MTH 552A: Quiz #2
Full Marks 10

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

(a) If the prior probabilities are $p(\Pi_1) = p(\Pi_2) = 0.3$ and $p(\Pi_3) = 0.4$; find the TPM minimizing partition.

(b) Find the TPM of the rule obtained in (a), with prior probabilities as in (a).

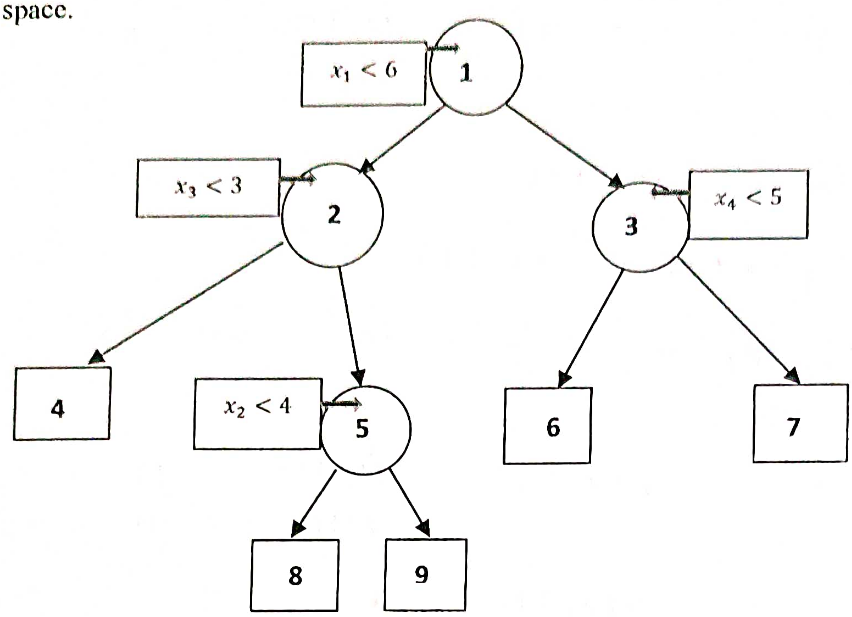
(c) If the prior probabilities of the 3 populations are assumed to be equal and the misclassification costs are given by $C(1|i) = 1, i = 2,3$; $C(2|i) = 2, i = 1,3$ and $C(3|i) = 3, i = 1,2$; find the ECM minimizing classification partition.

(d) Find the ECM of the rule obtained in (c), with prior probabilities as in (c).

2+2+3+3

MTH 552A: Quiz #3
Full Marks 10

Consider the following classification tree T for a 2-class (π_1, π_2) problem obtained from a learning sample of size 100 with 5-dimensional feature space.



For the constructed classification tree; $N(1) = 100, N(2) = 60, N(3) = 40, N(4) = 10, N(5) = 50,\ N(6) = 16,$ $N(7) = 24, N(8) = 20, N(9) = 30$ and $N_2(1) = 60, N_2(2) = 40, N_2(3) = 20, N_2(4) = 0, N_2(5) = 40, N_2(6) =$ $16,\ N_2(7) = 4, N_2(8) = 15, N_2(9) = 25$; where $N(t)$ is the number of training patterns reaching node t and $N_2(t)$ is the number of training patterns with label π_2 reaching node t.

(a) Assign class labels to the terminal nodes of ***T***.
(b) Classify the feature vector $(5,18,7,3,12)$ using the above tree.
(c) Find the node impurities of the terminal nodes.
(d) Find a measure of tree impurity.
(e) Find a measure of change in impurity function due to the split at node 2.
(f) Find pure nodes, if any, of ***T***.
(g) Find Gini index of node 5. **1+1+1+2+3+1+1**

Note: **Use misclassification error rate at node *t* as it's impurity measure wherever required,**

i.e. $Imp(t) = \sum_{i:x_i \in U(t)} I(y_i \neq j^*(t)) \big/ N(t)$; where $j^*(t) = \arg\max_i p(\pi_i | t)$.

MTH 552: STATISTICAL & AI TECHNIQUES IN DATA MINING
End semester Examination: Full Marks 100

[1] Let the covariance matrix of $p \times 1$ $(p > 1)$ random vector $\underline{X}$ be $\Sigma = (\sigma_{ij})$; where $\sigma_{ii} = 1$ for all $i = 1, \ldots, p$ and $\sigma_{ij} = \rho$ for all $i \neq j$ and $i, j = 1, \ldots, p$.

