Question Paper

Exam Date & Time: 05-Dec-2023 (02:30 PM - 05:30 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

VII Semester - End Semester Examination - Nov-Dec 2023

Pattern Recognition [ICT 4053]

Marks: 50 **Duration: 180 mins.**

Descriptive

Answer all the questions.

* Answer all questions.
* Assume the missing data suitably.
* Write neatly and legibly.
* Give suitable examples wherever necessary.

1) Give a neat diagram for generic pattern recognition system and briefly explain the following in context of this diagram: (5)

i) Sensing
ii) Segmentation and grouping
iii) Feature extraction
iv) Classification, and
v) Post-processing.

2) The discriminant function for the multivariate Gaussian density function is given by (3)

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_i)^T\Sigma_i^{-1}(\mathbf{x}-\boldsymbol{\mu}_i) - \frac{d}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| + \ln P(\omega_i).$$

Obtain the linear discriminnat function for the following cases:

i) Case 1: $\Sigma_i = \Sigma$
ii) Case 2: $\Sigma_i =$ arbitrary.

3) (2)

Consider the three-dimensional normal distribution $p(\mathbf{x}|\omega) \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$, where

$$\boldsymbol{\mu} = \begin{bmatrix}1\\2\\2\end{bmatrix} \text{ and } \Sigma = \begin{bmatrix}1 & 0 & 0\\0 & 5 & 2\\0 & 2 & 5\end{bmatrix}.$$

The whitening transformation matrix $\mathbf{A}_w$ is given by

$\lceil 1 \quad 0 \quad 0 \rceil$

$$\mathbf{A}_w = \begin{bmatrix} 0 & 1/\sqrt{6} & 1/\sqrt{14} \\ 0 & -1/\sqrt{6} & 1/\sqrt{14} \end{bmatrix}.$$

Apply this tranformation to $\mathbf{x}_0 = [0.8\ 0.2\ 2]^T$ to yield a transformed point $\mathbf{x}_w$. [Hint:$\mathbf{Y} = \mathbf{A}_w^T(\mathbf{x} - \boldsymbol{\mu}) \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$.]

4) Let each component x_i of $\mathbf{x}$ be binary valued (0 or 1) in a two-category problem with $P(\omega_1) = P(\omega_2) = 0.5$. Suppose that the probability of obtaining a 1 in any component is (5)

$$p_{i1} = p$$
$$p_{i2} = 1 - p,$$

and we assume for definiteness $p > 1/2$. The probability of error is known to approach zero as the dimensionality d approaches infinity.

i) Suppose that a single sample $\mathbf{x} = (x_1, \ldots, x_d)^T$ is drawn from category ω_1. Show that the maximum-likelihood estimate for p is given by

$$\hat{p} = \frac{1}{d}\sum_{i=1}^{d} x_i.$$

ii) Describe the behavior of $\hat{p}$ as d approaches infinity. Indicate why such behavior means that by letting the number of features increase without limit we can obtain an error-free classifier even though we have only one sample from each class.

5) Use Bayesian estimation technique to calculate the *a posteriori* density $p(\mu|\mathcal{D})$ for the univariate normal distribution. (3)

6) In the following, suppose a and b are positive constants greater than 1 and n a variable parameter. (2)

i) Is $b^{n+1} = O(b^n)$?

ii) Is $b^{an} = O(b^n)$?

7) (5)

Hidden Markov Model (HMM) can be represented as a probabilistic automata. Consider the probabilistic automata given in Fig.Q7, where the symbol $\bullet$ represents full stop. The emission symbol set and probability, B for the model is given as,

$$V = \{am, i, ict, department, of, student\},$$

and

$$B = \begin{bmatrix} 0.1 & 0.3 & 0.1 & 0.2 & 0.2 & 0.1 & 0 \\ 0.1 & 0.1 & 0.5 & 0.1 & 0.1 & 0.1 & 0 \\ 0.3 & 0.2 & 0.1 & 0.1 & 0.2 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

where the matrix index begins at 1. Assume that at $t = 0$ the model is in the starting state as indicated by the automata. Answer the following for the given HMM:

i) Find the probability that HMM generates for the sequence

$$\mathbf{V}^7 = \{i, am, student, of, ict, department, .\}.$$

ii) With a neat trellis diagram, show the sequence of hidden state which has lead to the generation of emitted symbol sequence in (i).

