MTH 552: STATISTICAL & AI TECHNIQUES IN DATAMINING
Endsem Examination: Full Marks: 100

[1] (a) Consider the following 4 objects, whose measurements on a single feature variable are

Object	O_1	O_2	O_3	O_4
Feature value	2	1	5	8

Using Ward's method, obtain a hierarchical clustering of the objects and draw the dendogram indicating the mergers and the ESS value at each level of merger.

(b) Let Π_1 and Π_2 be 2 p-dimensional populations $\Pi_i \equiv N_p(\underset{\sim}{\mu_i}, \Sigma)$; $i = 1,2$; $\underset{\sim}{\mu_i} \in \mathbb{R}^p$, $\Sigma > 0$.

(i) Find the relationship between the Bhattacharya distance (between Π_1 and Π_2),

$$J_B = -\log\left(\int \cdots \int \left(f(\underset{\sim}{x} \mid \Pi_1) f(\underset{\sim}{x} \mid \Pi_2)\right)^{1/2} d\underset{\sim}{x}\right),$$

and the Mahalanobis square distance $\Delta^2 = \left(\underset{\sim}{\mu_1} - \underset{\sim}{\mu_2}\right)^T \Sigma^{-1} \left(\underset{\sim}{\mu_1} - \underset{\sim}{\mu_2}\right)$.

(ii) Suppose $p(\Pi_1) = 1/3$ and $p(\Pi_2) = 2/3$. Prove or disprove: "The TPM corresponding to Fisher Linear Discriminant function based classifier is given by $\Phi(-\Delta/2)$, where $\Phi(.)$ denotes the distribution function of a standard normal distribution".

24 (8+8+8) marks

[2] (a) Let $(X_1, \ldots, X_n)$ be random sample from a population having a mixture normal density

$$p(x) = \sum_{j=1}^{g} \pi_j \, p(x \mid \underset{\sim}{\theta_j}),$$

where, $p(x \mid \underset{\sim}{\theta_j})$ is the density of $N(\mu_j, \sigma_j^2)$, $\underset{\sim}{\theta_j} = (\mu_j, \sigma_j)$ $j = 1, \ldots, g$. Formulate the maximum likelihood estimation of the parameters involved in the E-M framework and derive the E-M algorithm update equations for π_j and $\underset{\sim}{\theta_j}$ and hence outline the density estimation procedure based on a given set of observations.

(b) Present a detailed formulation of a basic SVM model for classification of a 2-class, linearly separable problem. Sketch the feature space (for a 2-dimensional feature vector), indicating clearly the separating hyperplane, the cannonical hyperplanes, the margin and the support vectors under the SVM setup.

18 (10+8) Marks

[3] Consider a 2-class (Π_1 and Π_2) classification problem where the class conditional densities are given by:

$$f(x \mid \Pi_1) = \begin{cases} 1, & 0 \le x \le 1, \\ 0, & \text{otherwise,} \end{cases} \quad \text{and} \quad f(x \mid \Pi_2) = \begin{cases} 0.5, & 0.5 \le x \le 2.5, \\ 0, & \text{otherwise.} \end{cases}$$

The prior probabilities of the 2 classes are $p(\Pi_1) = 0.25$ and $p(\Pi_2) = 0.75$.

(i) Find the Bayes classification rule and the corresponding total probability of misclassification.

(ii) Find the ECM of the ECM minimizing classification rule if the prior probabilities are $p(\Pi_1) = 0.25$ and $p(\Pi_2) = 0.75$ and the misclassification costs are $C(2 \mid 1) = 1$ and $C(1 \mid 2) = 0.5$.

10 marks

[4] Consider the 3 bivariate populations, Π_1, Π_2 and Π_3 with the following joint probability mass functions:

Π_1

$x_2 \backslash x_1$	1	2
1	0.5	0.2
2	0.1	0.2

Π_2

$x_2 \backslash x_1$	1	2
1	0.2	0.1
2	0.3	0.4

Π_3

$x_2 \backslash x_1$	1	2
1	0.25	0.25
2	0.25	0.25

The prior probabilities of the 3 populations are assumed to be equal and the misclassification costs are given by $C(1 \mid i) = 1, i = 2,3$; $C(2 \mid i) = 2, i = 1,3$ and $C(3 \mid i) = 3, i = 1,2$.

(a) Find the ECM minimizing classification partition.