(a) Prove or disprove "$\Sigma > 0$ for any positive integer p (> 1) iff $-1 < \rho < 1$".
(b) Prove or disprove "$\Sigma > 0$ for any positive integer p (> 1) iff $-1/(p-1) < \rho < 1$".
(c) Suppose $\rho = 0.5$, find the proportion of total variation in $\underline{X}$ explained by the first principal component derived from the covariance matrix of $\underline{X}$.
(d) Prove or disprove "for $\Sigma > 0$ the generalized variance of $\underline{X}$ is equal to the generalized variance of $\underline{Y}$; where $\underline{Y} = (Y_1, \ldots, Y_p)'$, $Y_1, \ldots, Y_p$ are the p principal components derived from the correlation matrix of $\underline{X}$".

12 marks

[2] The distance matrix corresponding to 5 multidimensional cases C_1, C_2, C_3, C_4, C_5 is given by

$$D = \begin{pmatrix} 0 & & & & \\ 10 & 0 & & & \\ 2 & 3 & 0 & & \\ 5 & 4 & 6 & 0 & \\ 8 & 12 & 11 & 7 & 0 \end{pmatrix}$$

Construct the dendogram tree corresponding to an agglomerative average linkage hierarchical clustering algorithm.

8 marks

[3] Consider the learning sample $\mathcal{L} = \{((1,2), \pi_1), ((2,3), \pi_2), ((3,2), \pi_2), ((-1,2), \pi_1)\}$ for 2-class (π_1, π_2) classification problem.

(a) Are the training patterns linearly separable?
(b) Sketch the solution region in the weight vector space of perceptron learning rule based linear classifier (without constant and without margin).
(c) Using the instantaneous mode perceptron learning rule for linear classifier (without constant and without margin) weight vector updation equation $\underline{w}_{k+1} = \underline{w}_k + 0.5\, \underline{z}_i$ ($\underline{z}_i$ is a pattern vector requiring updation); obtain the first 3 steps of iteration of the weight vector, starting from the initial weight vector $\underline{w}_0 = (1, 2)'$ and presenting the learning patterns sequentially. Is the updated weight vector, after the 3 steps, in the solution region?

16 (2+7+7) marks

[4] Let π_1 and π_2 be 2 p-dimensional populations, $\pi_i \equiv N_p(\underline{\mu}_i, \Sigma)$, $i = 1,2$; $\underline{\mu}_i \in \mathbb{R}^p, \Sigma > 0$. Let $\Delta^2 = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2)$ denote the Mahalanobis square distance between π_1 and π_2 and $J_B = -\log_e \left(\int \ldots \int \left(f(\underline{x}|\pi_1)\, f(\underline{x}|\pi_2) \right)^{1/2} \prod_{i=1}^{p} dx_i \right)$ denote the Bhattacharya distance between π_1 and π_2.

(a) Find the relationship between the J_B and Δ.
(b) Suppose the prior probabilities of the 2 populations are $p(\pi_1) = 1/3$ and $p(\pi_2) = 2/3$. Prove or disprove the statement : "The total probability of misclassification (TPM) corresponding to Bayes classifier is given by $\Phi(\Delta/2)$, where $\Phi(.)$ denotes the distribution function of a standard normal distribution".

14 (7+7) marks

[5] Consider a 3-class (π_1, π_2 and π_3) classification problem where the class conditional densities are given by: $f_1(x|\pi_1) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{otherwise}, \end{cases}$ $f_2(x|\pi_2) = \begin{cases} 2\, e^{-2x}, & x > 0 \\ 0, & \text{otherwise}, \end{cases}$ and

