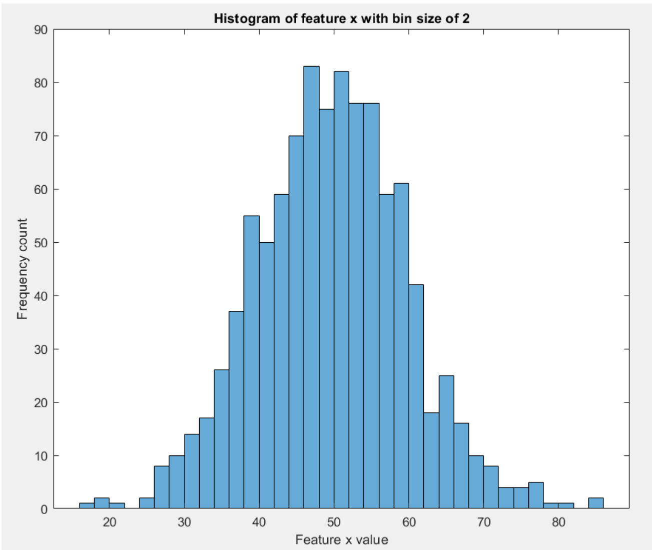
CSE 802 Homework 2

By Abhiram Durgaraju

Problem 1:

a)

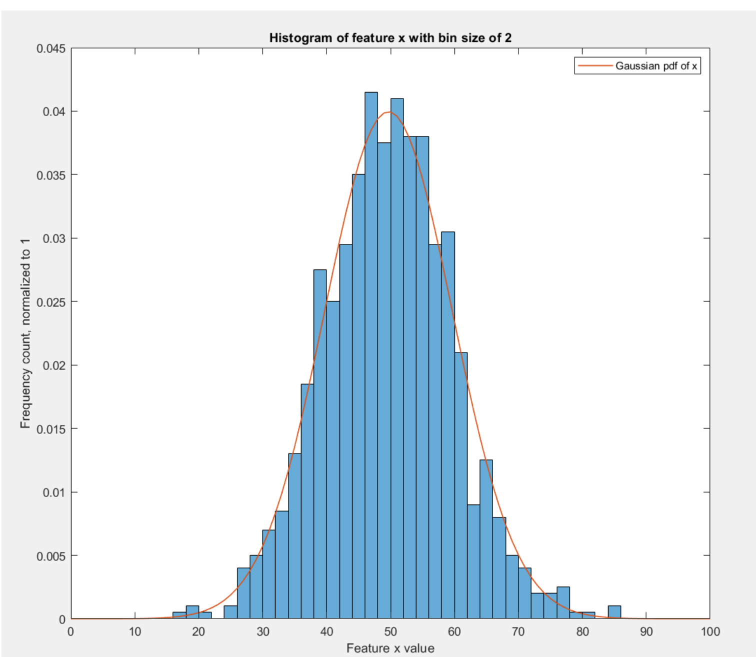


b) Used MATLAB to find mean and biased variance:

mean = 49.6737

biased variance = 99.6936

c)



Code used for problem 1:

```
%part a
A = readmatrix('hw02_data01.txt');
%h = histogram(A);
h = histogram(A, 'Normalization', 'pdf', 'HandleVisibility', 'off');
h.BinWidth = 2;
%title("Histogram of feature x with bin size of 2");
xlabel("Feature x value");
title("Histogram of feature x with bin size of 2");
%ylabel("Frequency count");
ylabel("Frequency count, normalized to 1");

%part b
M = mean(A);
V = var(A,1); %biased variance

%part c
pd = makedist('Normal', 'mu', M, 'sigma', sqrt(V));
x = [0:1:100];
%x = [floor(min(A)):1:ceil(max(A))];
y = pdf(pd, x);
hold on;
plot(x,y, 'LineWidth', 1, 'DisplayName', 'Gaussian pdf of x');
legend;
```

Problem 2:

All computations were done in MATLAB.

a) Determinant of the covariance matrix = 21
b) Inverse of the covariance matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.2381 & -0.0952 \\ 0 & -0.0952 & 0.2381 \end{bmatrix}$$

c) Eigenvector of covariance matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.7071 & 0.7071 \\ 0 & 0.7071 & 0.7071 \end{bmatrix}$$

Eigenvalues of the covariance matrix:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 7 \end{bmatrix}$$

d) Probability density at $(0, 0, 0)^t = 0.0073$
   Probability density at $(5, 5, 5)^t = 4.7271e\text{-}07$

e) Euclidean distance between the mean and $(5,5,5)^t$ is 6.9282
f) Mahalonobis distance between the mean and $(5,5,5)^t$ is 4.5356

Code for problem 2:

```
mu = [1 1 1];
cov_matrix = [1 0 0; 0 5 2; 0 2 5];
determ = det(cov_matrix);
invers = inv(cov_matrix);
[eigen_vec, eigen_val] = eig(cov_matrix);


x1 = [0,0,0];
multiGauss = (1/(sqrt(determ*(2*pi)^3)))*exp((-1/2)*(x1-mu)*invers*(x1-mu)');



x2 = [5,5,5];
multiGauss2 = (1/(sqrt(determ*(2*pi)^3)))*exp((-1/2)*(x2-mu)*invers*(x2-mu)');

euclid_dist = norm(x2 - mu);

mahalon = sqrt((x2 - mu)*invers*(x2-mu)');
```

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 $\Sigma_i = \Sigma$

The decision boundary between two distributions can be given by $g(x) = g_1(x) - g_2(x)$ where $g_1(x)$ is the discriminant function associated with class 1 (ω_1) and $g_2(x)$ is the discriminant function associated with class 2 (ω_1).

For case 2, this can be simplified to the following expression (as shown in textbook, equation (61-63) on page 23): $w^t(x - x_0) = 0$ where x_0 is the decision boundary. The full expression for x_0 is shown below:

$$x_0 = \frac{1}{2}(u_1 + u_2) - \frac{\ln[P(\omega_1) \,/\, P(\omega_2)]}{(u_1 - u_2)^t \Sigma^{-1} (u_1 - u_2)}(u_1 - u_2)$$

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

$$z = \frac{\ln[P(\omega_1) \,/\, P(\omega_2)]}{(u_1 - u_2)^t \Sigma^{-1} (u_1 - u_2)}$$

Therefore, x_0 can be expressed as follows:

$$x_0 = \frac{1}{2}(u_1 + u_2) - z(u_1 - u_2)$$

Notice that when the priors are equal:

$$\ln[p(\omega_1) \,/\, P(\omega_2)] = \ln(1) = 0$$

Therefore,

$$z = 0$$

$$x_0 = \frac{1}{2}(u_1 + u_2)$$

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 $x_0 > u_1$ and $x_0 < u_2$

Assume $u_1 > u_2$:

Situation 1:

When $x_0 > u_1$

$$\frac{1}{2}(u_1 + u_2) - z(u_1 - u_2) > u_1$$

$$-z(u_1 - u_2) > u_1 - \frac{1}{2}(u_1 + u_2)$$

$$-z(u_1 - u_2) > \frac{1}{2}(u_1 - u_2)$$

$$-z > \frac{1}{2}$$

$$z < -\frac{1}{2}$$

Situation 2:

