Homework 2

CSE 802 - Pattern Recognition and Analysis
Total Points: 100
Instructor: Dr. Arun Ross
Due Date: March 1, 12:40pm

Note:

1. **You are permitted to discuss the following questions with others in the class. However, you *must* write up your *own* solutions to these questions. Any indication to the contrary will be considered an act of academic dishonesty. Copying from *any source* constitutes academic dishonesty.**
2. **A neatly typed report is expected (alternately, you can neatly handwrite the report and then scan it). The report, in PDF format, must be uploaded in D2L by March 1, 12:40 pm. Late submissions will not be graded. In your submission, please include the names of individuals you discussed this homework with and the list of external resources (e.g., websites, other books, articles, etc.) that you used to complete the assignment (if any).**
3. **When solving equations or reducing expressions you must explicitly show every step in your computation and/or include the code that was used to perform the computation. Missing steps or code will lead to a deduction of points.**
4. **Code developed as part of this assignment must be (a) included as an appendix to your report or inline with your solution, and (b) archived in a single zip file and uploaded in D2L. Including the code without the outputs or including the outputs without the code will result in deduction of points.**

---

1. Consider a set of 1-dimensional feature values (i.e., points) pertaining to a class that is available here.

   (a) [3 points] Plot the histogram of these points using a bin size of 2, i.e., each bin should have a range of 2.

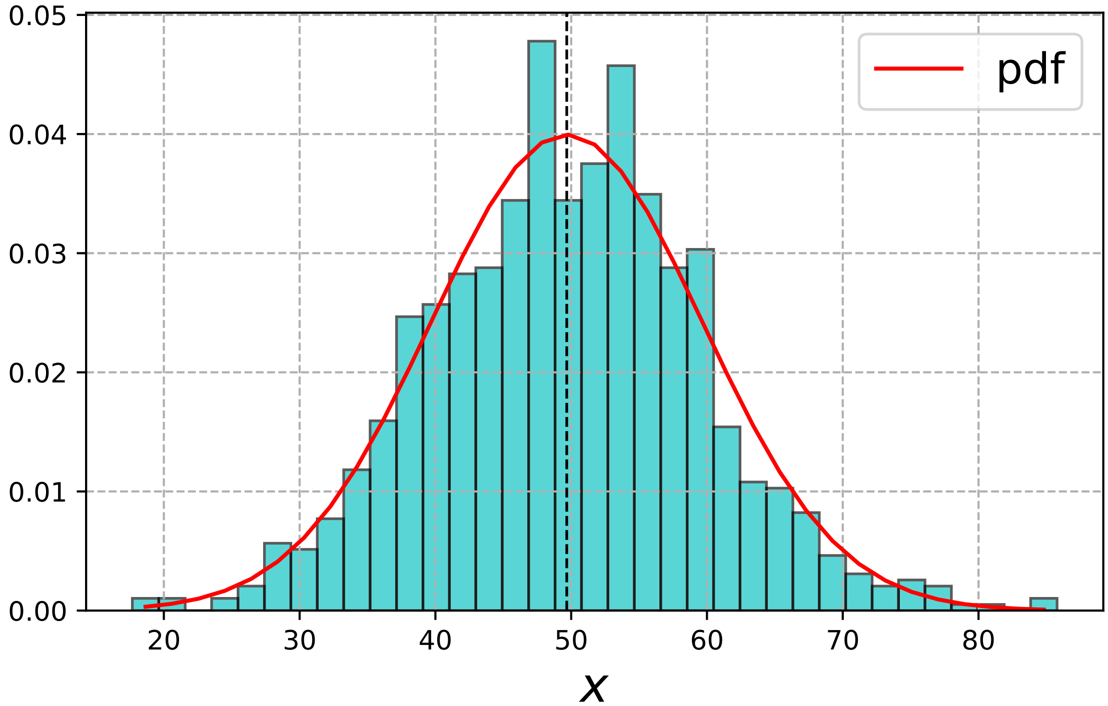

Figure 1: Histogram and a pdf of 1-dimensional feature values

(b) [4 points] Compute and report the mean and the biased variance of these points.
Solution:
Mean is 49.673675
Biased variance is 99.693624033375

(c) [3 points] Assuming that the given points are generated by an underlying Gaussian distribution, plot the pdf function on the same graph as (a).
Solution:
The plot is shown in Figure 1.