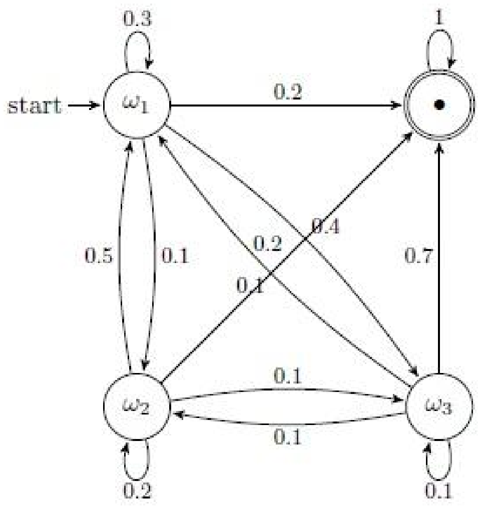

Fig.Q7

8) (3)

When an exhaustive search for clustering becomes infeasible, we use iterative optimization. The basic idea of iterative optimization is to find some reasonable initial partition and move samples, say $\hat{\mathbf{x}}$ from $\mathcal{D}_i$ to $\mathcal{D}_j$ if such a move improves the value of the criterion function, $\mathbf{J}_e$. Show that the transfer of $\hat{\mathbf{x}}$ from $\mathcal{D}_i$ to $\mathcal{D}_j$ is advantageous only if it meets the following criteria

$$\frac{n_i}{n_i - 1}||\hat{\mathbf{x}} - \mathbf{m}_i||^2 > \frac{n_j}{n_j + 1}||\hat{\mathbf{x}} - \mathbf{m}_j||^2.$$

9) Consider the following two-dimensional vectors from two categories: Which class should (2)

ω_1		ω_2	
x_1	x_2	x_1	x_2
7	2	5	6
5	1	7	5
2	3	4	5

be assigned to the point $(5, 4)$ by k-NN rule for $k = 3$?

10) Consider the quadratic discriminant function (5)

$$g(\mathbf{x}) = w_0 + \sum_{i=1}^{d} w_i x_i + \sum_{i=1}^{d}\sum_{j=1}^{d} w_{ij} x_i x_j$$

and define the symmetric, nonsingular matrix $\mathbf{W} = [w_{ij}]$. The basic character of the decision boundary can be described in terms of the scaled matrix

$$\overline{\mathbf{W}} = \frac{\mathbf{W}}{\mathbf{w}^T\mathbf{W}^{-1}\mathbf{w} - 4w_0}$$

and interpreted as follows

- If $\overline{\mathbf{W}} \propto \mathbf{I}$, then the decision boundary is hypersphere.
- If $\overline{\mathbf{W}}$ is positive definite, then the decision boundary is hyperellipsoid.
- If some eigen values of $\overline{\mathbf{W}}$ are positive and some negative, then the decision boundary is hyperhyperboloid.

Suppose $\mathbf{w} = \begin{bmatrix} 4 \\ 2 \\ -1 \end{bmatrix}$ and $\mathbf{W} = \begin{bmatrix} 2 & 1 & 0 \\ 4 & 3 & 1 \\ 1 & 0 & -2 \end{bmatrix}$. What is the character of the solution?