When $x_0 < u_2$

$$\frac{1}{2}(u_1 + u_2) - z(u_1 - u_2) < u_2$$

$$-z(u_1 - u_2) < u_2 - \frac{1}{2}(u_1 + u_2)$$

$$-z(u_1 - u_2) < \frac{1}{2}(u_2 - u_1)$$

$$-z(u_1 - u_2) < -\frac{1}{2}(u_1 - u_2)$$

$$-z < -\frac{1}{2}$$

$$z > \frac{1}{2}$$

When we assume the opposite ($u_2 > u_1$), the same result occurs:

$$-\frac{1}{2} > z > \frac{1}{2}$$

$$|z| > \frac{1}{2}$$

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

$$|z| > \frac{1}{2}$$

$$\left|\frac{\ln[P(\omega_1) \,/\, P(\omega_2)]}{(u_1 - u_2)^t \Sigma^{-1}(u_1 - u_2)}\right| > \frac{1}{2}$$

Problem 4:

To find the minimum overall risk R of a given problem, Bayes tells us that we need to compute the conditional risk $R(\alpha_i|x)$ and select the action α_i that gives the lowest risk. For a two-class problem, the decision boundary therefore is given by $R(\alpha_1|x) = R(\alpha_2|x)$

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

$$R(\alpha_1|x) = \lambda_{11}P(\omega_1|x) + \lambda_{12}P(\omega_2|x)$$

$$R(\alpha_2|x) = \lambda_{21}P(\omega_1|x) + \lambda_{22}P(\omega_2|x)$$

$$R(\alpha_1|x) = R(\alpha_2|x)$$

$$\lambda_{11}P(\omega_1|x) + \lambda_{12}P(\omega_2|x) = \lambda_{21}P(\omega_1|x) + \lambda_{22}P(\omega_2|x)$$

We have assumed from the problem that $\lambda_{11} = \lambda_{22} = 0$. Therefore, the expression can be reduced to:

$$\lambda_{12}P(\omega_2|x) = \lambda_{21}P(\omega_1|x)$$

We also know from Bayes formula that:

$$P(\omega_j|x) = \frac{p(x|\omega_j)P(\omega_j)}{p(x)}$$

This can be substituted back into the original expression:

$$\lambda_{12}P(\omega_2|x) = \lambda_{21}P(\omega_1|x)$$

$$\lambda_{12} \frac{p(x|\omega_2)P(\omega_2)}{p(x)} = \lambda_{21} \frac{p(x|\omega_1)P(\omega_1)}{p(x)}$$

We can cancel $p(x)$ on both sides:

$$\lambda_{12} p(x|\omega_2)P(\omega_2) = \lambda_{21} p(x|\omega_1)P(\omega_1)$$

We are given that $p(x|\omega_1) \sim N(0, \sigma^2)$ and $p(x|\omega_2) \sim N(1, \sigma^2)$. There we can substitute the Gaussian forms into the equation as follows:

$$\lambda_{12} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{x-1}{\sigma}\right)^2\right]} \right) P(\omega_2) = \lambda_{21} \left(\frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right]} \right) P(\omega_1)$$

We can cancel the term $\frac{1}{\sigma\sqrt{2\pi}}$

$$\lambda_{12} e^{\left[-\frac{1}{2}\left(\frac{x-1}{\sigma}\right)^2\right]} P(\omega_2) = \lambda_{21} e^{\left[-\frac{1}{2}\left(\frac{x}{\sigma}\right)^2\right]} P(\omega_1)$$

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

$$\ln[\lambda_{12} P(\omega_2)] - \frac{1}{2}\left(\frac{x-1}{\sigma}\right)^2 = \ln[\lambda_{21} P(\omega_1)] - \frac{1}{2}\left(\frac{x}{\sigma}\right)^2$$

$$\ln[\lambda_{12} P(\omega_2)] - \ln[\lambda_{21} P(\omega_1)] = \frac{1}{2}\left(\frac{x-1}{\sigma}\right)^2 - \frac{1}{2}\left(\frac{x}{\sigma}\right)^2$$

$$\ln[\lambda_{12} P(\omega_2)] - \ln[\lambda_{21} P(\omega_1)] = \frac{x^2 - 2x + 1 - x^2}{2\sigma^2}$$

$$\ln\left[\frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}\right] = \frac{x^2 - 2x + 1 - x^2}{2\sigma^2}$$

$$\ln\left[\frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}\right] = \frac{-2x + 1}{2\sigma^2}$$

$$\ln\left[\frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}\right] = \frac{-x}{\sigma^2} + \frac{1}{2}$$

$$\ln\left[\frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}\right] - \frac{1}{2} = \frac{-x}{\sigma^2}$$

$$\sigma^2 \left(\ln\left[\frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}\right] - \frac{1}{2} \right) = -x$$

$$-\sigma^2 \left(\ln\left[\frac{\lambda_{12} P(\omega_2)}{\lambda_{21} P(\omega_1)}\right] - \frac{1}{2} \right) = x$$

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

$$\tau = \frac{1}{2} - \sigma^2 \ln\left[\frac{\lambda_{12}P(\omega_2)}{\lambda_{21}P(\omega_1)}\right]$$

Problem 5:

a) As before, the Bayes decision boundary x can be calculated when the posterior probability of the two classes are equal:

$$P(\omega_1|x) = P(\omega_2|x)$$

Using Bayes formula:

$$\frac{p(x|\omega_1)P(\omega_1)}{p(x)} = \frac{p(x|\omega_2)P(\omega_2)}{p(x)}$$

Since the priors are equal and $p(x)$ is on both sides, the expression can be simplified to:

$$p(x|\omega_1) = p(x|\omega_2)$$

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

$$\frac{1}{\pi b}\left(\frac{1}{1+\left(\frac{x-a_1}{b}\right)^2}\right) = \frac{1}{\pi b}\left(\frac{1}{1+\left(\frac{x-a_2}{b}\right)^2}\right)$$

We can cancel the $\frac{1}{\pi b}$ term and then simplify to get the following:

$$1+\left(\frac{x-a_1}{b}\right)^2 = 1+\left(\frac{x-a_2}{b}\right)^2$$

$$1+\frac{x^2-2xa_1+{a_1}^2}{b} = 1+\frac{x^2-2xa_2+{a_2}^2}{b}$$

$$x^2-2xa_1+{a_1}^2 = x^2-2xa_2+{a_2}^2$$

$$-2xa_1+{a_1}^2 = -2xa_2+{a_2}^2$$

$$2xa_2-2xa_1 = {a_2}^2-{a_1}^2$$

$$2x(a_2-a_1) = {a_2}^2-{a_1}^2$$

$$2x(a_2-a_1) = (a_2-a_1)(a_2+a_1)$$

$$2x = (a_2+a_1)$$

$$x = \frac{a_2+a_1}{2}$$

The decision boundary is therefore computed as

$$x = \frac{a_2+a_1}{2}$$

b)