2. [10 points] Consider the problem of distinguishing between two classes of fish - ω_1 and ω_2 - based on their length. You can assume that these are the only two classes of fish in nature and that the length of the fish is a discrete integral value! The probability of observing a fish from class ω_1 is 0.6 and the probability of observing a fish from class ω_2 is 0.4. The probability of encountering a fish of length 10 inches, given that it is from class ω_1, is 0.2. Similarly, the probability of encountering a fish of length 10 inches, given that it is from class ω_2, is 0.4.

(a) Based on this information, compute the probability of encountering a fish of any class of length 10 inches?
Solution:
$P(\omega_1) = 0.6$, $P(\omega_2) = 0.4$
$p(x = 10|\omega_1) = 0.2$, $p(x = 10|\omega_2) = 0.4$
$p(x = 10) = p(x = 10|\omega_1)P(\omega_1) + p(x = 10|\omega_2)P(\omega_2) = (0.2)(0.6) + (0.4)(0.4) = 0.28$

Thus, we now have

$$P(\omega_1|x=10) = \frac{p(x=10|\omega_1)P(\omega_1)}{p(x=10)} = \frac{(0.2)(0.6)}{0.28} = 0.4286$$

and

$$P(\omega_2|x=10) = \frac{p(x=10|\omega_2)P(\omega_2)}{p(x=10)} = \frac{(0.4)(0.4)}{0.28} = 0.5714$$

(b) Based on the Bayes decision rule, to which class will a fish of length 10 inches be assigned to?

Solution:
A fish of length 10 inches will be assigned to a class ω_2

3. [15 points] Consider a 1-dimensional classification problem involving two categories ω_1 and ω_2 such that $P(\omega_1) = 2/3$ and $P(\omega_2) = 1/3$. Assume that the classification process can result in one of three actions:

α_1 - choose ω_1;
α_2 - choose ω_2;
α_3 - do not classify.

Consider the following loss function, λ:

$\lambda(\alpha_1|\omega_1) = \lambda(\alpha_2|\omega_2) = 0$;
$\lambda(\alpha_2|\omega_1) = \lambda(\alpha_1|\omega_2) = 1$;
$\lambda(\alpha_3|\omega_1) = \lambda(\alpha_3|\omega_2) = 1/4$.

For a given feature value x, assume that $p(x|\omega_1) = \frac{2-x}{2}$ and $p(x|\omega_2) = 1/2$. Here, $0 \leq x \leq 2$.

Based on the Bayes minimum risk rule, what action will be undertaken when encountering the value $x = 0.5$?

Solution: There are two actions that equally minimize, i.e., $R(\alpha_1|x)$ and $R(\alpha_3|x)$, and hence taking either one of the actions can be applied.

From

$$R(\alpha_i|x) = \sum_{j=1}^{C} \lambda_{ij} P(\omega_j|x)$$

and

$$\begin{aligned} p(x) &= \sum_{j=1}^{c} p(x|\omega_j)P(\omega_j) \\ &= p(x|\omega_1)P(\omega_1) + p(x|\omega_2)P(\omega_2) \\ &= (\frac{2-x}{2})(\frac{2}{3}) + (\frac{1}{2})(\frac{1}{3}) \\ &= \frac{4-2x+1}{6} \\ &= \frac{5-2x}{6} \end{aligned}$$

Thus,

$$
\begin{aligned}
R(\alpha_1|x=0.5) &= \lambda_{11}P(\omega_1|x) + \lambda_{12}P(\omega_2|x) \\
&= 0 + \lambda_{12}\frac{p(x|\omega_2)P(\omega_2)}{p(x)} \\
&= \frac{1(\frac{1}{2})(\frac{1}{3})}{\frac{5-2x}{6}} \\
&= (\frac{1}{6})(\frac{6}{5-2(\frac{1}{2})}) \\
&= \frac{1}{4}
\end{aligned}
$$

and

$$
\begin{aligned}
R(\alpha_2|x=0.5) &= \lambda_{21}P(\omega_1|x) + \lambda_{22}P(\omega_2|x) \\
&= \lambda_{21}\frac{p(x|\omega_1)P(\omega_1)}{p(x)} + 0 \\
&= \frac{\frac{4-2x}{6}}{\frac{5-2x}{6}} \\
&= (\frac{4-2x}{6})(\frac{6}{5-2x}) \\
&= (\frac{4-2(\frac{1}{2})}{6})(\frac{6}{5-2(\frac{1}{2})}) \\
&= \frac{3}{4}
\end{aligned}
$$

