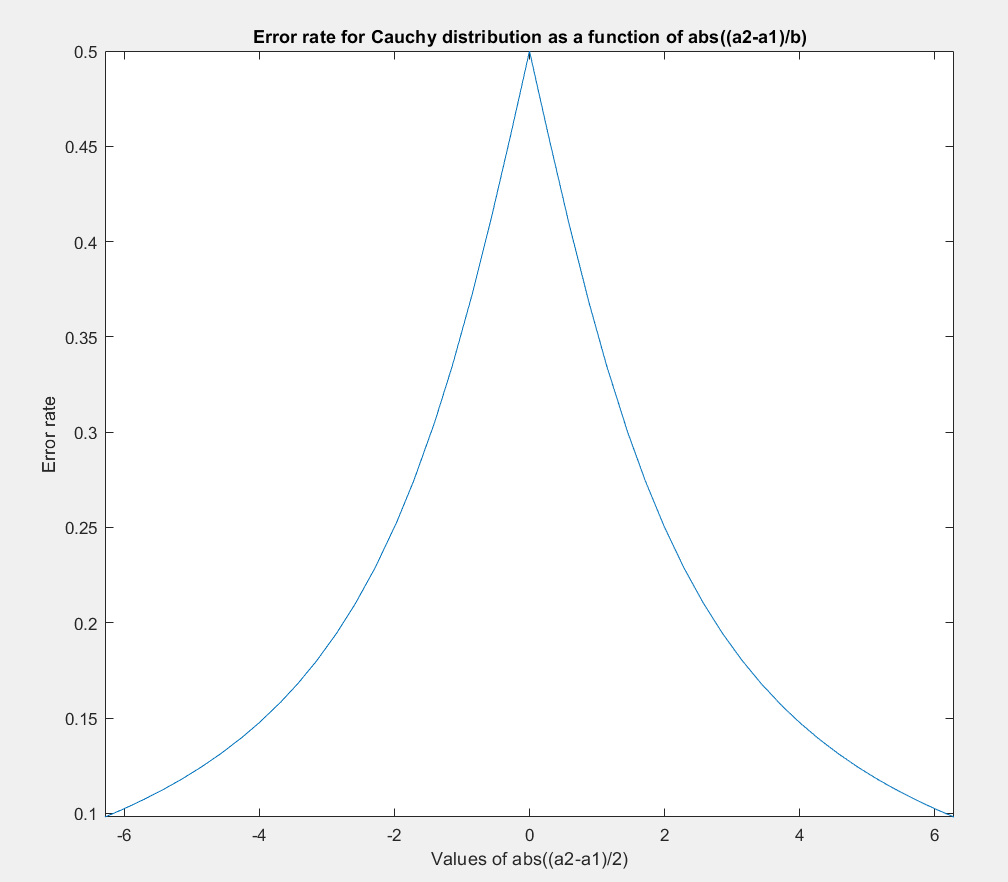
Resubstituting :

Remember originally we assumed that . If we assumed the opposite (, our integral bounds would swap. Therefore, our answer then would follow the same template:

Which is the same as:

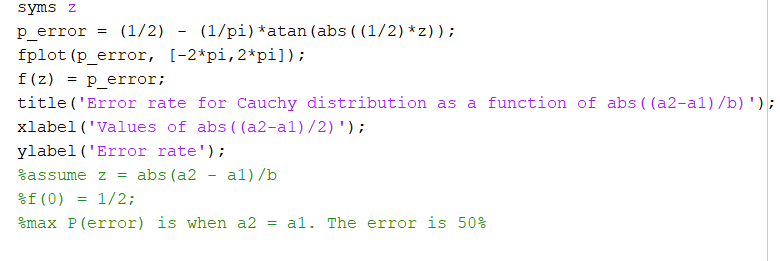
Therefore we can combine our findings from both assumptions to arrive at:

c)