The probability of misclassification $P(error)$ can be expressed according to the textbook, page 6, as follows:

$$P(error) = \int_{-\infty}^{\infty} P(error, x)dx = \int_{-\infty}^{\infty} P(error|x)p(x)dx$$

Given that we calculated the decision boundary to be $\frac{a2+a1}{2}$, we can write our conditional error based on the decision policy as follows, according to textbook page 5:

$$P(error|x) = \begin{Bmatrix} P(\omega_1|x) \;\; if \;\; x \geq \dfrac{a_2 + a_1}{2} \\ P(\omega_2|x) \;\; if \;\; x < \dfrac{a_2 + a_1}{2} \end{Bmatrix}$$

We need to integrate over the region overlapped by each density. For that we need to make an assumption about the relativity of a_2 and a_1

Let us assume of for this case that $a_2 > a_1$. Therefore we can describe $P(error)$ as a sum of two regions with bounds from $-\infty$ to $\frac{a_2+a_1}{2}$ and $\frac{a_2+a_1}{2}$ to ∞ as shown below:

$$P(error) = \int_{-\infty}^{\frac{a_2+a_1}{2}} P(\omega_2|x)p(x)dx + \int_{\frac{a_2+a_1}{2}}^{\infty} P(\omega_1|x)p(x)dx$$

Using Bayes formula we can subsite the posterior probability to cancel out the $p(x)$ term:

$$P(error) = \int_{-\infty}^{\frac{a_2+a_1}{2}} p(x|\omega_2)P(\omega_2)dx + \int_{\frac{a_2+a_1}{2}}^{\infty} p(x|\omega_1)P(\omega_1)dx$$

Assuming that $P(\omega_1) = P(\omega_2) = 0.5$, we place the them outside of the integral along with $\frac{1}{\pi b}$

$$P(error) = \frac{1}{2\pi b}\left(\int_{-\infty}^{\frac{a_2+a_1}{2}} \frac{1}{1+\left(\frac{x-a_2}{b}\right)^2}dx + \int_{\frac{a_2+a_1}{2}}^{\infty} \frac{1}{1+\left(\frac{x-a_1}{b}\right)^2}dx\right)$$

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

Let us $u = \frac{x-a_2}{b}$ and $v = \frac{x-a_1}{b}$

Then, $du = dv = \frac{1}{b}dx$

$$dx = b(du) = b(dv)$$

The new bounds on the integrals are defined as:

$$u\left(\frac{a_2 + a_1}{2}\right) = \frac{\frac{a_2 + a_1}{2} - a_2}{b} = \frac{\frac{a_2 + a_1 - 2a_2}{2}}{b} = \frac{a_1 - a_2}{2b}$$

$$v\left(\frac{a_2 + a_1}{2}\right) = \frac{\frac{a_2 + a_1}{2} - a_1}{b} = \frac{\frac{a_2 + a_1 - 2a_1}{2}}{b} = \frac{a_2 - a_1}{2b}$$

So, we can represent the original integral as follows:

$$P(error) = \frac{1}{2\pi b}\left(\int_{-\infty}^{\frac{a_1-a_2}{2b}} \frac{1}{1+(u)^2}(b)du + \int_{\frac{a_2-a_1}{2b}}^{\infty} \frac{1}{1+(v)^2}(b)dv\right)$$

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

$$P(error) = \frac{1}{2\pi}\left(\int_{-\infty}^{\frac{a_1-a_2}{2b}} \frac{1}{1+(u)^2}du + \int_{\frac{a_2-a_1}{2b}}^{\infty} \frac{1}{1+(v)^2}dv\right)$$

$$P(error) = \frac{1}{2\pi}\left(tan^{-1}(v)\Big|_{-\infty}^{\frac{a_1 - a_2}{2b}} + tan^{-1}(v)\Big|_{\frac{a_2 - a_1}{2b}}^{\infty}\right)$$

Since $\lim_{x\to\infty} tan^{-1}(x) = \frac{\pi}{2}$

$$P(error) = \frac{1}{2\pi}\left(tan^{-1}\left(\frac{a_1 - a_2}{2b}\right) + \frac{\pi}{2} + \frac{\pi}{2} - tan^{-1}\left(\frac{a_2 - a_1}{2b}\right)\right)$$

$$P(error) = \frac{1}{2} + \frac{1}{2\pi}\left(tan^{-1}\left(\frac{a_1 - a_2}{2b}\right) - tan^{-1}\left(\frac{a_2 - a_1}{2b}\right)\right)$$

Let $m = \frac{a_2 - a_1}{2b}$

Notice that, the above equation is in the form:

$$P(error) = \frac{1}{2} + \frac{1}{2\pi}\left(tan^{-1}(-m) - tan^{-1}(m)\right)$$

Recall the following trigonometric relation:

$$tan^{-1}(-m) = -tan^{-1}(m)$$

We can simplify our expression as:

$$P(error) = \frac{1}{2} + \frac{1}{2\pi}\left(-tan^{-1}(m) - tan^{-1}(m)\right)$$

$$P(error) = \frac{1}{2} + \frac{1}{2\pi}\left(-2tan^{-1}(m)\right)$$

$$P(error) = \frac{1}{2} - \frac{1}{\pi}\left(tan^{-1}(m)\right)$$

Resubstituting m:

$$P(error) = \frac{1}{2} - \frac{1}{\pi}\left(tan^{-1}\left(\frac{a_2 - a_1}{2b}\right)\right)$$

Remember originally we assumed that $a_2 > a_1$. If we assumed the opposite ($a_1 > a_2$), our integral bounds would swap. Therefore, our answer then would follow the same template:

$$P(error) = \frac{1}{2} - \frac{1}{\pi}\left(tan^{-1}\left(\frac{a_1 - a_2}{2b}\right)\right)$$

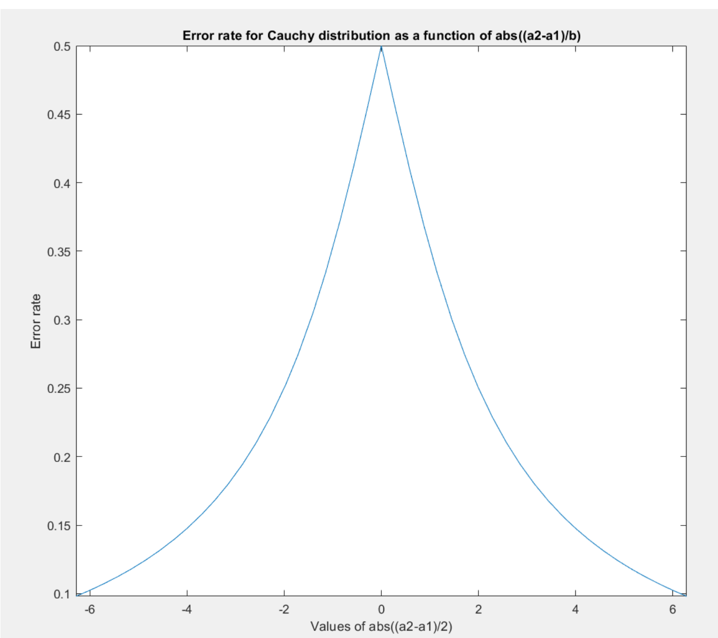
Which is the same as:

$$P(error) = \frac{1}{2} - \frac{1}{\pi}\left(tan^{-1}\left(\frac{-(a_2 - a_1)}{2b}\right)\right)$$