and

$$
\begin{aligned}
R(\omega_3|x=0.5) &= \lambda_{31}P(\omega_1|x) + \lambda_{32}P(\omega_2|x) \\
&= \lambda_{31}\frac{p(x|\omega_1)P(\omega_2)}{p(x)} + \lambda_{32}\frac{p(x|\omega_2)P(\omega_2)}{p(x)} \\
&= (\frac{1}{4})(\frac{4-2x}{6})(\frac{6}{5-2x}) + (\frac{1}{4})(\frac{1}{6})(\frac{6}{5-2x}) \\
&= (\frac{1}{4})(\frac{4-2(\frac{1}{2})}{5-2(\frac{1}{2})}) + (\frac{1}{4})(\frac{1}{5-2(\frac{1}{2})}) \\
&= \frac{3}{16} + \frac{1}{16} \\
&= \frac{4}{16} = \frac{1}{4}
\end{aligned}
$$

4. Consider a two-class problem with the following class-conditional pdfs (Gaussian density functions):

$$p(x \mid \omega_1) \sim N(50, 5)$$

and

$$p(x \mid \omega_2) \sim N(30, 10).$$

(a) [4 points] Plot the two class-conditional pdfs in the interval $x \in [10, 125]$ on the same graph.

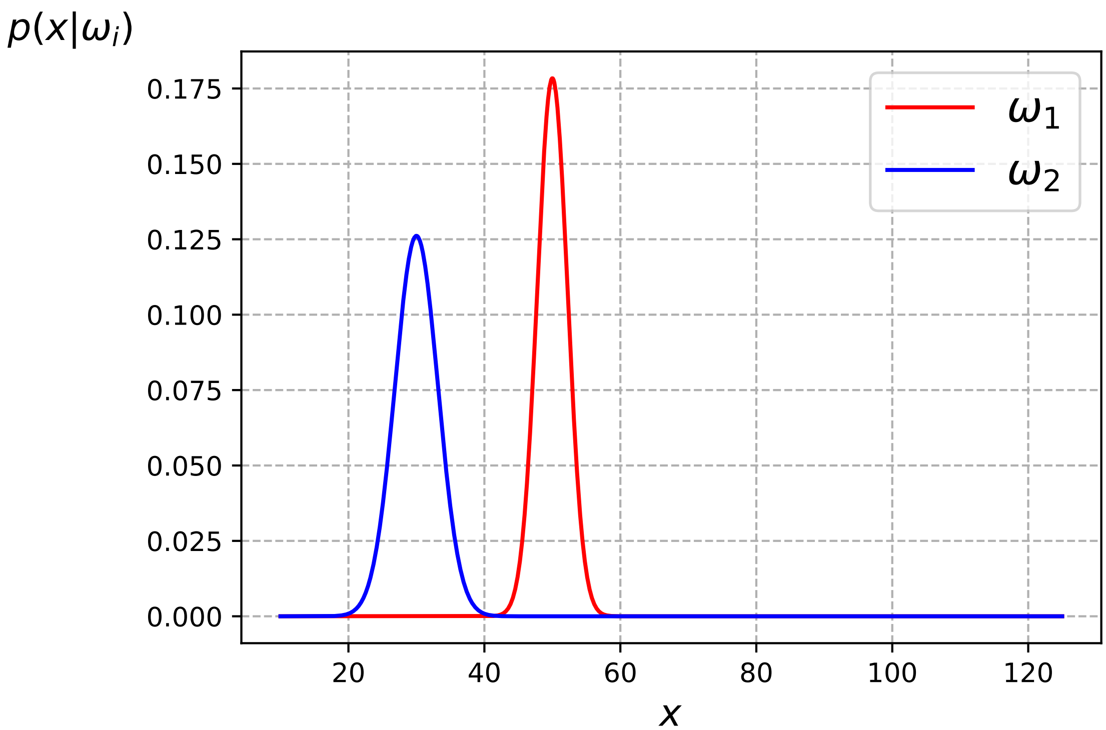

Figure 2: Class-conditional pdfs of two classes

(b) [4 points] In another graph, plot the likelihood ratio, $\frac{p(x|\omega_1)}{p(x|\omega_2)}$, in the range $x \in [10, 125]$.