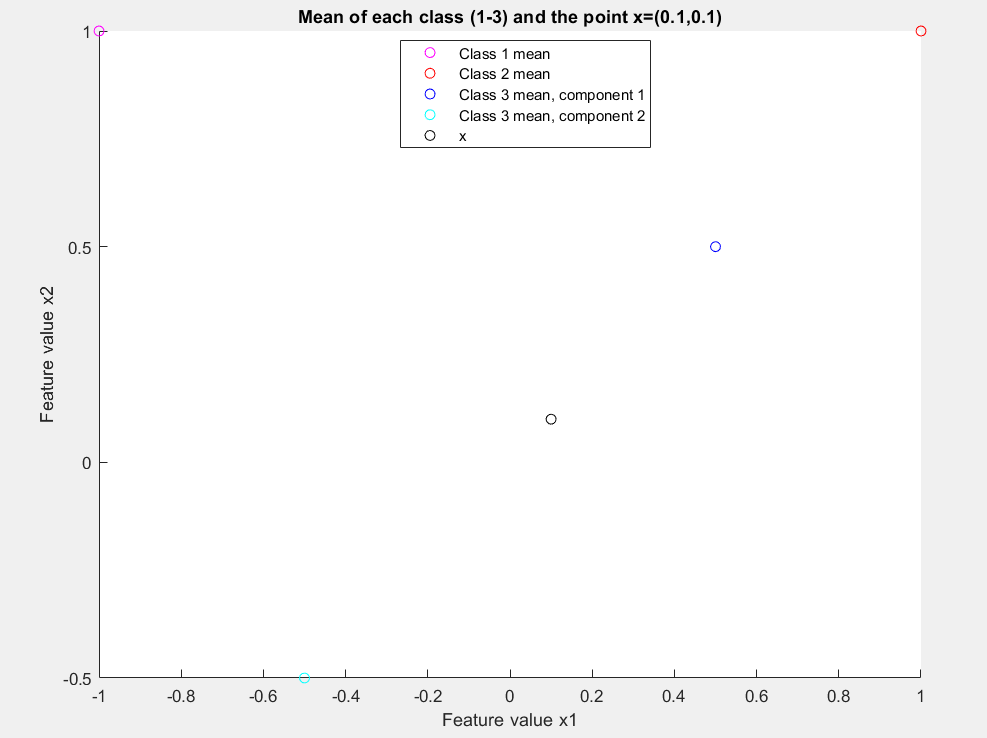


d) The maximum value of P(error) is , as seen clearly on the plot above. This will occur in the condition that , because . This makes sense intuitively because when the two values are equal, the classification is purely based on the priors. Since the priors are equal, the error must be 0.5

Code for problem 5:



Problem 6:



1. Bayes decision rule is as follows:

Decide if ;

Decide if ;

Otherwise, decide

Since the priors are equal and the is common to both terms, we can simply the above rule using Bayes formula to be:

Decide if ;

Decide if ;

Otherwise, decide

Using MATLAB, the conditional densities were calculated for (0.1,0.1) as follows:

Since was the highest of the 3 classes, we should assign to class 3.

Code for problem 6:



Problem 7:

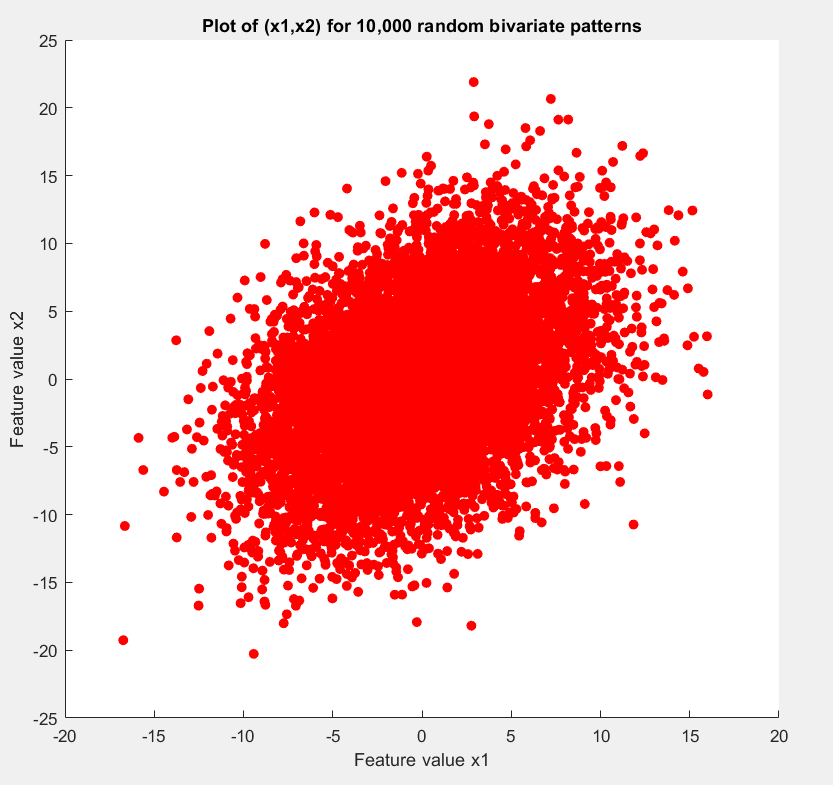
1. The white transform of is given by:

Where is a matrix of orthonormal eigenvectors and is the diagonal matrix of corresponding eigenvalues.

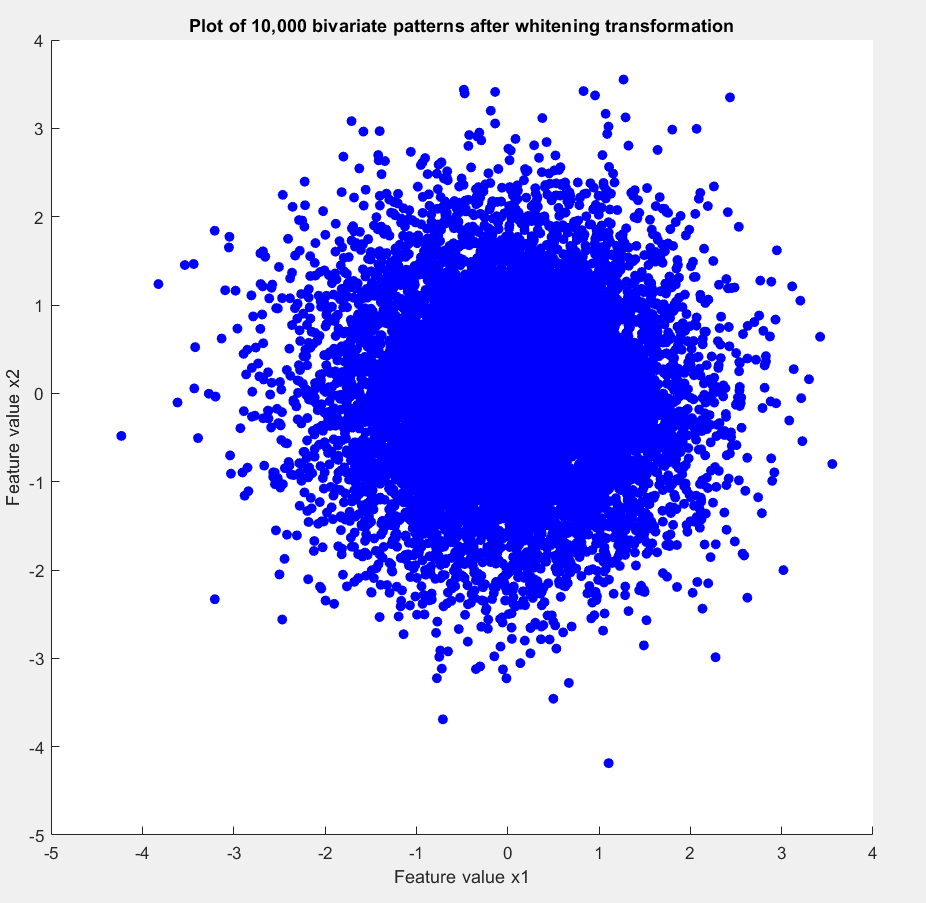
Using MATLAB, the whitening transform is given by the following:

1. The transformed density function is given by:

The mean remains unchanged. The covariance matrix is now identity matrix. Each individual pattern is multiplied by the transpose of the whitening transform.

