Roll No.: _

Amrita Vishwa Vidyapeetham

Amrita School of Engineering, Coimbatore

B.Tech First Assessment Examinations – August 2022

Seventh Semester

Computer Science Engineering

19CSE432 Pattern Recognition

Duration: Two hours

Maximum: 50 Marks

Course Outcomes (COs):

CO	Course Outcomes
CO01	Understand basic concepts in pattern recognition
CO02	Understand discriminant functions and apply them for applications..
CO03	Understand and apply Parametric techniques of Pattern recognition..
CO04	Apply Non parametric techniques of PR and analyze their performance..
CO05	Understand the supervised and unsupervised learning algorithms and apply them for real world problems.

Answer all questions

1) Represent two PR applications with their input and output in tabular form .[2] [CO01][BTL1]
Any two among the following: - 2 * 1 = 2

Problem	Input	Output
Speech recognition	Speech waveforms	Spoken words, speaker identity
Non-destructive testing	Ultrasound, eddy current, acoustic emission waveforms	Presence/absence of flaw, type of flaw
Medical waveform analysis	EKG, EEG waveforms	Types of cardiac conditions, classes of brain conditions
Remote sensing	Multispectral images	Terrain forms, vegetation cover
Aerial reconnaissance	Visual, infrared, radar images	Tanks, airfields
Character recognition (page readers, zip code, license plate)	scanned image	Alphanumeric characters

2) List some essential modules required and challenges to overcome, for a Number Plate Recognition System. [2] [CO01][BTL2]
Modules Required: 0.25 * 4 = 1

i. **Acquisition**
ii. **Enhancement**
iii. **segmentation**
iv. **character recognition**

challenges : Any two of the following : 0.5 *2 = 1

I. **Poor image resolution**
II. **Motion blur**
III. **low contrast**
IV. **view point variation and occlusion**
V. **Different fonts and background**

3) Describe the term "Curse of dimensionality" and "Conditional Risk" with appropriate mathematical notations [2][CO01][BTL2]

Curse of Dimensionality: Error rate may in fact increase with too many features in thecase of small number of training samples - 1 mark

Conditional Risk: - 1 mark

The risks associated with taking action α_i given data x

$$R\left({}^{\alpha_i}/_{x}\right) = \sum_{j=1}^{c} \lambda\left({}^{\alpha_i}/_{\omega_j}\right) P\left({}^{\omega_j}/_{x}\right)$$

4) You are assigned a new project which involves helping a food delivery company save more money. The problem is, company's delivery team aren't able to deliver food on time. As a result, their customers get unhappy. To keep them happy, they end up delivering food for free. Is this a Machine Learning problem ? If yes, what are the necessary requirements for formulating a solution, if no what category of the problems does it fall into? . Explain in one or two sentences. [2][CO01][BTL2]

This is not a machine learning problem. This is a route optimization problem. A machine learning problem consist of three things:

i. **There exist a pattern.**
ii. **You cannot solve it mathematically (even by writing exponential equations).**
iii. **You have data on it.**

Always look for these three factors to decide if machine learning is a tool to solve a particular problem.

5) Feature x is normally distributed for class A with $\mu_{A=0}$ and $\sigma_{A=1}$. For class B, x is normally distributed with $\mu_{B=1}$ and $\sigma_{B=1}$. The prior probabilities are P(A) = 0.6 and p(B) =0.4. Where are the optimal decision regions? [3] [CO02][BTL3]

identification of x being a univariate distribution - 1 Mark

Usage of formula for univaraite distribution - $\boldsymbol{\frac{1}{\sigma_{G\sqrt{2\pi}}} e^{\frac{-1}{2}\left(\frac{x-\mu_G}{\sigma_G}\right)^2}}$ **-1 Mark**

identification of decision regions in terms of x calculating Conditional Risk(R) and concluding in terms of R - 1 mark

6) Consider a c-class classification problem; under what conditions would the optimal classifier be equivalent to [4][CO02][BTL3]

a. The minimum distance classifier

$\sum_i = \boldsymbol{\sigma^2 I}$ **Assuming equal priors - 2**

b. **The Mahalanobis distance classifier** **- 2**

$$\sum_i = \sum$$

7) Suppose a bank classifies customers as either good or bad credit risks. On the basis of extensive historical data, the bank has observed that 1% of good credit risks and 10% of bad credit risks overdraw their account in any given month. A new customer opens a checking account at this bank. On the basis of a check with a credit bureau, the bank believe that there is a 70% chance the customer will turn out to be a good credit risk. [10] CO02][BTL3]

   a) Suppose that this customer's account is overdrawn in the first month. How does this alter the bank's opinion of this customer's creditworthiness?

Let G and O represent the following events: G: customer is considered to be a good credit risk. O: customer overdraws checking account. From the bank's historical data, we have P(O|G) = 0.01 and P(O|G˜) = 0.1, and we also know that the bank's initial opinion about the customer's creditworthiness is given by P(G) = 0.7. - 1

Given the information that the customer has an overdraft in the first month, the bank's revised opinion about the customer's creditworthiness is given by the conditional probability P(G|O). Using Bayes' theorem and the law of total probability,

Stating Bayes formula - 1

Application - 2

correct answer - 1

$$\begin{aligned} P(G|O) &= \frac{P(O|G)P(G)}{P(O)} \\ &= \frac{P(O|G)P(G)}{P(O|G)P(G) + P(O|\tilde{G})P(\tilde{G})} \\ &= \frac{0.01 \times 0.7}{0.01 \times 0.7 + 0.1 \times 0.3} \\ &\approx 0.189 \end{aligned}$$

   b) Given (a), what would be the bank's opinion of the customer's creditworthiness at the end of the second month if there was not an overdraft in the second month?
   According to (a), at the beginning of the second month, the prior probability of the customer being a good risk is 0.189. Since the customer does not have an overdraft 1 in the second month, the bank's opinion about the customer's creditworthiness at the end of the month is given by - 1