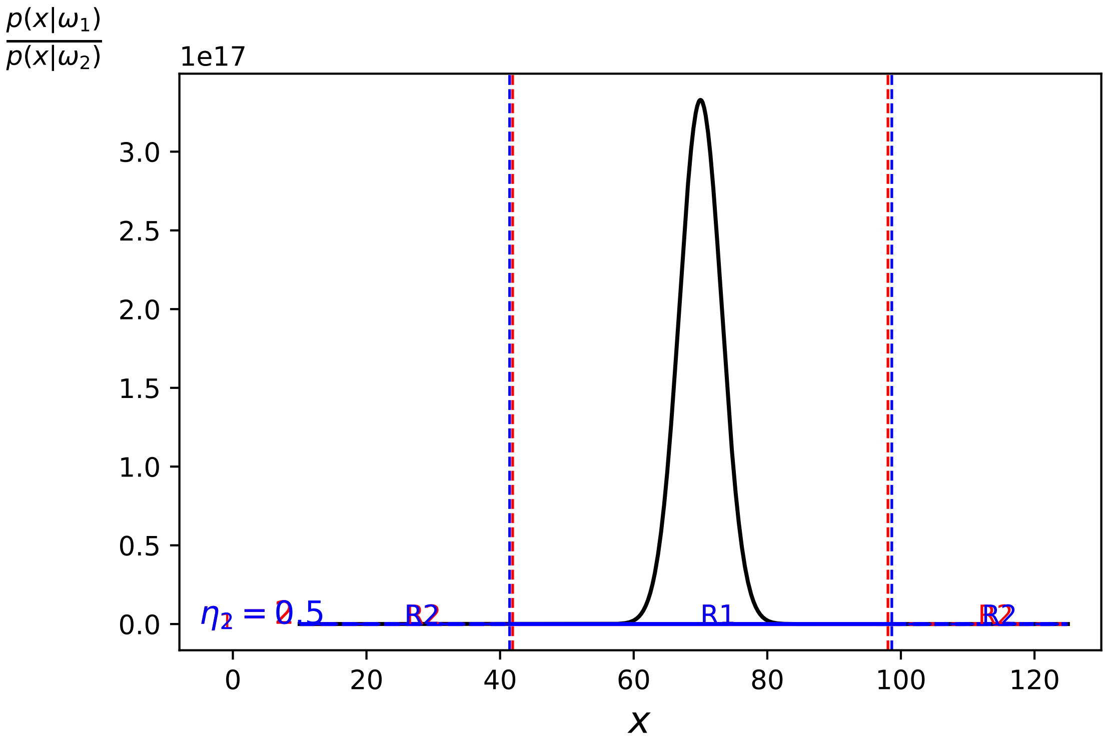

Figure 3: The likelihood ratio, $\frac{p(x|\omega_1)}{p(x|\omega_2)}$

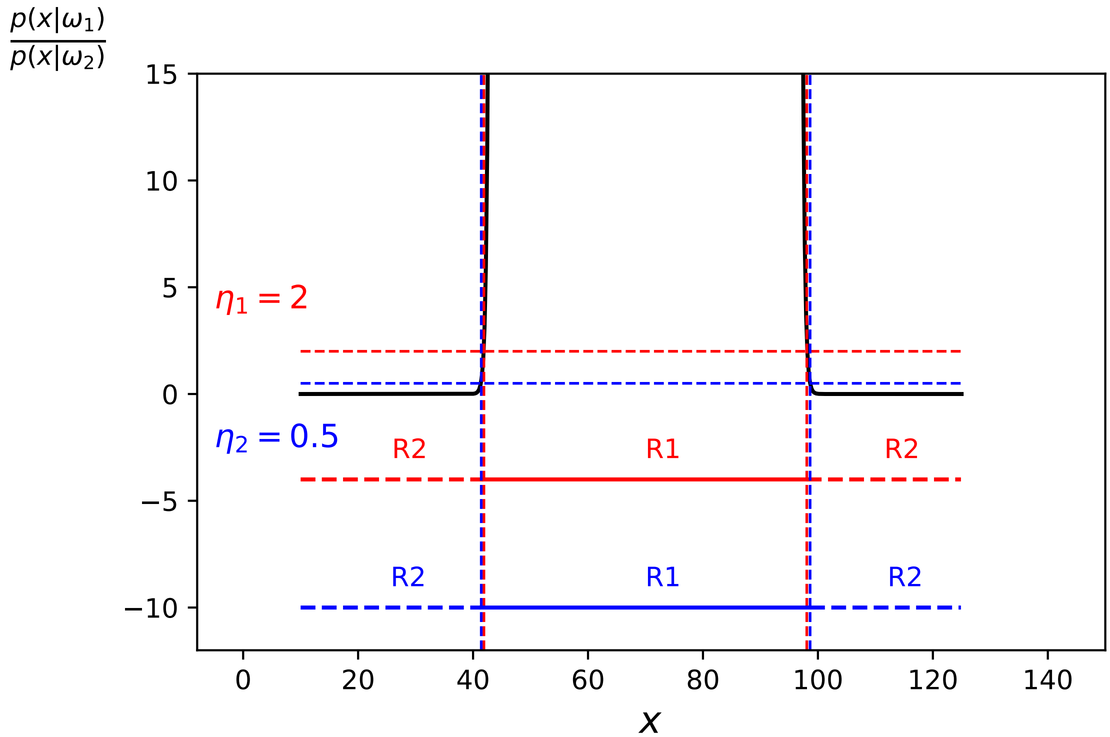

Figure 4: The likelihood ratio $\frac{p(x|\omega_1)}{p(x|\omega_2)}$, please ignore the negative value on the y-axis, this is just to show the decision regions.

(c) [5 points] Recall that the likelihood ratio can be compared against a threshold, say η, in order to assign a feature value x to one of the two classes, ω_1 or ω_2. Assuming that $\lambda_{11} = \lambda_{22} = 0$, $\lambda_{12} = 2$, $\lambda_{21} = 1$, $P(\omega_1) = 0.5$ and $P(\omega_2) = 0.5$, what is the value of η? Show this threshold in graph 4b and mark the decision regions corresponding to ω_1 and ω_2.

Solution:

$$\eta = \frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})} \frac{p(\omega_2)}{p(\omega_1)} = \frac{2-0}{1-0} \frac{0.5}{0.5} = 2$$

(d) [5 points] Now suppose that $\lambda_{11} = \lambda_{22} = 0$, $\lambda_{12} = 1$, $\lambda_{21} = 2$, $P(\omega_1) = 0.5$ and $P(\omega_2) = 0.5$, what is the value of η? Show this threshold also in graph 4b and mark the decision regions corresponding to ω_1 and ω_2.

Solution:

$$\eta = \frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21} - \lambda_{11})} \frac{p(\omega_2)}{p(\omega_1)} = \frac{1-0}{2-0} \frac{0.5}{0.5} = 0.5$$