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

$$P(error) = \frac{1}{2} - \frac{1}{\pi}\left(tan^{-1}\left|\frac{a_2 - a_1}{2b}\right|\right)$$

c)



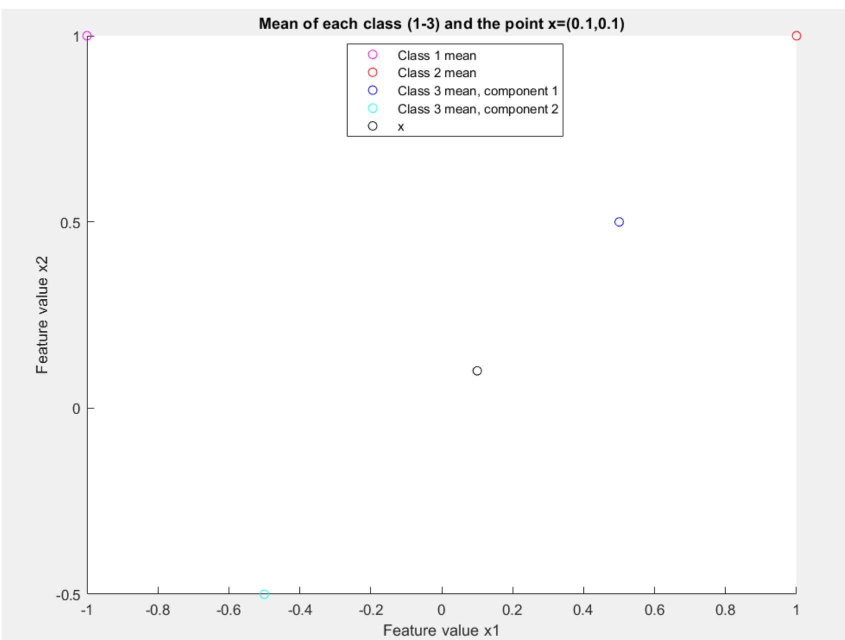
d) The maximum value of P(error) is $\frac{1}{2}$, as seen clearly on the plot above. This will occur in the condition that $a_2 = a_1$, because $\arctan(0) = 0$. This makes sense intuitively because when the two values are equal, the classification is purely based on the priors. Since the priors are equal, the error must be 0.5

Code for problem 5:

```
syms z
p_error = (1/2) - (1/pi)*atan(abs((1/2)*z));
fplot(p_error, [-2*pi,2*pi]);
f(z) = p_error;
title('Error rate for Cauchy distribution as a function of abs((a2-a1)/b)');
xlabel('Values of abs((a2-a1)/2)');
ylabel('Error rate');
%assume z = abs(a2 - a1)/b
%f(0) = 1/2;
%max P(error) is when a2 = a1. The error is 50%
```

Problem 6:

a)



b) Bayes decision rule is as follows:
Decide ω_1 if max $(P(\omega_1|x), P(\omega_2|x), P(\omega_3|x)) = P(\omega_1|x)$;
Decide ω_2 if max $(P(\omega_1|x), P(\omega_2|x), P(\omega_3|x)) = P(\omega_2|x)$;
Otherwise, decide ω_3

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

Decide ω_1 if $\max\left(p(x|\omega_1), p(x|\omega_2), p(x|\omega_3)\right) = p(x|\omega_1)$;

Decide ω_2 if $\max\left(p(x|\omega_1), p(x|\omega_2), p(x|\omega_3)\right) = p(x|\omega_2)$;

Otherwise, decide ω_3

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

$$p(x|\omega_1) = 0.0475$$
$$p(x|\omega_2) = 0.0708$$
$$p(x|\omega_3) = \frac{1}{2}(0.1356) + \frac{1}{2}(0.1110) = 0.1233$$

Since $p(x|\omega_3)$ was the highest of the 3 classes, we should assign to class 3.

Code for problem 6:

```
hold on;
plot(-1,1, 'o', 'MarkerEdgeColor', 'm', 'DisplayName', 'Class 1 mean'); %class w1
plot(1,1, 'o', 'MarkerEdgeColor', 'r', 'DisplayName', 'Class 2 mean'); %class w2
plot(0.5,0.5,'o', 'MarkerEdgeColor', 'b', 'DisplayName', 'Class 3 mean, component 1'); %,GMM 1st part
plot(-0.5,-0.5,'o', 'MarkerEdgeColor', 'c', 'DisplayName', 'Class 3 mean, component 2'); %GMM 2nd part
plot(0.1,0.1, 'o', 'MarkerEdgeColor', 'k', 'DisplayName', 'x'); %x
xlabel('Feature value x1');
ylabel('Feature value x2');
title('Mean of each class (1-3) and the point x=(0.1,0.1)');
legend;

mu = [-1,-1];
cov = [1 0;0 1];
determ = det(cov);
invers = inv(cov);
x1 = [0.1, 0.1];
multiGauss = (1/(sqrt(determ*(2*pi)^2)))*exp((-1/2)*(x1-mu)*invers*(x1-mu)'); %class 1 desnity

mu = [1,1];
multiGauss2 = (1/(sqrt(determ*(2*pi)^2)))*exp((-1/2)*(x1-mu)*invers*(x1-mu)'); %class 2 density

mu = [0.5,0.5];
multiGauss3_1 = (1/(sqrt(determ*(2*pi)^2)))*exp((-1/2)*(x1-mu)*invers*(x1-mu)'); %GMM 1st component

mu = [-0.5,-0.5];
multiGauss3_2 = (1/(sqrt(determ*(2*pi)^2)))*exp((-1/2)*(x1-mu)*invers*(x1-mu)'); %GMM 2nt component

multiGauss3 = 0.5*multiGauss3_1 + 0.5*multiGauss3_2; %Total GMM has the highest probability, so we assign to class 3
```

Problem 7:

a) The white transform A_w of x is given by:

$$A_w = \Phi \Lambda^{-1/2}$$

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

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

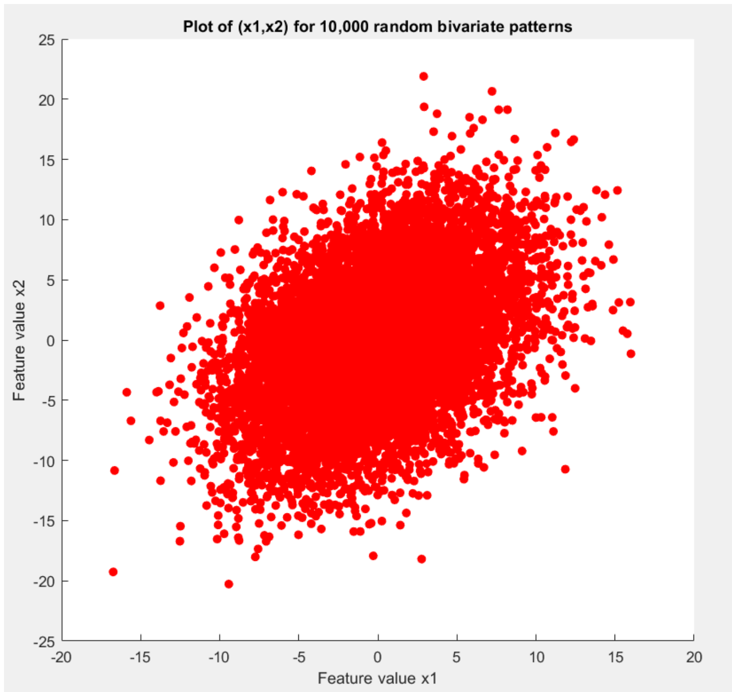
$$A_w = \begin{bmatrix} -0.2288 & 0.0874 \\ 0.1414 & 0.1414 \end{bmatrix}$$

b) The transformed density function is given by:

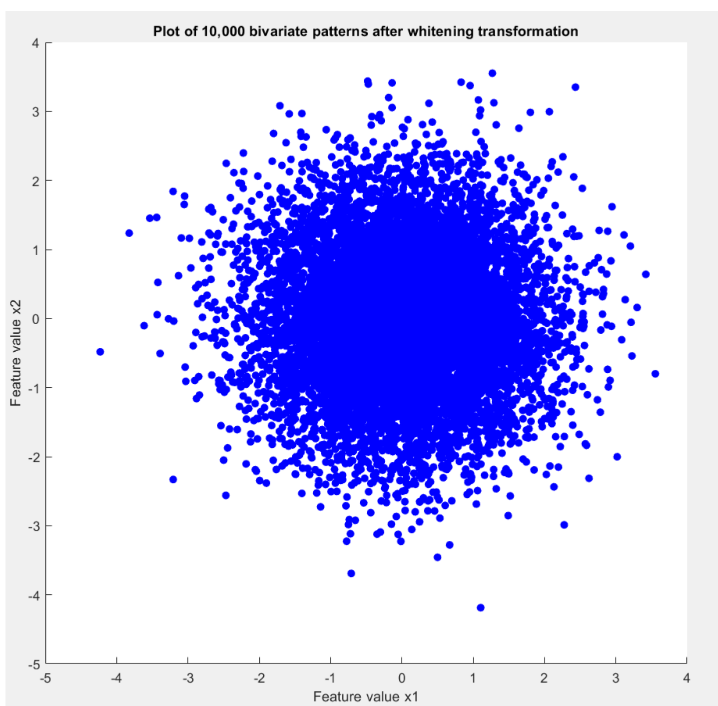
$$P_{transform}(x) \sim N\left(\begin{bmatrix} -0.2288 & 0.0874 \\ 0.1414 & 0.1414 \end{bmatrix}^t \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$$

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

c)



d)



d) 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:

```
%part a
mu = [0 0]';
cov = [20 10; 10 30];
[eigenvec, eigenval] = eig(cov);
whitening = eigenvec*eigenval^(-1/2); %correct whitening is actually the transpose of this

%part b = desnity function has same mu, but cov = identity matrix

%part c
rng('default');
R = mvnrnd(mu, cov, 10000);
scatter(R(:,1), R(:,2), 'filled', 'MarkerFaceColor', 'r');
%hold on;
W = zeros(10000,2);
xlabel('Feature value x1');
ylabel('Feature value x2');
title('Plot of (x1,x2) for 10,000 random bivariate patterns');

%part d
for i=1:10000
    transform = whitening'*R(i,:)';
    W(i,:) = transform';
end
scatter(W(:,1), W(:,2), 'filled', 'MarkerFaceColor', 'b');
xlabel('Feature value x1');
ylabel('Feature value x2');
title('Plot of 10,000 bivariate patterns after whitening transformation');
%part e: W should be spherical in distribution with covariance = I
```

Problem 8:

a)

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

$$P(\omega_1|x) = P(\omega_2|x)$$

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

$$p(x|\omega_1) = p(x|\omega_2)$$

Using the textbook, the multivariate Gaussian is given by:

$$p(x|\omega) = \frac{1}{2\pi^{d/2}|\Sigma|^{1/2}} e^{\left(-\frac{1}{2}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-u\right)^t \Sigma^{-1}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-u\right)\right)}$$

$$p(x|\omega_1) = \frac{1}{(2\pi)^{2/2}(4)^{1/2}} e^{\left(-\frac{1}{2}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}0\\0\end{bmatrix}\right)^t \begin{bmatrix}1/2 & 0\\0 & 1/2\end{bmatrix}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}0\\0\end{bmatrix}\right)\right)}$$

$$p(x|\omega_2) = \frac{1}{(2\pi)^{2/2}(1)^{1/2}} e^{\left(-\frac{1}{2}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}2\\2\end{bmatrix}\right)^t \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}2\\2\end{bmatrix}\right)\right)}$$

Set them equal to each other and cancel the common term $\frac{1}{(2\pi)^{2/2}}$

$$\frac{1}{2} e^{\left(-\frac{1}{2}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}0\\0\end{bmatrix}\right)^t \begin{bmatrix}1/2 & 0\\0 & 1/2\end{bmatrix}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}0\\0\end{bmatrix}\right)\right)} = e^{\left(-\frac{1}{2}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}2\\2\end{bmatrix}\right)^t \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}2\\2\end{bmatrix}\right)\right)}$$

Take natural logarithm of both sides:

$$\ln\left(\frac{1}{2}\right) - \frac{1}{2}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}0\\0\end{bmatrix}\right)^t \begin{bmatrix}1/2 & 0\\0 & 1/2\end{bmatrix}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}0\\0\end{bmatrix}\right) = -\frac{1}{2}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}2\\2\end{bmatrix}\right)^t \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-\begin{bmatrix}2\\2\end{bmatrix}\right)$$

This can be simplified by multiplying the matrices:

$$\ln\left(\frac{1}{2}\right) - \frac{1}{2}\left(\frac{1}{2}x_1{}^2 + \frac{1}{2}x_2{}^2\right) = -\frac{1}{2}(x_1{}^2 - 4x_1 + 4 + x_2{}^2 - 4x_2 + 4)$$

$$\ln\left(\frac{1}{2}\right) - \frac{1}{4}x_1{}^2 - \frac{1}{4}x_2{}^2 = -\frac{1}{2}x_1{}^2 + 2x_1 - 2 - \frac{1}{2}x_2{}^2 + 2x_2 - 2$$

$$\ln\left(\frac{1}{2}\right) = -\frac{1}{4}x_1{}^2 + 2x_1 - \frac{1}{4}x_2{}^2 + 2x_2 - 4$$

$$\ln\left(\frac{1}{2}\right) = -\frac{1}{4}({x_1}^2 - 8x_1 + 8) - \frac{1}{4}({x_2}^2 - 8x_2 + 8)$$