11) (3)

Apply PNN training for the given data points (show your calculation for each data point)

$$\omega_1 : \{(3,3),\ (3,5),\ (4,4),\ (2,4)\}$$
$$\omega_2 : \{(7,3),\ (6,4),\ (8,4),\ (7,5)\}.$$

PNN Training

initialize $j = 0, n = \#patterns, a_{ji} = 0$ for $j = 1, \dots, n; i = 1, \dots, c$

repeat

    $j \leftarrow j + 1$

    normalize: $x_{jk} \leftarrow x_{jk} / \left(\sum_{i=1}^{d} x_{ji}^2\right)^{1/2}$

    train: $w_{jk} \leftarrow x_{jk}$

    if $\mathbf{x}_j \in \omega_i$ **then**

        $a_{ji} \leftarrow 1$

until $j = n$

12) Consider a Markov model with given set of states $S = \{s_1, s_2, \dots, s_{|S|}\}$, wherein we can choose a series over time $\vec{z} \in S^T$. Assume that the transition matrix from a weather system is given by (2)

$$A = \begin{array}{c} \\ s_0 \\ s_{sun} \\ s_{cloud} \\ s_{rain} \end{array} \begin{array}{c} \begin{matrix} s_0 & s_{sun} & s_{cloud} & s_{rain} \end{matrix} \\ \begin{bmatrix} 0 & 0.4 & 0.5 & 0.1 \\ 0 & 0.5 & 0.2 & 0.3 \\ 0 & 0.2 & 0.6 & 0.2 \\ 0 & 0.1 & 0.7 & 0.2 \end{bmatrix} \end{array}$$

Compute the probability for sequence of observation

$$\vec{z} = \{z_1 = s_{sun}, z_2 = s_{cloud}, z_3 = s_{cloud}, z_4 = s_{rain}, z_5 = s_{cloud}\}.$$

.

13) (5)

Consider the set of two-dimensional patterns given in Table.Q13. Use agglomerative

Table: Q13

Sample	x_1	x_2	Sample	x_1	x_2
1	1	1	8	5	1.5
2	1	2	9	4.5	2
3	2	1	10	4	4
4	2	1.5	11	4.5	4
5	3	2	12	4.5	5
6	4	1.5	13	4	5
7	4	2	14	5	5

(single-link) clustering to obtain three clusters. Draw the resulting dendrogram. Use the city-block distance between $\mathbf{x}_i$, and $\mathbf{x}_j$, which is defined as $d(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{d} |\mathbf{x}_{ik} - \mathbf{x}_{jk}|$, where d is the dimensionality of samples.

14) Consider the given binary feature vectors: (3)

$$\mathbf{x}_1 = [0\ 1\ 0\ 1\ 1\ 0\ 1\ 0]^T,\ \mathbf{x}_2 = [0\ 1\ 1\ 1\ 1\ 0\ 1\ 1\ 1]^T,\ \mathbf{x}_3 = [1\ 0\ 1\ 1\ 1\ 0\ 1\ 1]^T.$$

Determine the following:

i) Similarity for the pairs of feature vectors $\{\mathbf{x}_1, \mathbf{x}_2\}$, and $\{\mathbf{x}_2, \mathbf{x}_3\}$ using

$$s(\mathbf{x}_i, \mathbf{x}_j) = \frac{\mathbf{x}_i^T \mathbf{x}_j}{d}.$$

ii) Tanimoto distance for pairs of feature vector $\{\mathbf{x}_1, \mathbf{x}_3\}$. The Tanimoto distance is defined as

$$s(\mathbf{x}_i, \mathbf{x}_j) = \frac{\mathbf{x}_i^T \mathbf{x}_j}{\mathbf{x}_i^T \mathbf{x}_i + \mathbf{x}_j^T \mathbf{x}_j - \mathbf{x}_i^T \mathbf{x}_j}.$$

15) Let $\mathbf{x}_1 = \binom{5}{6}, \mathbf{x}_2 = \binom{2}{5}, \mathbf{x}_3 = \binom{1}{2}$, and $\mathbf{x}_4 = \binom{6}{1}$. Find trace criterion J_d for the partition $\mathcal{D}_1 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}, \mathcal{D}_2 = \{\mathbf{x}_4\}$. The trace criterion is given by (2)

$$J_d = |\mathbf{S}_W| = |\sum_{i=1}^{c} \mathbf{S}_\text{i}|.$$

-----End-----