(e) [2 points] Explain in words, the underlying reason for the change in decision regions in 4c and 4d.

Solution: Our decision boundary from 4c is determined by $\eta = 2$, in which the loss incurred for taking action α_1 when the true class is ω_2 is greater than converse (i.e., $\lambda_{12} > \lambda_{21}$). Now in 4d, our loss value from taking action α_2 when the true class is ω_1 is greater than the converse (i.e., $\lambda_{21} > \lambda_{12}$), leading to a smaller $\eta = 0.5$, and hence R2 becomes smaller and R1 becomes bigger. This can be intuitively explained that to minimize the overall risk, we need to select the action that minimizes the conditional risk. i.e., the risk of taking α_2 (classifying ω_2) when the true class is ω_1 needs to be minimized.

5. [10 points] In many pattern classification problems, the classifier is allowed to reject an input pattern by not assigning it to any one of the c classes. If the cost of rejection is not too high, it may be a desirable action in some cases. Let

$$\lambda(\alpha_i \mid \omega_j) = \begin{cases} 0, & i = j \quad (i, j = 1, \dots c) \\ \lambda_r, & i = c + 1 \\ \lambda_s, & i \neq j \quad (i, j = 1, \dots c), \end{cases}$$

where λ_r is the loss incurred for rejecting an input pattern and λ_s is the loss incurred for misclassifying the input pattern (known as substitution error).