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

$$\ln\left(\frac{1}{2}\right) - 4 = -\frac{1}{4}({x_1}^2 - 8x_1 + 16) - \frac{1}{4}({x_2}^2 - 8x_2 + 16)$$

$$-4\left(\ln\left(\frac{1}{2}\right) - 4\right) = ({x_1}^2 - 8x_1 + 16) - ({x_2}^2 - 8x_2 + 16)$$

$$16 + 4\ln(2) = ({x_1}^2 - 8x_1 + 16) - ({x_2}^2 - 8x_2 + 16)$$

$$16 + 4\ln(2) = (x_1 - 4)^2 - (x_2 - 4)^2$$

Our Bayes decision boundary is therefore a circle with radius centered at (4,4) and a radius of $\boldsymbol{\sqrt{16 + 4\ln(2)}}$

We can find the decision rule by first figuring out the inequality 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.

$$p\left(\begin{bmatrix}0\\0\end{bmatrix}\middle|\,\omega_1\right) = \frac{1}{(2\pi)^{2/2}(4)^{1/2}} e^{\left(-\frac{1}{2}\left(\begin{bmatrix}0\\0\end{bmatrix}-\begin{bmatrix}0\\0\end{bmatrix}\right)^t \begin{bmatrix}1/2 & 0\\ 0 & 1/2\end{bmatrix}\left(\begin{bmatrix}0\\0\end{bmatrix}-\begin{bmatrix}0\\0\end{bmatrix}\right)\right)} = \frac{1}{(2\pi)^{2/2}(4)^{1/2}}$$

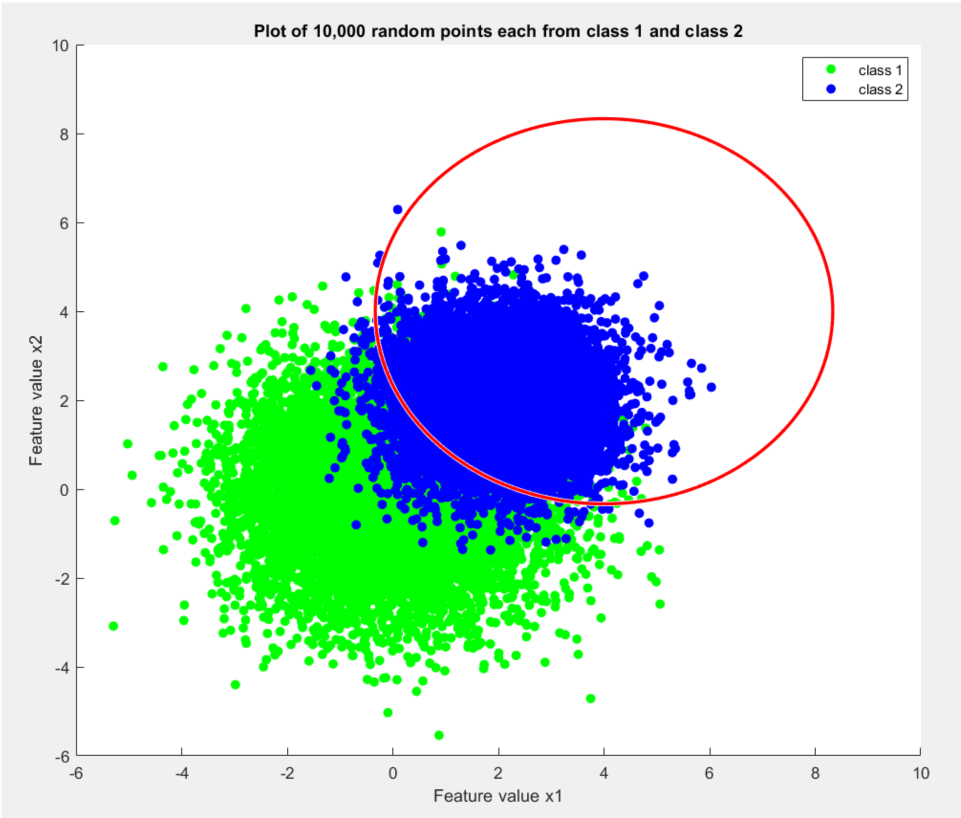
$$p\left(\begin{bmatrix}0\\0\end{bmatrix}\middle|\,\omega_2\right) = \frac{1}{(2\pi)^{2/2}(1)^{1/2}} e^{\left(-\frac{1}{2}\left(\begin{bmatrix}0\\0\end{bmatrix}-\begin{bmatrix}2\\2\end{bmatrix}\right)^t \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}\left(\begin{bmatrix}0\\0\end{bmatrix}-\begin{bmatrix}2\\2\end{bmatrix}\right)\right)} = \frac{1}{(2\pi)^{2/2}(1)^{1/2}} e^{-4}$$

Since $p\left(\begin{bmatrix}0\\0\end{bmatrix}\middle|\,\omega_1\right) \geq p\left(\begin{bmatrix}0\\0\end{bmatrix}\middle|\,\omega_2\right)$ 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 $\boldsymbol{\omega_1}$**if** $\boldsymbol{(x_1 - 4)^2 - (x_2 - 4)^2 \geq 16 + 4\ln(2)}$**; otherwise decide** $\boldsymbol{\omega_2}$

b)



c) 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:

$$\frac{1406 + 818}{10000 + 10000} = \frac{2224}{20000} = 0.1112$$

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:

```
mu = [0 0];
cov = [2 0;0 2];
determ = det(cov);
invers = inv(cov);

mu2 = [2 2];
cov2 = [1 0;0 1];
determ2 = det(cov2);
invers2 = inv(cov2);

%part c
rng('default');
R = mvnrnd(mu, cov, 10000);
scatter(R(:,1), R(:,2), 'filled', 'MarkerFaceColor', 'g');

hold on;

W = mvnrnd(mu2, cov2, 10000);
scatter(W(:,1), W(:,2), 'filled', 'MarkerFaceColor', 'b');
title('Plot of 10,000 random points each from class 1 and class 2');
xlabel('Feature value x1');
ylabel('Feature value x2c');
legend('class 1', 'class 2');


tp = 0;
fp = 0;
tn = 0;
fn = 0;
```

```
for i=1:10000
    %the first 10000 points, generated in R (class 1)
    lhs = (1/(sqrt(determ*(2*pi)^2)))*exp((-1/2)*(R(i,:)-mu)*invers*(R(i,:)-mu)');
    rhs = (1/(sqrt(determ2*(2*pi)^2)))*exp((-1/2)*(R(i,:)-mu2)*invers2*(R(i,:)-mu2)');
    if lhs >= rhs %then assign to class 1. since these points are from R, this is a true positive (tp)
        tp = tp + 1;
    else %then assign to class 2. since these points are from R, this is a false positive (fp)
        fp = fp + 1;
    end
    lhs = (1/(sqrt(determ*(2*pi)^2)))*exp((-1/2)*(W(i,:)-mu)*invers*(W(i,:)-mu)');
    rhs = (1/(sqrt(determ2*(2*pi)^2)))*exp((-1/2)*(W(i,:)-mu2)*invers2*(W(i,:)-mu2)');
     if lhs >= rhs %then assign to class 1. since these points are from W, this is a false negative (fn)
        fn = fn + 1;
     else %then assign to class 2. Since these points are from W, this is a true negative(tn)
        tn = tn + 1;
    end
end

%confusion matrix: tp = 8594, fp = 1406; fn = 818, tn = 9182
%error for class 1: 14.06%
%error for class 2: 8.18%
%total error from the 20,000 samples = 2224/20000 = 11.12%


%ezplot(fin, [[-1,11],[-1,11]]);
viscircles([4,4], sqrt(16+4*log(2)));
```