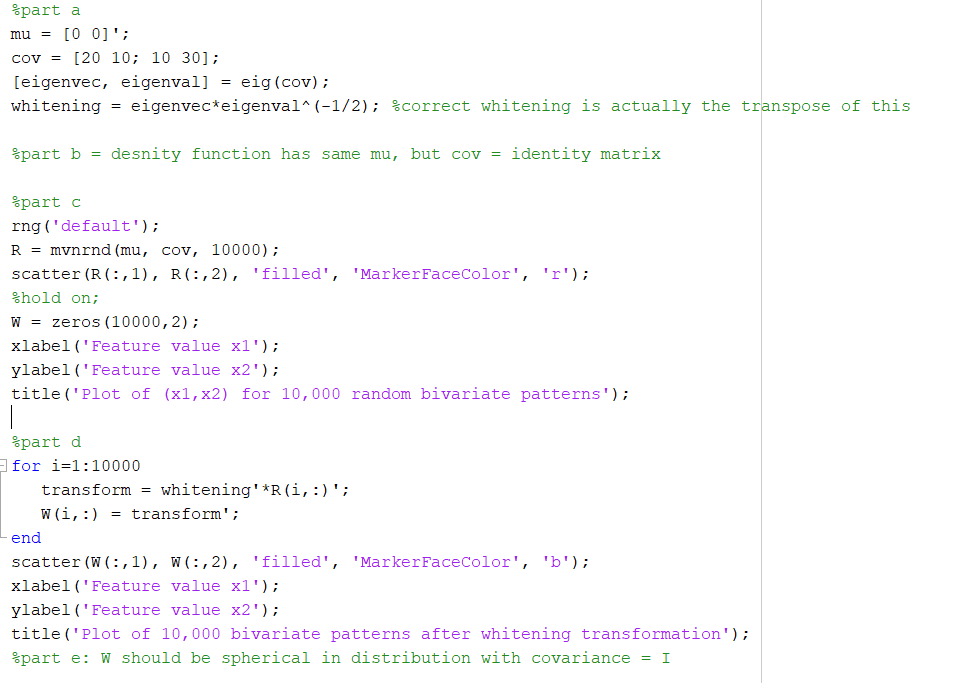


d)



1. The patterns in 7d are uniformly distributed around the center like a sphere, and the variance of each variable is equal. This is to be expected from since the covariance matrix is the identity matrix. The mean is still unchanged, at [0,0]. The patterns in 7c clearly show a skewed distribution since the features are not independent. As feature value x1 increases, the features of x2 also increases, resulting in the elongated shape shown in 7c. This is also to be expected since the off-diagonals in the covariance matrix are both equal to 10, indicating a positive relationship between the two features.

Code used for problem 7:



Problem 8:

a)

As before, the Bayes decision boundary can be calculated when the posterior probability of the two classes is equal:

Again, since we are assuming equal priors and is common to both sides, this can be reduced using Bayes formula to:

Using the textbook, the multivariate Gaussian is given by:

Set them equal to each other and cancel the common term

Take natural logarithm of both sides:

This can be simplified by multiplying the matrices:

Subtract 4 from both sides to get an equation for the circle:

**Our Bayes decision boundary is therefore a circle with radius centered at (4,4) and a radius of**

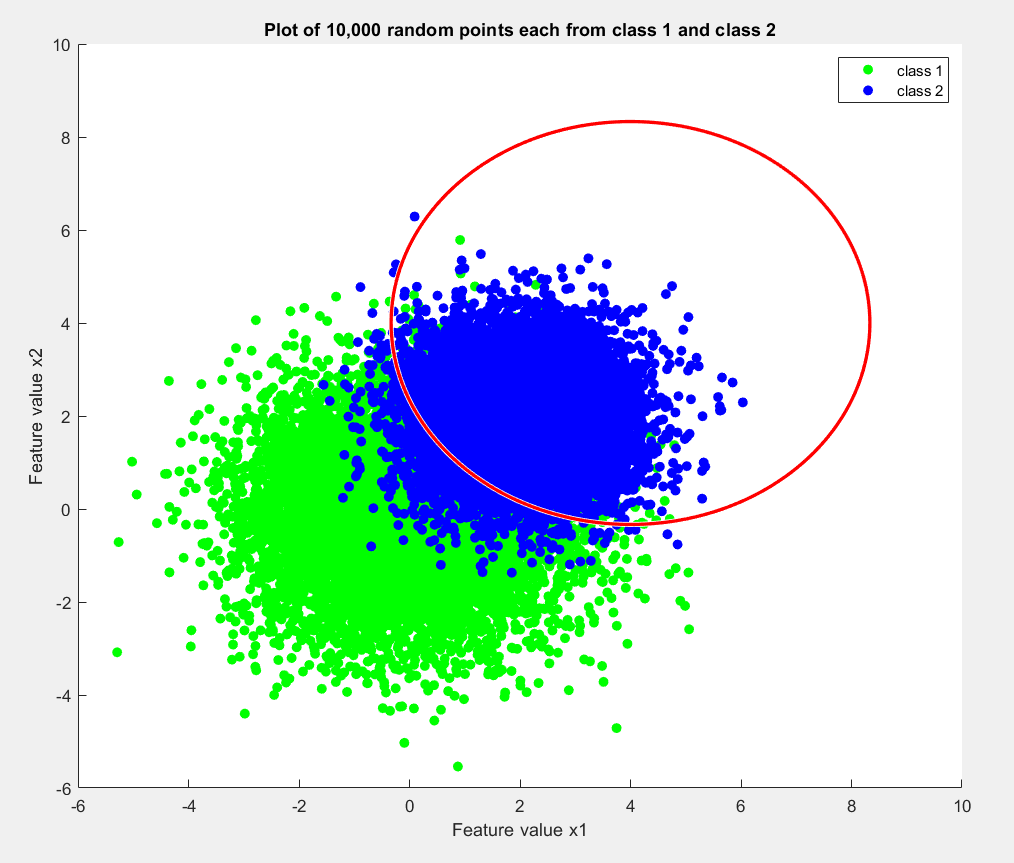
We can find the decision rule by first figuring out the ine quality for the decision boundary. We can take a sample point outside of the circle and compare the class-conditional densities. The point (0,0) lies outside the circle.

Since we can conclude that all patterns greater than or equal to the boundary will be classified as a class 1 by our classifier.

Our Bayes decision rule is therefore:

**Decide if; otherwise decide**

b)