(a) Show that the minimum risk rule results in the following decision policy.
Assign pattern $\boldsymbol{x}$ to class ω_i if $P(\omega_i \mid \boldsymbol{x}) \geq P(\omega_j \mid \boldsymbol{x}) \forall j$ **AND** $P(\omega_i \mid \boldsymbol{x}) \geq 1 - \lambda_r/\lambda_s$, else reject it.
Solution:
The conditional risk is:

$$\begin{aligned} R(\alpha_i|x) &= \sum_{j=1}^{c} \lambda_{ij} P(\omega_j|x) \\ &= 0 + \sum_{j!=i} \lambda_s P(\omega_j|x) \\ &= \lambda_s(1 - P(\omega_i|x)) \end{aligned}$$

Assign x to the class ω_i if i has the highest posterior probability, and the overall risk/loss is less than the loss incurred for rejecting x:

$$\begin{aligned} \lambda_s(1 - P(\omega_i|x)) &< \lambda_r \\ (1 - P(\omega_i|x)) &< \lambda_r/\lambda_s \\ P(\omega_i|x) &\geq 1 - \lambda_r/\lambda_s \end{aligned}$$

Thus, to minimize the average probability of error, we select the i that maximizes the posterior probability $P(\omega_i|x)$. This results in the following rule:

$$\omega_i \text{ if } P(\omega_i \mid \boldsymbol{x}) \geq P(\omega_j \mid \boldsymbol{x}) \forall j \textbf{ AND } P(\omega_i \mid \boldsymbol{x}) \geq 1 - \lambda_r/\lambda_s\text{; otherwise reject it}$$

(b) Explain what happens to the decision policy if $\lambda_r = 0$? Similarly, explain what happens to the decision policy if $\lambda_r > \lambda_s$?
Solution:
In the case $\lambda_r = 0$, there is no loss incurred for rejecting an input pattern. Also, the last inequality is never satisfied. Thus, the decision rule will always reject an input pattern x. But if we assume that the input pattern x still belongs to one of the c classes, the last inequality shows that we should reject or classify only if ω_i has the maximum posterior probability.

In the case $\lambda_r > \lambda_s$, there is a high loss of rejecting. This means the last inequality is always satisfied. Thus the decision rule is to select the ω_i that maximizes $P(\omega_i|x)$.

6. [10 points] Consider a two-class one-dimensional classification problem with the following class-conditional densities:

$$p(x|\omega_1) = 2 - 2x, \quad x \in [0, 1]$$
$$p(x|\omega_2) = 2x, \quad x \in [0, 1]$$

(a) Plot these two densities in the same graph.

