Homework 2

CSE 802 - Pattern Recognition and Analysis
Total Points: 100
Instructor: Dr. Arun Ross
Due Date: February 21, 2023, 11pm ET

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 11:00pm on February 21, 2023. 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.**
5. **The report (PDF) and code (ZIP) must be uploaded as two separate files in D2L.**

---

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. Normalize the histogram so that the sum of the histogram values equals 1.

   (b) [4 points] Compute and report the mean and the (biased) variance of these points.

   (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). (Hint: Use the computed values for the mean and variance as parameters of the Gaussian).

   (d) [3 points] Assuming that the given points are generated by an underlying **Laplacian distribution**, plot the pdf function on the same graph as (a). (Hint: Use the computed values for the mean and variance to compute the parameters of the Laplacian).

   (e) [2 points] By visual examination, which of the two parametric distributions better fit the data?

   Note: In this problem, you are using already available formulae to compute the parameters of the Gaussian and Laplacian. Later on in the course, we will derive such formulae based on the maximum likelihood estimation technique.

2. 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) [5 points] Based on this information, compute the probability of encountering a fish of any class of length 10 inches?
   (b) [5 points] Based on the Bayes decision rule, to which class will a fish of length 10 inches be assigned to?

3. [10 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, $0 \le x \le 2$.

   Based on the Bayes minimum risk rule, what action will be taken when encountering the value $x = 0.45$?

4. 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) [2 points] Plot these two densities in the same graph.

   (b) [3 points] Let $P(\omega_1) = P(\omega_2) = 1/2$. Assuming a 0-1 loss function, compute the Bayes decision boundary and derive the Bayes decision rule. Mark the decision boundary and decision regions on the figure in (a).

   (c) [3 points] 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 $\lambda_{21} = 1$. Derive the Bayes decision boundary and 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).

   (d) [2 points] Intuitively explain why the boundaries in (b) and (c) are different.

5. 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., $|\Sigma|$.

(b) [2 points] Compute and report the inverse of the covariance matrix, i.e., Σ^{-1}.

(c) [3 points] Compute and report the eigen-vectors and eigen-values of the covariance matrix.

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

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

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

6. [10 points] Consider the following class-conditional densities for a three-class problem involving two-dimensional features:

$$p(\boldsymbol{x}|\omega_1) \sim N\left((-1,-1)^t, I\right);$$
$$p(\boldsymbol{x}|\omega_2) \sim N\left((1,1)^t, I\right);$$
$$p(\boldsymbol{x}|\omega_3) \sim \frac{1}{2}N\left((0.5,0.5)^t, I\right) + \frac{1}{2}N\left((-0.5,-0.5)^t, I\right).$$

(Here, class ω_3 conforms to a **Gaussian Mixture Model (GMM)** with two components - one component is $N\left((0.5,0.5)^t, I\right)$ and the other component is $N\left((-0.5,-0.5)^t, I\right)$ - whose weights are equal (i.e., $\frac{1}{2}$). Note that a GMM is not a Gaussian Distribution).

(a) In a 2D graph, mark the mean of ω_1, ω_2, and the two components of ω_3. In the same graph, mark the point $\boldsymbol{x} = (0.1, 0.1)^t$.

(b) Assuming a 0-1 loss function and equal priors, determine the class to which you will assign the two-dimensional point $\boldsymbol{x} = (0.1, 0.1)^t$ based on the Bayes decision rule.

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 $\lambda_{11} = \lambda_{22} = 0$.

8. [20 points] The **iris (flower) dataset** consists of 150 4-dimensional patterns (i.e., feature vectors) belonging to three classes (setosa=1, versicolor=2, and virginica=3). There are 50 patterns per class. The 4 features correspond to sepal length in cm (x_1), sepal width in cm (x_2), petal length in cm (x_3), and petal width in cm (x_4). Note that the class labels are indicated at the end of every pattern.

Assume that each class can be modeled by a multivariate Gaussian density, i.e., $p(\boldsymbol{x}|\omega_j) \sim \mathrm{N}(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, $j = 1,2,3$. Write a program to design a Bayes classifier and test it by following the steps below:

(a) Train the classifier: Using the first 25 patterns of each class (training data), compute $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$, $j = 1,2,3$. **Report** these values.

(b) Design the Bayes classifier: Assuming that the three classes are equally probable and a 0-1 loss function, write a program that inputs a 4-dimensional pattern $\boldsymbol{x}$ and assigns it to one of the three classes based on the maximum posterior rule, i.e., assign $\boldsymbol{x}$ to ω_{j^*} if,

$$j^* = \arg\max_{j=1,2,3} \{P(\omega_j|\boldsymbol{x})\}.$$

(c) Test the classifier: Classify the remaining 25 patterns of each class (test data) using the Bayes classifier constructed above and report the confusion matrix for this three-class problem. What is the empirical error rate on the test set?