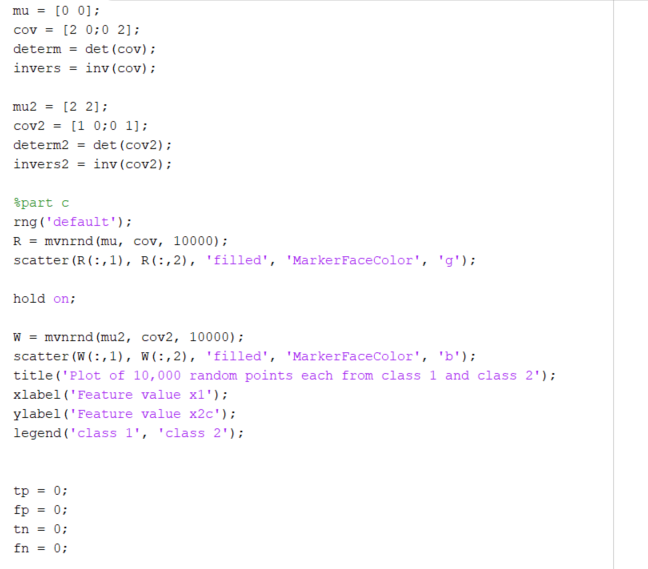


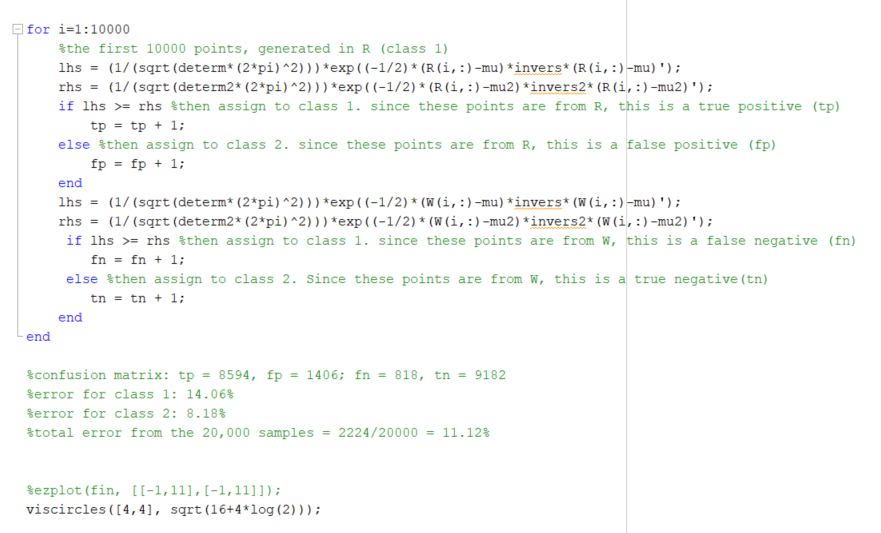
1. A total of 1406 patterns were misclassified as class 2 when they were truly class 1. A total of 818 patterns were misclassified as class 1 when they were truly class 2. The empirical error is therefore:

The confusion matrix is given as follows:

|  |  |  |
| --- | --- | --- |
|  | Predicted class 1 | Predicted class 2 |
| True class 1 | 8594 | 1406 |
| True class 2 | 818 | 9182 |

Code for problem 8:





Problem 9:

1. As before the Bayes decision boundary can be calculated when the posterior probabilities are equal.

Since the priors are equal and is common to both sides, we can again reduce the expression using Bayes formula to:

We can cancel out the common term and take the natural logarithm of both sides:

At this point, the left hand side is and right hand side is . To get the boundary we can continue the above expression. But, to get the decision rule, we need to keep careful track of the inequality. The normal decision rule is: Decide if . So , we can write this as:

Since we divide by on both side, the inequality flips.

**Therefore, our decision boundary is**

Our decision rule is:

**Decide if , otherwise decide**

1. From the textbook, page 31, we can see that the Chernoff bound is given by:

where

and

I used MATLAB to simplify the expression with the given means and covariance in our problem to:

To find the minimum beta, we need to first take the derivative. The above equation is in the form:

Therefore, the first derivative is:

Since cannot be 0, the only way for is if:

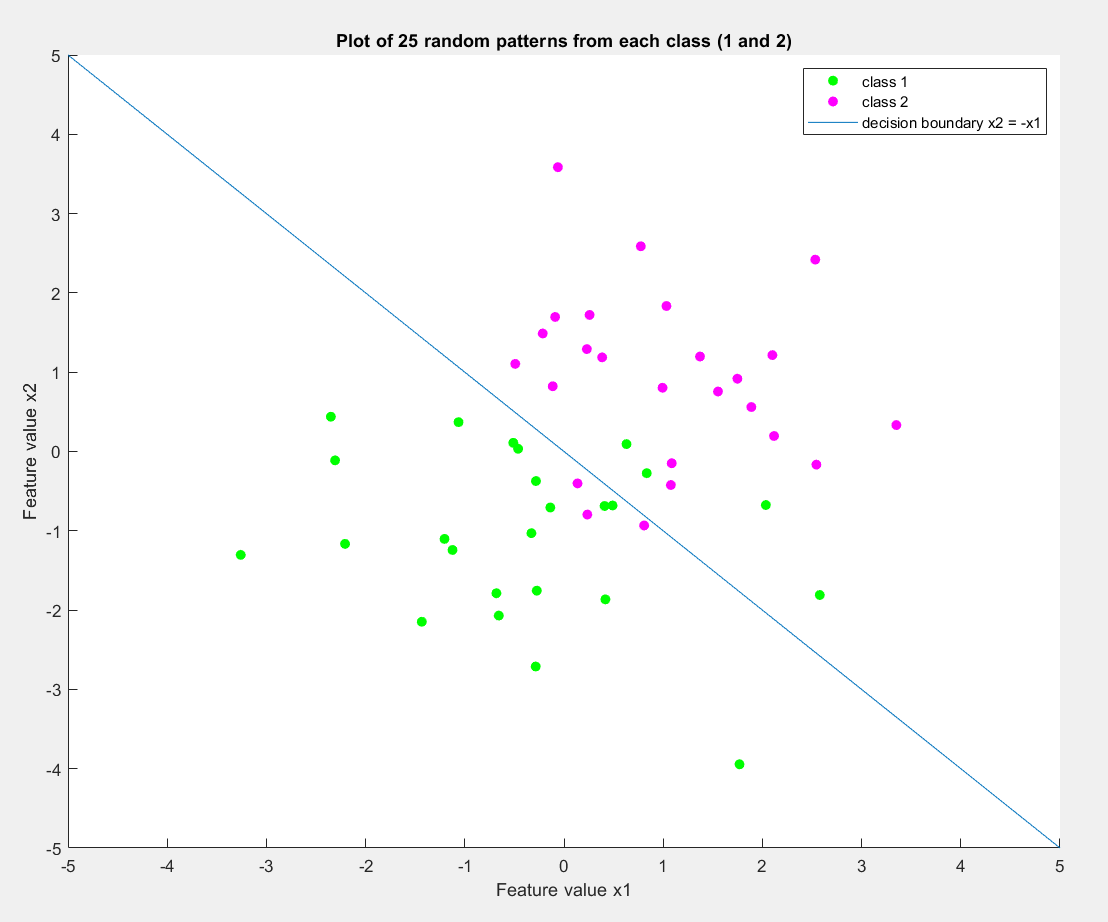
So, the original expression reduces to:

Chernoff bound is therefore:

The Bhattacharya bound can be calculating by simply assuming that

Therefore, the Bhattacharya bound is identical to the Chernoff bound in this case:

c)



d)

|  |  |  |
| --- | --- | --- |
|  | Predicted class 1 | Predicted class 2 |
| True class 1 | 21 | 4 |
| True class 2 | 3 | 22 |

Total misclassified = 4 + 3 = 7

Empirical error =

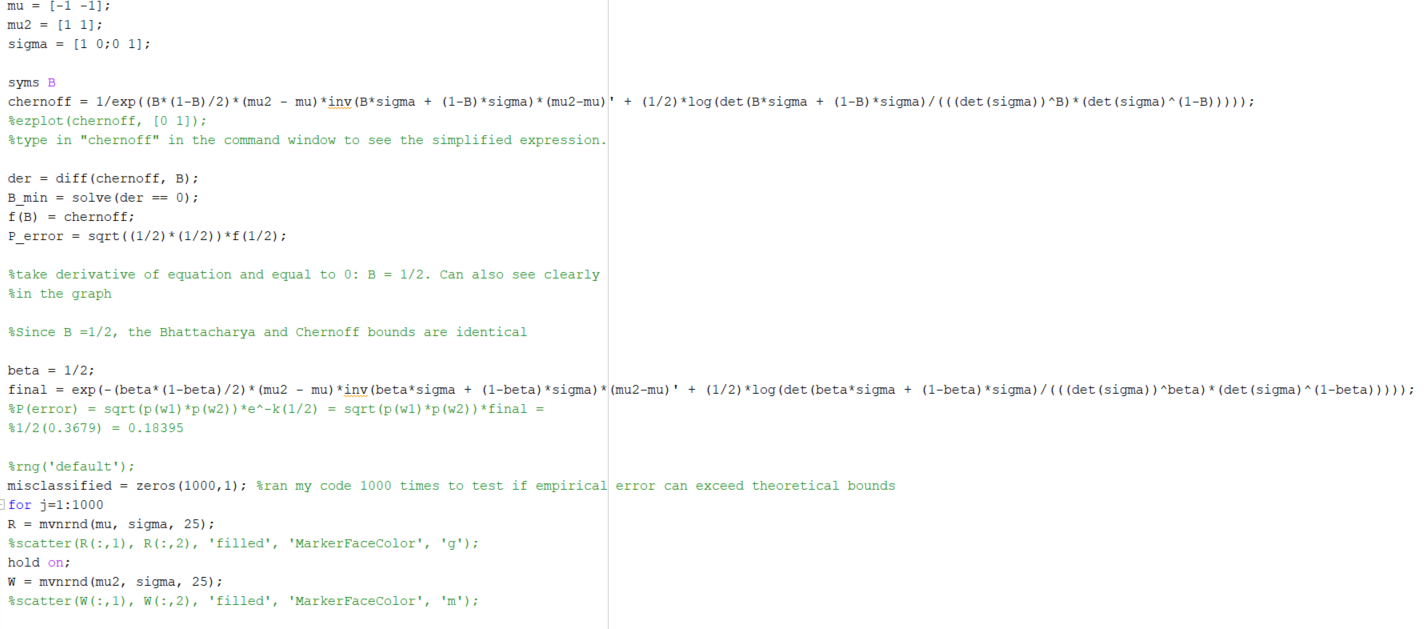
e)