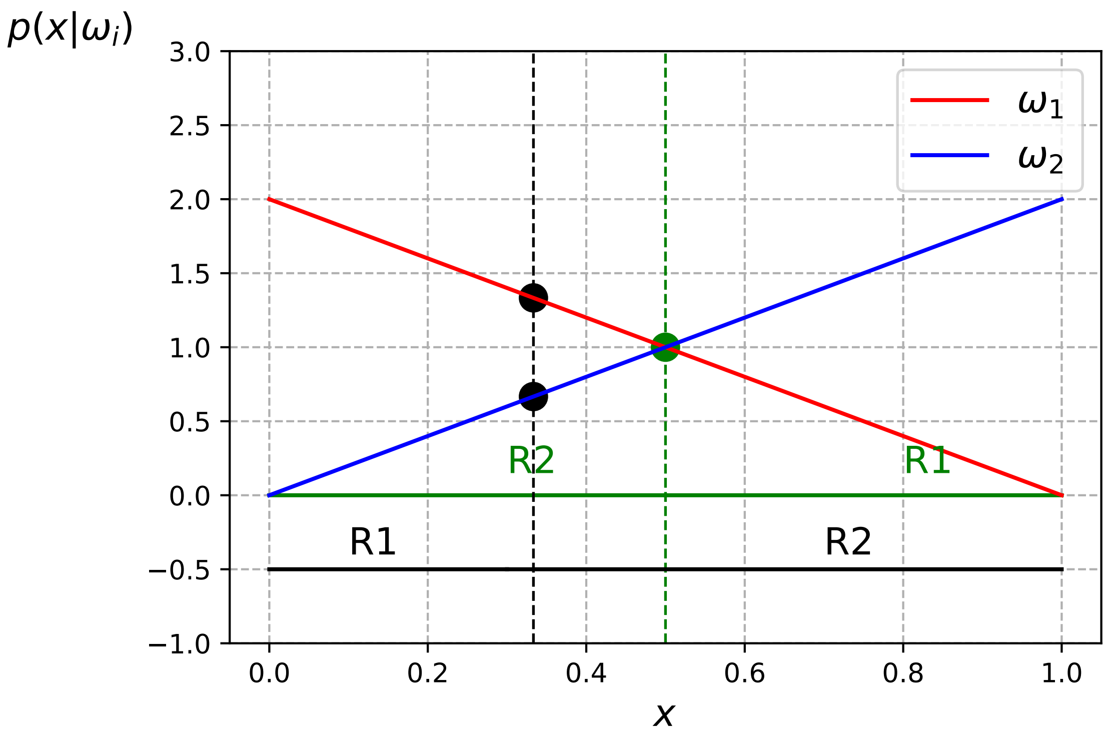

Figure 5: Two densities with respective decision boundaries and decision regions.

(b) Let $P(\omega_1) = P(\omega_2) = 1/2$. Assuming a 0-1 loss function, compute the Bayes decision boundary and write down the Bayes decision rule. Mark the decision boundary and decision regions on the figure in (a).

Solution:
Bayes decision boundary is derived from:

$$p(x|\omega_1) = p(x|\omega_2)$$
$$2 - 2x = 2x$$
$$x = 0.5$$

and Bayes decision rule is:

$$\omega_1 \text{if } p(x|\omega_1) > p(x|\omega_2); \text{ otherwise } \omega_2$$

(c) Let $P(\omega_1) = P(\omega_2) = 1/2$. Suppose the loss function is defined as follows: $\lambda_{11} = \lambda_{22} = 0$, $\lambda_{12} = 2$ and $\lambda_{21} = 1$. Compute the Bayes decision boundary and write down the Bayes decision rule. Note that λ_{ij} represents the loss incurred when a sample from class ω_j is classified as ω_i. Mark the decision boundary and decision regions on the figure in (a).

Solution:
Bayes decision boundary is derived from:

$$\begin{aligned}
R(\alpha_1|x) &= R(\alpha_2|x) \\
\lambda_{11}P(\omega_1|x) + \lambda_{12}P(\omega_2|x) &= \lambda_{21}P(\omega_1) + \lambda_{22}P(\omega_2|x) \\
\lambda_{12}\frac{p(x|\omega_2)P(\omega_2)}{p(x)} &= \lambda_{21}\frac{p(x|\omega_1)P(\omega_1)}{p(x)} \\
2(2x)(\frac{1}{2}) &= (2-2x)(\frac{1}{2}) \\
x &= \frac{1}{3}
\end{aligned}$$

and the fundamental Bayes decision rule is:

$$\omega_1 \text{ if } R(\alpha_1|x) < R(\alpha_2|x)\text{; otherwise } \omega_2$$

However, we are looking at the two densities graph, and there are a variety of ways of expressing the minimum-risk decision rule. Thus, by employing Bayes's formula, we will replace the above rule with prior probabilities and the conditional densities. This results in the following Bayes decision rule.

$$\omega_1 \text{ if } (\lambda_{21} - \lambda_{11})p(x|\omega_1)P(\omega_1) \ > \ (\lambda_{12} - \lambda_{22})p(x|\omega_2)P(\omega_2)\text{; otherwise } \omega_2$$

(d) Intuitively explain why the boundaries in (b) and (c) are different.

Solution:
In (b) the prior probabilities are equal, i.e., $P(\omega_1) = P(\omega_2)$, this does not provide information about the true class. Also, a zero-one loss function is employed. In this case, the decision is based entirely on the likelihoods $p(x|\omega_i)$. Now in (c), the loss function is given, in which λ_{ij} telling the loss incurred for taking α_i when the true class is ω_j. We can now also see that loss incurred for classifying ω_1 when the true class is ω_2 is higher than the converse. Thus, to minimize the overall risk, we need to minimize the risk from misclassifying ω_2, leading to Bayes decision boundary shifting to the left. Any a particular x falling on the left of the Bayes decision boundary belongs to ω_1, and falling on the right belongs to ω_2.

