Question Paper

Exam Date & Time: 10-Jan-2024 (02:30 PM - 05:30 PM)

MANIPAL ACADEMY OF HIGHER EDUCATION

BTech VII Semester - Makeup Examination - Jan 2024

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 block diagram depicting the design cycle of a pattern recognition system, and explain the following terms in context of design cycle: (5)

i) Data collection
ii) Feature choice
iii) Model choice
iv) Training
v) Evaluation and
vi) Computational complexity.

2) The minimum-error rate classification can be achieved by the use of discriminant function (3)

$$g_i(\mathbf{x}) = \ln p(\mathbf{x}|\omega_i) + \ln P(\omega_i).$$

Rewrite the given discriminnat function considering the density $p(\mathbf{x}|\omega_i)$ as a multivariate normal, that is $p(\mathbf{x}|\omega_i) \sim \mathcal{N}(\boldsymbol{\mu}_i, \Sigma)$ and obtain the linear discriminant function for $\Sigma_i = \sigma^2 I$.

3) Consider the three-dimensional normal distribution $p(\mathbf{x}|\omega) \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where (2)

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

Find the probability density at the point $\mathbf{x}_0 = [0.5\ 2\ 0.8]^T$.

4) Let x have an exponential density (5)

$$p(x|\theta) = \begin{cases} \theta \exp(-\theta x) & x \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

i) Plot $p(x|\theta)$ versus x for $\theta = 1$

i) Plot $p(x|\theta)$ versus x for $\theta = 1$.

ii) Plot $p(x|\theta)$ versus θ, $(0 \leq \theta \leq 5)$, for $x = 2$.

iii) Suppose that n samples $x_1, \ldots, x_n$ are drawn independently according to $p(x|\theta)$. Show that the maximum-likelihood estimate for θ is given by

$$\hat{\theta} = \frac{1}{\frac{1}{n}\sum_{k=1}^{n} x_k}.$$

5) The purpose of this problem is to derive the Bayesian classifier for the d dimensional multivariate Bernoulli case. Work with each class separately, interpreting $P(\mathbf{x}|\mathcal{D})$ to mean $P(\mathbf{x}|\mathcal{D}_i, \omega_i)$. Let the conditional probability for a given category be given by (3)

$$P(\mathbf{x}|\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{x_i}(1-\theta_i)^{1-x_i},$$

and let $\mathcal{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ be a set of n samples independently drawn according to this probability density. Assuming a uniform *a priori* distribution for $\boldsymbol{\theta}$ and using the identity

$$\int_0^1 \theta^m (1-\theta)^n d\theta = \frac{m!n!}{(m+n+1)!},$$

show that

$$p(\boldsymbol{\theta}|\mathcal{D}) = \prod_{i=1}^{d} \frac{(n+1)!}{s_i!(n-s_i)!}\theta_i^{s_i}(1-\theta_i)^{n-s_i}.$$

Note that $P(\mathcal{D}/\boldsymbol{\theta}) = \prod_{i=1}^{d} \theta_i^{s_i}(1-\theta_i)^{n-s_i}$, and $\mathrm{s} = (s_1, \ldots, s_d)^T$ is the sum of n samples.

6) When do maximum-likelihood and Bayes methods differ? (2)

7) Describe various types of hierarchical clustering techniques. (5)

8) (3)

Consider the criterion function

$$J(\mathbf{w}) = \frac{(\mu_1 - \mu_2)}{\sigma_1^2 + \sigma_2^2} = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_\mathbf{w} \mathbf{w}}$$

which is maximized by $\mathbf{w} = (\Sigma_1 + \Sigma_2)^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)$. Determine the maximum value of $J(\mathbf{w})$ for the data points given in Table.Q3B.

Table: Q3B

ω_1		ω_2	
x_1	x_2	x_1	x_2
2	4	5	6

3	5	5	4
3	4	6	5
3	3	6	6

9) Write the pseudocode for PNN training. (2)

10) Consider a HMM, $\boldsymbol{\theta} = (A, B)$, where A, and B are the transition and emission probability matrices which are given by (5)

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.2 & 0.3 & 0.1 & 0.4 \\ 0.2 & 0.5 & 0.2 & 0.1 \\ 0.8 & 0.1 & 0.0 & 0.1 \end{pmatrix}$$

and

$$B = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0.3 & 0.4 & 0.1 & 0.2 \\ 0 & 0.1 & 0.1 & 0.7 & 0.1 \\ 0 & 0.5 & 0.2 & 0.1 & 0.2 \end{pmatrix}.$$

Assume that the initial state at $t = 0$ be ω_1. What is the probability that the model generates the sequence $\mathbf{V}^7 = \{v_2, v_1, v_3, v_2, v_2, v_4, v_0\}$? Assume that the matrix indexes begin at 0. [Note: v_0 denotes the visible symbol emitted by the accepting/final state]

11) (3)

Consider the sum-of-error criterion function

$$J_s = \|\mathbf{Ya} - \mathbf{b}\|^2 = \sum_{i=1}^{n} (\mathbf{a}^T \mathbf{y}_i - b_i)^2.$$

Let $b_i = 1 \; \forall i$ and consider the following training samples:

$$\omega_1 : \{(1,5),\ (2,9),\ (-5,-3)\}$$
$$\omega_2 : \{(2,-3),\ (-1,-4),\ (0,2)\}.$$

Calculate the Hessian matrix for this problem.

12) State the limited horizon assumption for Markov model. (2)

13) Consider the set of two-dimensional patterns given in Table.Q5A. Use agglomerative (5)

Table: Q5A

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

(complete-link) clustering to obtain four 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 vectors: (3)

$$\mathbf{x}_1 = [5,\ 3,\ 2,\ 4]^T,\ \mathbf{x}_2 = [4,\ 1,\ 2,\ 3]^T,\ \mathbf{x}_3 = [2,\ 4,\ 3,\ 5]^T.$$

Show that for all the pair of vectors formed using $\mathbf{x}_1, \mathbf{x}_2$ and $\mathbf{x}_3$, the following identity holds:

$$d_2(\mathbf{x}_i, \mathbf{x}_j) < d_1(\mathbf{x}_i, \mathbf{x}_j) \text{ for } i, j = 1, 2, 3 (i \neq j),$$

where d_2 and d_1 are the Euclidean and Manhattan distance respectively. [Hint: $d_1(\mathbf{x}_i, \mathbf{x}_j) = \sum_{k=1}^{d} |\mathbf{x}_{ik} - \mathbf{x}_{jk}|$].

15) (2)

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 the sum-of -square error J_e criterion for the partition $\mathcal{D}_1 = \{\mathbf{x}_1, \mathbf{x}_2\}, \mathcal{D}_2 = \{\mathbf{x}_3, \mathbf{x}_4\}$. The sum-of-square criterion is defined as

$$J_e = tr[\mathbf{S}_w] = \sum_{i=1}^{c} tr[\mathbf{S}_i] = \sum_{i=1}^{c} \sum_{\mathbf{x} \in \mathcal{D}_i} \|\mathbf{x} - \mathbf{m}_i\|^2.$$

-----End-----