Problem 9:

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

$$P(\omega_1|x) = P(\omega_2|x)$$

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

$$p(x|\omega_1) = p(x|\omega_2)$$

$$p(x|\omega) = \frac{1}{2\pi^{d/2}|\Sigma|^{1/2}} e^{\left(-\frac{1}{2}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-u\right)^t \Sigma^{-1}\left(\begin{bmatrix}x_1\\x_2\end{bmatrix}-u\right)\right)}$$

$$p(x|\omega_1) = \frac{1}{2\pi^{2/2}(1)^{1/2}} e^{\left(-\frac{1}{2}\left(\begin{bmatrix}x_1+1\\x_2+2\end{bmatrix}\right)^t I\left(\begin{bmatrix}x_1+1\\x_2+1\end{bmatrix}\right)\right)}$$

$$p(x|\omega_2) = \frac{1}{2\pi^{2/2}(1)^{1/2}} e^{\left(-\frac{1}{2}\left(\begin{bmatrix}x_1-1\\x_2-2\end{bmatrix}\right)^t I\left(\begin{bmatrix}x_1-1\\x_2-1\end{bmatrix}\right)\right)}$$

$$\frac{1}{2\pi^{2/2}(1)^{1/2}} e^{\left(-\frac{1}{2}\left(\begin{bmatrix}x_1+1\\x_2+2\end{bmatrix}\right)^t I\left(\begin{bmatrix}x_1+1\\x_2+1\end{bmatrix}\right)\right)} = \frac{1}{2\pi^{2/2}(1)^{1/2}} e^{\left(-\frac{1}{2}\left(\begin{bmatrix}x_1-1\\x_2-2\end{bmatrix}\right)^t I\left(\begin{bmatrix}x_1-1\\x_2-1\end{bmatrix}\right)\right)}$$

We can cancel out the common term $\frac{1}{2\pi^{2/2}(1)^{1/2}}$ and take the natural logarithm of both sides:

$$-\frac{1}{2}\left(\begin{bmatrix}x_1+1\\x_2+2\end{bmatrix}\right)^t I\left(\begin{bmatrix}x_1+1\\x_2+1\end{bmatrix}\right) = -\frac{1}{2}\left(\begin{bmatrix}x_1-1\\x_2-2\end{bmatrix}\right)^t I\left(\begin{bmatrix}x_1-1\\x_2-1\end{bmatrix}\right)$$

$$-\frac{1}{2}({x_1}^2 + 2x_1 + 1 + {x_2}^2 + 2x_1 + 1) = -\frac{1}{2}({x_1}^2 - 2x_1 + 1 + {x_2}^2 - 2x_1 + 1)$$

At this point, the left hand side is $p(x|\omega_1)$ and right hand side is $p(x|\omega_2)$. 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 $\omega_1$if $p(x|\omega_1) \geq p(x|\omega_2)$. So , we can write this as:

$$-\frac{1}{2}({x_1}^2 + 2x_1 + 1 + {x_2}^2 + 2x_1 + 1) \geq -\frac{1}{2}({x_1}^2 - 2x_1 + 1 + {x_2}^2 - 2x_1 + 1)$$

Since we divide by $-\frac{1}{2}$on both side, the inequality flips.

$$({x_1}^2 + 2x_1 + 1 + {x_2}^2 + 2x_1 + 1) \leq ({x_1}^2 - 2x_1 + 1 + {x_2}^2 - 2x_1 + 1)$$

$$4x_1 + 4x_2 \leq 0$$
$$x_1 + x_2 \leq 0$$
$$x_2 \leq -x_1$$

Therefore, our decision boundary is $\boldsymbol{x_2 = -x_1}$

Our decision rule is:

Decide ω_1 if $x_2 \leq -x_1$, otherwise decide ω_2

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

$$P(error) = P^{\beta}(\omega_1)P^{1-\beta}(\omega_2)\int p^{\beta}(x|\omega_1)p^{1-\beta}(x|\omega_2)dx$$

where

$$\int p^{\beta}(x|\omega_1)p^{1-\beta}(x|\omega_2)dx = e^{-k(\beta)}$$

and

$$k(\beta) = \frac{\beta(1-\beta)}{2}(u_2-u_1)^t[\beta\Sigma_1 + (1-\beta)\Sigma_2]^{-1}(u_2-u_1) + \frac{1}{2}\ln\frac{|\beta\Sigma_1 + (1-\beta)\Sigma_2|}{|\Sigma_1|^{\beta}|\Sigma_2|^{1-\beta}}$$

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

$$e^{-k(\beta)} = e^{-(4\beta(\beta-1))}$$

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

$$f(\beta) = e^{g(\beta)}$$
$$f'(\beta) = e^{g(\beta)}g'(\beta)$$

Therefore, the first derivative is:

$$f'(\beta) = e^{(4\beta(\beta-1))}(8\beta-4) = 0$$

Since $e^{-(4\beta(\beta-1))}$ cannot be 0, the only way for $f'(\beta) = 0$ is if:

$$(8\beta-4) = 0$$
$$\beta = \frac{1}{2}$$

So, the original expression reduces to:

$$e^{(4\beta(\beta-1))}$$
$$e^{(4\left(\frac{1}{2}\right)(\frac{1}{2}-1))}$$
$$e^{-1}$$

Chernoff bound is therefore:

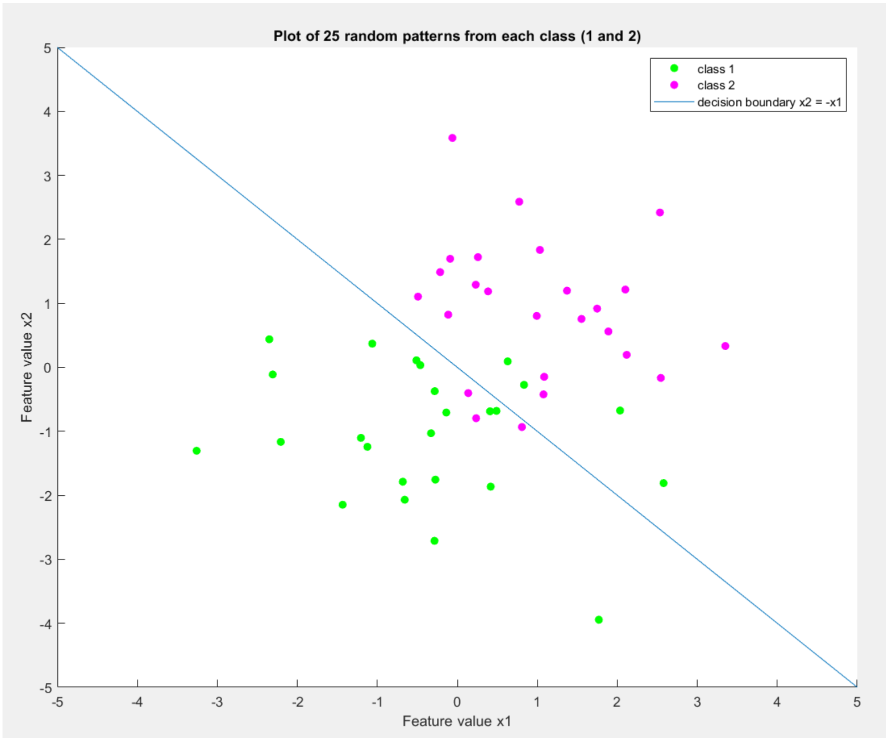
$$P(error) = P^{\beta}(\omega_1)P^{1-\beta}(\omega_2)e^{-k(\beta)}$$
$$P(error) = \sqrt{P(\omega_1)}\sqrt{P(\omega_2)}e^{-1}$$
$$P(error) = \sqrt{\frac{1}{2}}\sqrt{\frac{1}{2}}e^{-1}$$
$$P(error) = \frac{1}{2}e^{-1} = \mathbf{0.18395}$$