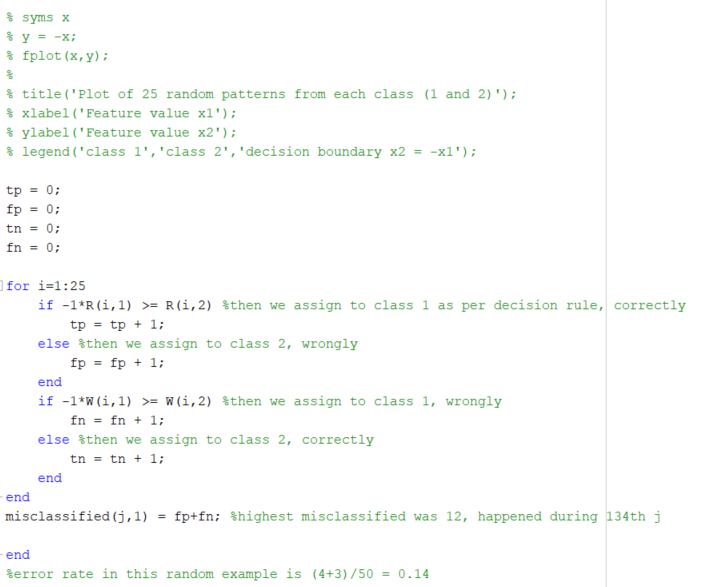
Since the error bound is 0.18395, we need to misclassify patterns to exceed the theoretical bounds for this problem. I ran my program without default rng to generate 50 patterns 1000 times using a loop. The most misclassifications were 11, which happened during the 659th execution of the code. This clearly exceeded the theoretical bound set by Chernoff and Bhattacharya, but it was quire rare. The bounds were only exceeded 4 times in 1000 executions.

The empirical error rate can exceed the theoretical bounds on probability misclassification because we are always dealing with a finite sample size. Especially on a sample size of 25 per class, it is possible to exceed the theoretical bounds because of potentially “bad” rng. The theoretical bounds assume we have access to infinite samples.

Summary: With sufficiently low sample size and sufficiently large repeated generation of patterns, it is possible to exceed theoretical bounds.

Code for problem 9:





Problem 10:

The Bayes minimum risk rule states that we must calculate the conditional risk for each possible action and pick the action with the lowest risk, denoted as

As before, the conditional risk is given by;

Since , the first term can be cancelled out. We can use Bayes formula to write posterior probability in terms of the likelihood and the priors. Since is simply a scaling factor and is common to all 3 actions, it will be ignored:

As before, since , the second term can be cancelled out. Bayes formula is used again:

We are given that x = 0.5, so:

It is given that , so:

The cost of and are identical. Since no information was given on resolution of ties such as in the decision policy, either of these two actions can be undertaken and are optimal for