(b) Find the misclassification rate of the rule obtained in (a) for the following pre-classified examples:

$$\ell = \{((1,2),\Pi_1),((1,1),\Pi_1),((1,1),\Pi_1),((1,1),\Pi_3),((2,2),\Pi_2),((2,1),\Pi_3),((1,2),\Pi_3),((2,2),\Pi_3)\}$$

12 (10+2) marks

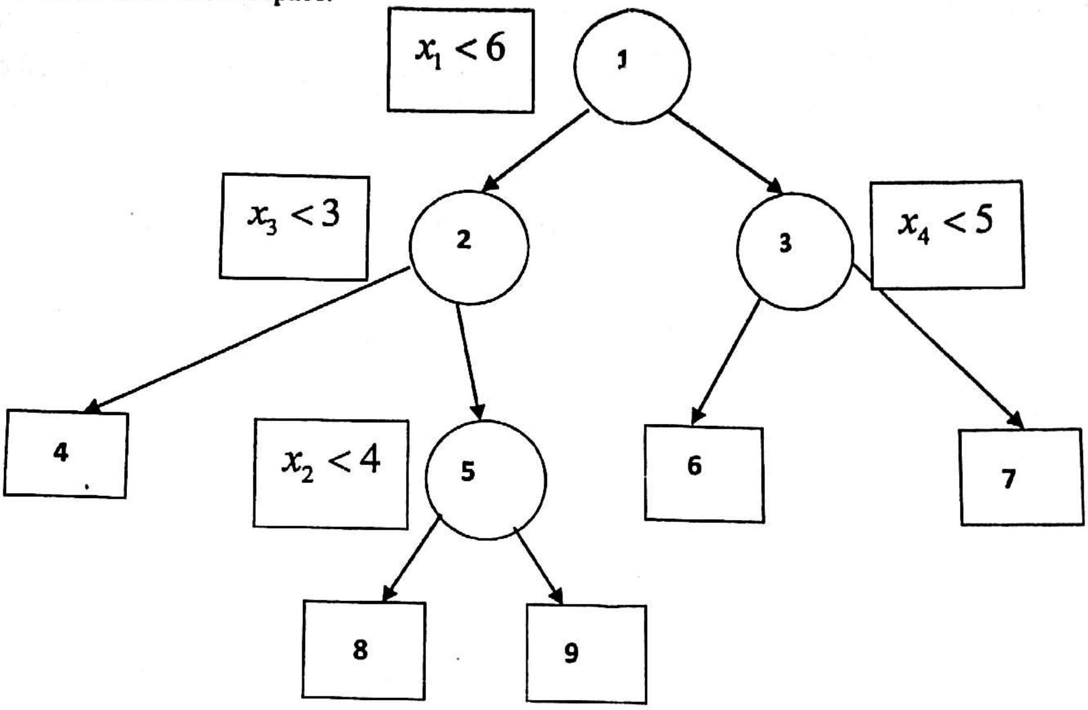
5

[4] (a) Design the architecture of a single hidden layer neural network model for a C-class $(\Pi_1,\ldots,\Pi_C)$ classification problem, indicating clearly the input and output at each of the processing units. Discuss how such a classifier operates.

(b) Consider a neural network model with NO hidden layers for a 2-class classification problem with L-dimensional feature vector. Find the relationship between the ANN classifier model and a 2-class logistic discrimination model.

12 (6+6) marks

[6] Consider the following classification tree T for a 2-class (Π_1,Π_2) problem obtained from a learning sample of size 100 with 4-dimensional feature space.



For the constructed classification tree,

t	1	2	3	4	5	6	7	8	9
$N(t)$	100	60	40	10	50	16	24	20	30
$N_2(t)$	60	40	20	0	40	16	4	15	25

$N(t)$: # of training patterns reaching node t and $N_2(t)$: # of training patterns with label π_2 reaching node t.

(a) Assign class labels to the terminal nodes.

(b) Classify the feature vector $(3,21,9,1)$ using the above tree.

(c) Find the node impurities of the terminal nodes.

(d) Find a measure of tree impurity.

(e) Find the strength of all the internal nodes.

(f) Under a weakest link pruning approach, obtain the first pruned subtree, T_1.

(g) Which of the two trees, T_1 or T, is preferable if the cost of complexity per node, α, is 0.01.

24 (2+2+2+3+8+3+4) Marks

[Use misclassification error at node t as it's impurity measure wherever required, i.e.

$$\text{Imp}(t) = \left(\sum_{x_i \in U(t)} I\left(y_i \neq j^*(t)\right)\right) \Big/ N(t), \text{ where } j^*(t) = \arg\max_i \hat{p}(\pi_i \mid t)]$$