1. Problem 13 (page 69)

13. In many pattern classification problems one has the option either to assign the pattern to one of $c$ classes, or to reject it as being unrecognizable. If the cost for rejects is not too high, rejection may be a desirable action. Let

$$\lambda(\alpha_i | \omega_j) = \begin{cases} 
0 & i = j \quad i, j = 1, ..., c \\
\lambda_r & i = c + 1 \\
\lambda_s & \text{otherwise,}
\end{cases}$$

where $\lambda_r$ is the loss incurred for choosing the $(c+1)$th action, rejection, and $\lambda_s$ is the loss incurred for making a substitution error. Show that the minimum risk is obtained if we decide $\omega_i$ if $P(\omega_i | x) \geq P(\omega_j | x)$ for all $j$ and if $P(\omega_i | x) \geq 1 - \lambda_r / \lambda_s$, and reject otherwise. What happens if $\lambda_r = 0$? What happens if $\lambda_r > \lambda_s$?

2. Problem 23 (page 71)

23. Consider the three-dimensional normal distribution $p(x | \omega) \sim N(\mu, \Sigma)$ where

$$\mu = \begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}$$

and

$$\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{pmatrix}.$$

(a) Find the probability density at the point $x_0 = (0.5, 0, 1)^T$.

(b) Construct the whitening transformation $A_w$. Show your $\Lambda$ and $\Phi$ matrices. Next, convert the distribution to one centered on the origin with covariance matrix equal to the identity matrix, $p(x | \omega) \sim N(0, \mathbf{I})$.

(c) Apply the same overall transformation to $x_0$ to yield a transformed point $x_w$.

(d) By explicit calculation, confirm that the Mahalanobis distance from $x_0$ to the mean $\mu$ in the original distribution is the same as for $x_w$ to 0 in the transformed distribution.

(e) Does the probability density remain unchanged under a general linear transformation? In other words, is $p(x_0 | N(\mu, \Sigma)) = p(T^T x_0 | N(T^T \mu, T^T \Sigma T))$ for some linear transform $T$? Explain.

(f) Prove that a general whitening transform $A_w = \Phi \Lambda^{-1/2}$ when applied to a Gaussian distribution insures that the final distribution has covariance proportional to the identity matrix $\mathbf{I}$. Check whether normalization is preserved by the transformation.

3. Problem 37 (or 36 in some prints) (page 75)
36. Consider a two-category classification problem in two dimensions with \( p(x|\omega_1) \sim N(0, I) \), \( p(x|\omega_2) \sim N \left( \begin{pmatrix} 1 \\ 1 \end{pmatrix}, I \right) \), and \( P(\omega_1) = P(\omega_2) = 1/2 \).

(a) Calculate the Bayes decision boundary.

(b) Calculate the Bhattacharyya error bound.

(c) Repeat the above for the same prior probabilities, but \( p(x|\omega_1) \sim N \left( 0, \begin{pmatrix} 2 & 0.5 \\ 0.5 & 2 \end{pmatrix} \right) \) and \( p(x|\omega_2) \sim N \left( \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix} \right) \).

4. Computer Exercise 2 (page 80)

<table>
<thead>
<tr>
<th>sample</th>
<th>( \omega_1 )</th>
<th>( \omega_2 )</th>
<th>( \omega_3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( x_1 )</td>
<td>( x_2 )</td>
<td>( x_3 )</td>
</tr>
<tr>
<td>1</td>
<td>-5.01</td>
<td>-8.12</td>
<td>-3.68</td>
</tr>
<tr>
<td>2</td>
<td>-5.43</td>
<td>-3.48</td>
<td>-3.54</td>
</tr>
<tr>
<td>3</td>
<td>1.08</td>
<td>-5.52</td>
<td>1.66</td>
</tr>
<tr>
<td>4</td>
<td>0.86</td>
<td>-3.78</td>
<td>-4.11</td>
</tr>
<tr>
<td>5</td>
<td>-2.67</td>
<td>0.63</td>
<td>7.39</td>
</tr>
<tr>
<td>6</td>
<td>4.94</td>
<td>3.29</td>
<td>2.08</td>
</tr>
<tr>
<td>7</td>
<td>-2.51</td>
<td>2.09</td>
<td>-2.59</td>
</tr>
<tr>
<td>8</td>
<td>-2.25</td>
<td>-2.13</td>
<td>-6.94</td>
</tr>
<tr>
<td>9</td>
<td>5.56</td>
<td>2.86</td>
<td>-2.26</td>
</tr>
<tr>
<td>10</td>
<td>1.03</td>
<td>-3.33</td>
<td>4.33</td>
</tr>
</tbody>
</table>

2. Use your classifier from Problem ?? to classify the following 10 samples from the table above in the following way. Assume that the underlying distributions are normal.

(a) Assume that the prior probabilities for the first two categories are equal \( P(\omega_1) = P(\omega_2) = 1/2 \) and \( P(\omega_3) = 0 \) and design a dichotomizer for those two categories using only the \( x_1 \) feature value.

(b) Determine the empirical training error on your samples, i.e., the percentage of points misclassified.

(c) Use the Bhattacharyya bound to bound the error you will get on novel patterns drawn from the distributions.

(d) Repeat all of the above, but now use two feature values, \( x_1 \), and \( x_2 \).

(e) Repeat, but use all three feature values.

(f) Discuss your results. In particular, is it ever possible for a finite set of data that the empirical error might be larger for more data dimensions?