IITM-CS5691 : Pattern Recognition and Machine Learning — Release Date: August 30, 2023
Assignment 1 — Due Date : September 14, 2023, 23:59

Roll No: CS23E001 — Name: Shuvrajeet Das
Collaborators (if any):
References/sources (if any):

- Use LaTeX to write-up your solutions (in the solution blocks of the source LaTeX file of this assignment), and submit the resulting pdf files (one per question) at Crowdmark by the due date. (Note: **No late submissions** will be allowed, other than one-day late submission with 10% penalty or four-day late submission with 30% penalty! Instructions to join Crowdmark and submit your solution to each question within Crowdmark **TBA** later).
- For the programming question, please submit your code (rollno.ipynb file and rollno.py file in rollno.zip) directly in moodle, but provide your results/answers (including Jupyter notebook **with output**) in the pdf file you upload to Crowdmark.
- Collaboration is encouraged, but all write-ups must be done individually and independently, and mention your collaborator(s) if any. Same rules apply for codes written for any programming assignments (i.e., write your own code; we will run plagiarism checks on codes).
- If you have referred a book or any other online material or LLMs (Large Language Models like ChatGPT) for obtaining a solution, please cite the source. Again don't copy the source *as is* - you may use the source to understand the solution, but write-up the solution in your own words (this also means that you cannot copy-paste the solution from LLMs!). Please be advised that *the lesser your reliance on online materials or LLMs* for answering the questions, *the more your understanding* of the concepts will be and *the more prepared you will be for the course exams*.
- Points will be awarded based on how clear, concise and rigorous your solutions are, and how correct your answer is. The weightage of this assignment is 12% towards the overall course grade.

---

1. (8 points) [REVEREND BAYES DECIDES FURTHER!]

   (a) (2 points) For a two-class optimal Bayes classifer h, the decision region is given by: $R_i = \{x \in \mathbb{R} : h(x) = C_i\}$. Is R_1 always a single interval (based on a single cutoff separating the C_1 and C_2 class) or can R_1 be composed of more than one discontiguous interval? If yes for latter, give an example by plotting the pdfs $p(x, C_1)$ and $p(x, C_2)$ against x.

   (b) (2 points) For a binary classifer h, let $L = \begin{bmatrix} p & q \\ r & s \end{bmatrix}$ be the loss matrix; and $C_{\text{train}} = \begin{bmatrix} 100 & 10 \\ 20 & 120 \end{bmatrix}$, and $C_{\text{test}} = \begin{bmatrix} 90 & 45 \\ 30 & 85 \end{bmatrix}$ be the confusion matrix when h is applied on the training and test data respectively. All three matrices have ground-truth classes t along the rows and predictions h along the columns in the same order for the two classes. Express your estimate of the expected loss of h in terms of p to s above.

   (c) (4 points) Consider the dataset introduced in the table below, where the task is to predict whether a person is ill. We use a representation based on three features per subject to describe an individual person. These features are "running nose (N)", "coughing (C)", and "reddened skin (R)", each of which can take the value true ('+') or false ('–'). (i) Classify the data point ($d_7 : N = -, C = +, R = -$) using a Naive Bayes classifier. As part of your solution, also write down the (ii) Naive Bayes assumption and (iii) Naive Bayes classifier, along with (iv) which distribution's MLE formula you used to estimate the class conditionals.

Training Example	N (running nose)	C (coughing)	R (reddened skin)	Classification
d_1	+	+	+	positive (ill)
d_2	+	+	–	positive (ill)
d_3	–	–	+	positive (ill)
d_4	+	–	–	negative (healthy)
d_5	–	–	–	negative (healthy)
d_6	–	+	+	negative (healthy)

Solution:

The solution of question (a)

In a two-class optimal Bayes classifier, the decision regions are typically separated by a single cutoff point, resulting in a single interval for each region. This is often the case when the class-conditional probability density functions (pdfs) are well-behaved and exhibit a clear separation between the two classes.