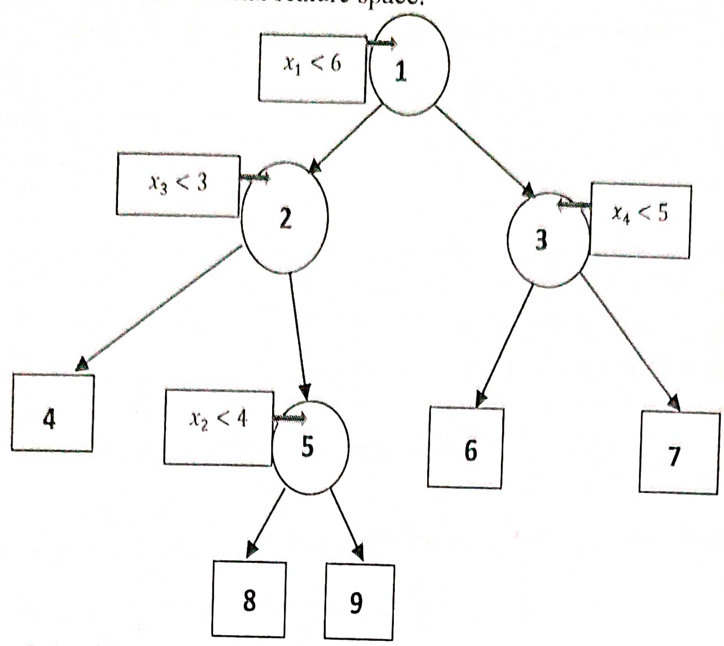
$$f_3(x|\pi_3) = \begin{cases} 3\, e^{-3x}, & x > 0 \\ 0, & \text{otherwise}. \end{cases}$$

The prior probabilities are such that $p(\pi_1) = p(\pi_2) = p(\pi_3)$.

(a) Find the TPM minimizing classification partition.
(b) Find $P(\pi_2|\pi_1)$.
(c) Find the TPM of the rule obtained in (a).
($\log_e 2 = 0.693, \log_e 3 = 1.099, \log_e 5 = 1.609, \log_e 7 = 1.946$)

16 (5+3+8) marks

[6] Consider the following classification tree T for a 3-class (π_1, π_2, π_3) problem obtained from a learning sample of size 100 with 5-dimensional feature space.



For the constructed classification tree; $N(1) = 100, N(2) = 60, N(3) = 40, N(4) = 20, N(5) = 40,$ $N(6) = 16,\ N(7) = 24, N(8) = 10, N(9) = 30$ and $N_2(1) = 40, N_2(2) = 30, N_2(3) = 10, N_2(4) = 0, N_2(5) = 30, N_2(6) = 10,\ N_2(7) = 0, N_2(8) = 1, N_2(9) = 29$ and $N_3(1) = 30, N_3(2) = 10,\ N_3(3) = 20, N_3(4) = 0, N_3(5) = 10, N_3(6) = 4, N_3(7) = 16, N_3(8) = 9, N_3(9) = 1$; where $N(t)$ is the number of training patterns reaching node t and $N_j(t)$ is the number of training patterns with label π_j reaching node t.

(a) Find the strength of all the internal modes.

(b) Under the weakest link pruning approach, obtain the first pruned subtree, T_1, of T.

(c) Which of the 2 trees, T_1 or T, is preferable, if the cost of complexity per node, α, is (i) 0.07 and (ii) 0.05.

Note: Use misclassification error rate at node *t* as it's impurity measure wherever required,

i.e. $Imp(t) = \sum_{i:x_i \in U(t)} I(y_i \neq j^*(t)) \Big/ N(t)$; where $j^*(t) = \arg\max_i p(\pi_i|t)$. **17 (8+3+6) marks**

[7] Suppose the trained set of weights of a 2-2-2 (2 inputs at input layer-single hidden layer with 2 neurons-single output layer with 2 neurons) feedforward neural network model using identity transfer function at all the hidden unit nodes, designed for a 2-class (π_1, π_2) classification problem, is given by;

Input to hidden layer weights: $W = \begin{pmatrix} 1 & 2 \\ 3 & 2 \end{pmatrix}$; w_{ij} is the weight connecting i^{th} input node and j^{th} hidden layer neuron.

Hidden to output layer weights: $B = \begin{pmatrix} 2 & 4 \\ 3 & 1 \end{pmatrix}$; $b_{ij}$is the weight connecting i^{th} hidden layer node and j^{th} output node.

If node #1 at the output layer is attached to the π_1 class and node #2 attached to the π_2 class, predict the class membership of the feature vector $\underline{x}^0 = (3,4)'$ using the trained network. **7 marks**

[8] Consider the following transactions database with 5 records

Trans-ID	Items
C_1	Hummus, Wine, Egg
C_2	Chips, Wine, Nut
C_3	Hummus, Chips, Wine, Nut
C_4	Chips, Nut
C_5	Hummus, Chips, Wine, Nut

Apply apriori algorithm to derive all association rules satisfying minimum support level of 60% and minimum confidence level 80%. Explain the steps and mention clearly where exactly apiori algorithm is being used for generating rules.

10 marks