7. [10 points] Consider a two-class problem with the following class-conditional probability density functions (pdfs):

$$p(x \mid \omega_1) \sim N(0, \sigma^2)$$

and

$$p(x \mid \omega_2) \sim N(1, \sigma^2).$$

Show that the threshold, τ, corresponding to the Bayes decision boundary is:

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

where we have assumed that $\lambda_{11} = \lambda_{22} = 0$.

Solution:
Bayes decision boundary is derived from:

$$
\begin{aligned}
R(\alpha_1|x) &= R(\alpha_2|x) \\
\lambda_{11}P(\omega_1|x) + \lambda_{12}P(\omega_2|x) &= \lambda_{21}P(\omega_1) + \lambda_{22}P(\omega_2|x) \\
\lambda_{12}\frac{p(x|\omega_2)P(\omega_2)}{p(x)} &= \lambda_{21}\frac{p(x|\omega_1)P(\omega_1)}{p(x)} \\
\lambda_{12}P(\omega_2)\frac{1}{\sqrt{2\pi}\sigma}e^{\left[\frac{-1}{2}\frac{x^2}{\sigma^2}\right]} &= \lambda_{21}P(\omega_1)\frac{1}{\sqrt{2\pi}\sigma}e^{\left[\frac{-1}{2}\frac{(x-1)^2}{\sigma^2}\right]} \\
\ln\left[\frac{\lambda_{12}P(\omega_2)}{\lambda_{21}P(\omega_1)}\right] &= \frac{1}{2\sigma^2}(x^2-2x+1-x^2) \\
-2x+1 &= 2\sigma^2\ln\left[\frac{\lambda_{12}P(\omega_2)}{\lambda_{21}P(\omega_1)}\right] \\
x &= \frac{1}{2}-\sigma^2\ln\left[\frac{\lambda_{12}P(\omega_2)}{\lambda_{21}P(\omega_1)}\right]
\end{aligned}
$$

8. Consider the three-dimensional normal distribution $p(\boldsymbol{x}) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu} = (1,1,1)^t$ and $\boldsymbol{\Sigma} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 2 \\ 0 & 2 & 5 \end{bmatrix}$.

   (a) [2 points] Compute and report the determinant of the covariance matrix, i.e., $|\Sigma|$.
   Solution:
   $|\Sigma| = 21$

   (b) [2 points] Compute and report the inverse of the covariance matrix, i.e., Σ^{-1}.
   Solution:
   $$\Sigma^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.23809524 & -0.0952381 \\ 0 & -0.0952381 & 0.23809524 \end{bmatrix}$$

   (c) [3 points] Compute and report the eigen-vectors and eigen-values of the covariance matrix.
   Solution:
   $eigen-values = \begin{bmatrix} 7 & 3 & 1 \end{bmatrix}$
   $$eigen-vectors = \begin{bmatrix} 0 & 0 & 1 \\ 0.70710678 & 0.70710678 & 0 \\ 0.70710678 & -0.70710678 & 0 \end{bmatrix}$$

   (d) [3 points] Compute and report the density value at $(0,0,0)^t$ and at $(5,5,5)^t$.

   Solution:
   density value at $(0,0,0)^t$ is 0.00016093
   density value at $(5,5,5)^t$ is 2.23101257e-53

   (e) [2 points] Compute the Euclidean Distance between $\boldsymbol{\mu}$ and the point $(5,5,5)^t$.
   Solution:
   Euclidean Distance is 6.928203230275509

Squared Euclidean Distance is 48

(f) [3 points] Compute the Mahalanobis Distance between $\boldsymbol{\mu}$ and the point $(5, 5, 5)^t$.
Solution:
Mahalanobis Distance is 4.535573676110727
Squared Mahalanobis Distance is 20.57142857

A Appendices

1. Python code in Jupyter Notebook (with outputs) for Q1 is located at notebooks/q1.ipynb
2. Python code in Jupyter Notebook (with outputs) for Q4 is located at notebooks/q4.ipynb
3. Python code in Jupyter Notebook (with outputs) for Q6 is located at notebooks/q6.ipynb
4. Python code in Jupyter Notebook (with outputs) for Q8 is located at notebooks/q8.ipynb