However, there can be scenarios where the decision region R1 may be composed of more than one discontiguous interval if the class-conditional pdfs overlap in such a way that it's not possible to define a single cutoff point to separate the classes. This situation is more common when the class-conditional pdfs overlap significantly.

Let's illustrate this with an example by plotting the pdfs $p(x, C_1)$ and $p(x, C_2)$ against x:

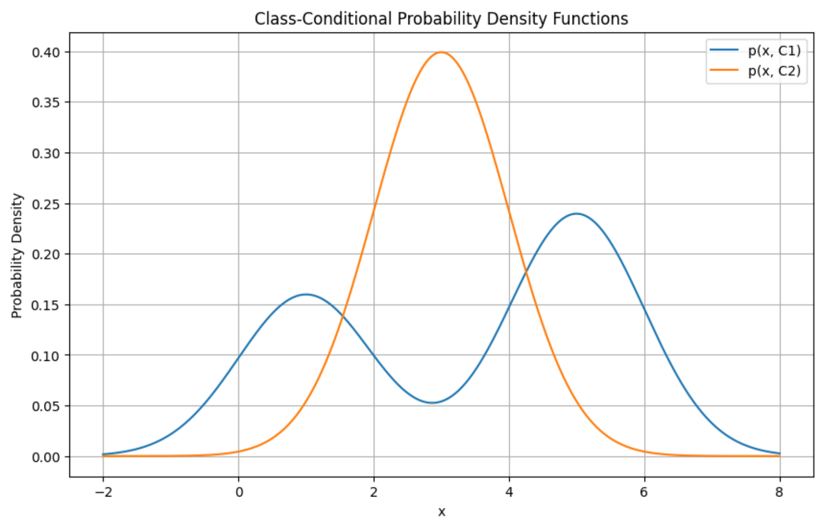
Suppose we have two classes, C1 and C2, and their class-conditional pdfs are given as follows:

For Class C1: $p(x, C1) = 0.4 * N(1, 1) + 0.6 * N(5, 1)$

For Class C2: $p(x, C2) = N(3, 1)$

Here, $N(\mu, \sigma^2)$ represents the normal distribution. In this example, the pdf for Class C_1 is a mixture of two Gaussian distributions with different means and weights.

Let's plot these pdfs:



In this plot, you can see that the two class-conditional pdfs overlap significantly, and there isn't a single cutoff point that cleanly separates the two classes. Therefore, in such cases, the decision region R1 for Class C1 would consist of multiple discontiguous intervals.

The solution of question (b)

The expected loss for the train is,
$\mathcal{E}_{train} = 100.p + 10.q + 20.r + 120.s$

The expected loss for the train is,
$\mathcal{E}_{test} = 90.p + 45.q + 30.r + 85.s$

The solution to question (c)

Let's assume the dataset $D = \{\{1, 1, 1\}, \{1, 1, 0\}, \{0, 0, 1\}, \{1, 0, 0\}, \{0, 0, 0\}, \{0, 1, 1\}\}$ and
$Y = \{\{1, 1, 1, 0, 0, 0\}\}$ where '+' denotes 1 and '-' denotes 0
we know posterior $\propto$ likelihood x prior,
Assuming the priors for each class 0 is 0.5 and class 1 is 0.5 (since both classes are in the same quantities)

Now, likedlihood is,

$$\mathcal{L}_n(X_1, X_2, X_3, X_4, X_5, X_6, \theta) = \prod_{i=1}^{6} p_\theta(x_i)$$

i.e. for class 1, the likelihood is 0.5, and for class 0 is 0.5

Now from the Bayes rule, the desired class is argmax of the posterior of the 2 classes but in this, the posterior probability is the same so we can choose any 1 or 0.

(i) The classified point is 1 or 0 for anyone whom you chose.

(ii) The Naive Bayes assumption is that prior is $\frac{N_{samples=C_i}}{N_{samples}}$ where $i \in \{0, 1\}$

(iii) The Naive Bayes Classifier $P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$

(iv) The Distribution is assumed to be Bernoulli.