Stating Bayes formula - 1

Application - 2

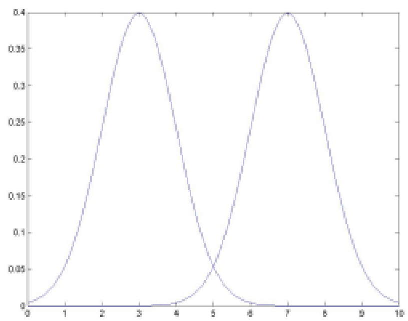
correct answer - 1

$$\begin{aligned} P(G|\tilde{O}) &= \frac{P(\tilde{O}|G)P(G)}{P(\tilde{O})} \\ &= \frac{P(\tilde{O}|G)P(G)}{P(\tilde{O}|G)P(G) + P(\tilde{O}|\tilde{G})P(\tilde{G})} \\ &= \frac{0.99 \times 0.189}{0.99 \times 0.189 + 0.9 \times 0.811} \\ &\approx 0.204 \end{aligned}$$

8) Consider a two-category classification problem with one-dimensional Gaussian distributions $p(x/w_i) \sim N(\mu_i, \sigma^2)$, i = 1, 2 (i.e. they have same variance but different means).

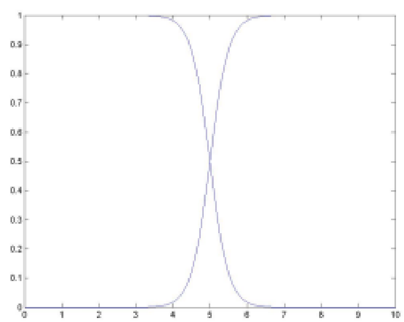
   a) Sketch the two densities $p(x/w_1)$ and $p(x/w_2)$ in one figure

   Sketching both the densities - 2.5



   b) Sketch the two posterior probabilities $P(w_1|x)$ and $P(w_2|x)$ in one figure [5] [CO02][BTL3]

   Sketching both the posterior probabilities - 2.5



9) Imagine you are on a game show and you are given the choice of three doors: behind one door is a car and behind the other two doors are goats. You have the opportunity to select a door (say No. 1). Then the host, who knows exactly what is behind each door and will not reveal the car, will open a different door (i.e., one that has a goat). The host then asks you if you want to switch your selection to the last remaining door. [10] [CO02][BTL3]

a) Formulate the problem using the Bayes rule, i.e., what are the random variables and the input data. What are the meaning of the prior and the posterior probabilities in this problem (one sentence each)

Without the loss of generality, suppose that we chose door 1 at the beginning. Random Variables: Ci represents the state that the car is behind door i, i ∈ [1, 2, 3], and H_J represents the state that the host opens door j, j ∈ [2, 3] - 1

Input Data: The door that the host opens - 1

Prior Probability: The probability of winning the car before the host opens the door. - 1 Posterior Probability: The probability of winning the car after the host opens the door.- 1

b) What are the probability values for the prior?

$$\boldsymbol{P(C1) = \frac{1}{3}, P(C2) = \frac{1}{3}, P(C3) = \frac{1}{3}.}$$

c) What are the probability values for the likelihood?

P(H2|C1) = $\frac{1}{2}$, P(H3|C1) = $\frac{1}{2}$, P(H2|C2) = 0, P(H3|C2) = 1, P(H2|C3) = 1, P(H3|C3) = 0.

10) Assume a two-class problem with equal a priori class probabilities and Gaussian class-conditional densities as follows:

$p(x/w_1) = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} a & c \\ c & b \end{bmatrix}\right)$ $p(x/w_2) = N\left(\begin{bmatrix} d \\ e \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right)$ where $a \times b - c \times c = 1$

a) Find the equation of the decision boundary between these two classes in terms of the given parameters, after choosing a logarithmic discriminant function.

stating the logarithmic discriminant function - 3
stating the decision boundary condition - 2
calculations - 2

Given two normal density and equal prior probabilities, we choose the discriminant function as

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_\mathbf{i})^t \Sigma_i^{-1}(\mathbf{x} - \mu_\mathbf{i}) - \frac{1}{2} ln|\Sigma_i|$$

then the decision boundary is given by $g_1(\mathbf{x}) = g_2(\mathbf{x})$, and after simplified the function, we have

$$(b-1)x_1^2 + (a-1)x_2^2 - 2cx_1x_2 + 2dx_1 + 2ex_2 - d^2 - e^2 = 0$$

b) Determine the constraints on the values of a, b, c, d and e, such that the resulting discriminant function results with a linear decision boundary. [10] [CO02][BTL3]

stating of the covariance matrix condition - 1
stating of the conditions on a,b and c - 2

To get a linear decision boundary, we need $\Sigma_1 = \Sigma_2$. Thus we have $a = b = 1$ and $c = 0$.

****

Course Outcome /Bloom's Taxonomy Level (BTL) Mark Distribution Table

CO	Marks	BTL	Marks
CO01	**8**	BTL 1	**2**
CO02	**42**	BTL 2	**6**
CO03	**0**	BTL 3	**42**
CO04	**0**	BTL 4	
CO05		BTL 5	