The Bhattacharya bound can be calculating by simply assuming that $\beta = \frac{1}{2}$

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

$$P(error) = \sqrt{P(\omega_1)P(\omega_2)}e^{-k(\frac{1}{2})} = \frac{1}{2}e^{-1} = \mathbf{0.18395}$$

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 = $\frac{7}{50} = 0.14$

e)

Since the error bound is 0.18395, we need to misclassify $\lceil 0.18395 * 50 \rceil = 10$ 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:

```
mu = [-1 -1];
mu2 = [1 1];
sigma = [1 0;0 1];

syms B
chernoff = 1/exp((B*(1-B)/2)*(mu2 - mu)*inv(B*sigma + (1-B)*sigma)*(mu2-mu)' + (1/2)*log(det(B*sigma + (1-B)*sigma)/(((det(sigma))^B)*(det(sigma)^(1-B)))));
%ezplot(chernoff, [0 1]);
%type in "chernoff" in the command window to see the simplified expression.

der = diff(chernoff, B);
B_min = solve(der == 0);
f(B) = chernoff;
P_error = sqrt((1/2)*(1/2))*f(1/2);

%take derivative of equation and equal to 0: B = 1/2. Can also see clearly
%in the graph

%Since B =1/2, the Bhattacharya and Chernoff bounds are identical

beta = 1/2;
final = exp(-(beta*(1-beta)/2)*(mu2 - mu)*inv(beta*sigma + (1-beta)*sigma)*(mu2-mu)' + (1/2)*log(det(beta*sigma + (1-beta)*sigma)/(((det(sigma))^beta)*(det(sigma)^(1-beta)))));
%P(error) = sqrt(p(w1)*p(w2))*e^-k(1/2) = sqrt(p(w1)*p(w2))*final =
%1/2(0.3679) = 0.18395

%rng('default');
misclassified = zeros(1000,1); %ran my code 1000 times to test if empirical error can exceed theoretical bounds
for j=1:1000
R = mvnrnd(mu, sigma, 25);
%scatter(R(:,1), R(:,2), 'filled', 'MarkerFaceColor', 'g');
hold on;
W = mvnrnd(mu2, sigma, 25);
%scatter(W(:,1), W(:,2), 'filled', 'MarkerFaceColor', 'm');

% syms x
% y = -x;
% fplot(x,y);
%
% title('Plot of 25 random patterns from each class (1 and 2)');
% xlabel('Feature value x1');
% ylabel('Feature value x2');
% legend('class 1','class 2','decision boundary x2 = -x1');

tp = 0;
fp = 0;
tn = 0;
fn = 0;

for i=1:25
    if -1*R(i,1) >= R(i,2) %then we assign to class 1 as per decision rule, correctly
        tp = tp + 1;
    else %then we assign to class 2, wrongly
        fp = fp + 1;
    end
    if -1*W(i,1) >= W(i,2) %then we assign to class 1, wrongly
        fn = fn + 1;
    else %then we assign to class 2, correctly
        tn = tn + 1;
    end
end
misclassified(j,1) = fp+fn; %highest misclassified was 12, happened during 134th j

end
%error rate in this random example is (4+3)/50 = 0.14
```

Problem 10:

The Bayes minimum risk rule states that we must calculate the conditional risk for each possible action $R(\alpha_1|x)$ and pick the action with the lowest risk, denoted as R^*

As before, the conditional risk is given by;

$$R(\alpha_i|x) = \sum_{j=1}^{c} \lambda(\alpha_i|\omega_j)P(\omega_j|x)$$

$$R(\alpha_1|x) = \lambda_{11}P(\omega_1|x) + \lambda_{12}P(\omega_2|x)$$

Since $\lambda_{11} = 0$, 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 $p(x)$ is simply a scaling factor and is common to all 3 actions, it will be ignored:

$$R(\alpha_1|x) = \lambda_{12}P(\omega_2|x) = (1)\frac{p(x|\omega_2)P(\omega_2)}{p(x)} = (1)p(x|\omega_2)P(\omega_2) = (1)\left(\frac{1}{2}\right)\left(\frac{1}{3}\right) = \frac{1}{6}$$

$$R(\alpha_2|x) = \lambda_{21}P(\omega_1|x) + \lambda_{22}P(\omega_2|x)$$

As before, since $\lambda_{22} = 0$, the second term can be cancelled out. Bayes formula is used again:

$$R(\alpha_2|x) = \lambda_{21}P(\omega_1|x) = (1)\frac{p(x|\omega_1)P(\omega_1)}{p(x)} = (1)p(x|\omega_1)P(\omega_1) = (1)\left(\frac{2-x}{2}\right)\left(\frac{2}{3}\right)$$

We are given that x = 0.5, so:

$$R(\alpha_2|x) = (1)\left(\frac{2-0.5}{2}\right)\left(\frac{2}{3}\right) = \frac{3}{6} = \frac{1}{2}$$

$$R(\alpha_3|x) = \lambda_{31}P(\omega_1|x) + \lambda_{32}P(\omega_2|x)$$

It is given that $\lambda_{31} = \lambda_{32} = \frac{1}{4}$, so:

$$R(\alpha_3|x) = \frac{1}{4}P(\omega_1|x) + \frac{1}{4}P(\omega_2|x)$$

$$R(\alpha_3|x) = \frac{1}{4}p(x|\omega_1)P(\omega_1) + \frac{1}{4}p(x|\omega_2)P(\omega_2)$$

$$R(\alpha_3|x) = \frac{1}{4}\left(\frac{2-x}{2}\right)\left(\frac{2}{3}\right) + \frac{1}{4}\left(\frac{1}{2}\right)\left(\frac{1}{3}\right)$$

$$R(\alpha_3|x) = \frac{1}{4}\left(\frac{2-0.5}{2}\right)\left(\frac{2}{3}\right) + \frac{1}{24} = \frac{3}{24} + \frac{1}{24} = \frac{4}{24} = \frac{1}{6}$$

The cost of α_1 and α_3 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 $x = 0.5$