23CSPC306

USN | | | | | | | | | | |

MANGALORE INSTITUTE OF TECHNOLOGY & ENGINEERING

(A Unit of Rajalaxmi Education Trust®, Mangalore)
Autonomous Institute affiliated to VTU, Belagavi, Approved by AICTE, New Delhi
Accredited by NAAC with A+ Grade & ISO 9001:2015 Certified Institution

Model Question Paper

Sixth Semester B.E Degree Examination

Machine Learning

Time: 3 Hours (180 Minutes) **Max. Marks: 100**

Note: 1. Answer any FIVE full questions, choosing ONE full question from each module.
2. M: Marks, L: RBT (Revised Bloom's Taxonomy) level, C: Course outcomes.

<table>
<tr><th colspan="3">Module -1</th><th>M</th><th>L</th><th>C</th></tr>
<tr><td rowspan="2">Q1</td><td>a.</td><td>List out and briefly explain the types of Machine Learning.</td><td>7</td><td>L2</td><td>CO1</td></tr>
<tr><td>b.</td><td>Illustrate the Machine Learning process model with suitable diagram.</td><td>7</td><td>L2</td><td>CO1</td></tr>
<tr><td></td><td>c.</td><td>Consider the set V = {80,90, 92, 94}.
i) Apply Min-Max procedure and map the values to a new range 0–1
ii) Convert the set to z-score</td><td>6</td><td>L3</td><td>CO1</td></tr>
<tr><td colspan="6">OR</td></tr>
<tr><td rowspan="2">Q2</td><td>a.</td><td>List and explain the visualization aids available for univariate data analysis with example for each.</td><td>7</td><td>L2</td><td>CO1</td></tr>
<tr><td>b.</td><td>List and explain 6V's of Big Data</td><td>7</td><td>L2</td><td>CO1</td></tr>
<tr><td></td><td>c.</td><td>For a given univariate dataset S= {5, 10, 15, 20, 25, 30} of marks. Find five-point summary and plot the box chart.</td><td>6</td><td>L3</td><td>CO1</td></tr>
<tr><td colspan="6">Module- 2</td></tr>
<tr><td rowspan="2">Q3</td><td>a.</td><td>Write Find-S algorithm. Apply the algorithm to obtain the hypothesis for the dataset given in Table Q.3(a)
<table>
<tr><th>Sky</th><th>Air temp</th><th>Humidity</th><th>Wind</th><th>Water</th><th>Forecast</th><th>Enjoy sport</th></tr>
<tr><td>Sunny</td><td>Warm</td><td>Normal</td><td>Strong</td><td>Warm</td><td>Same</td><td>YES</td></tr>
<tr><td>Sunny</td><td>Warm</td><td>High</td><td>Strong</td><td>Warm</td><td>Same</td><td>YES</td></tr>
<tr><td>Rainy</td><td>Cold</td><td>High</td><td>Strong</td><td>Warm</td><td>Change</td><td>NO</td></tr>
<tr><td>Sunny</td><td>Warm</td><td>High</td><td>Strong</td><td>Cool</td><td>Change</td><td>YES</td></tr>
</table>
Table Q.3(a)</td><td>10</td><td>L3</td><td>CO2</td></tr>
<tr><td>b.</td><td>Apply candidate elimination algorithm on the dataset given in Table Q.3(b) to obtain the complete version space.
<table>
<tr><th>CGPA</th><th>Interactiveness</th><th>Practical knowledge</th><th>Communication skills</th><th>Logical thinking</th><th>Job offer</th></tr>
<tr><td>≥ 9</td><td>Yes</td><td>Excellent</td><td>Good</td><td>Fast</td><td>YES</td></tr>
<tr><td>≥ 9</td><td>Yes</td><td>Good</td><td>Good</td><td>Fast</td><td>YES</td></tr>
<tr><td>≥ 8</td><td>No</td><td>Good</td><td>Good</td><td>Fast</td><td>NO</td></tr>
<tr><td>≥ 9</td><td>Yes</td><td>Good</td><td>Good</td><td>Slow</td><td>YES</td></tr>
</table>
Table Q.3(b)</td><td>10</td><td>L3</td><td>CO2</td></tr>
<tr><td colspan="6">OR</td></tr>
<tr><td>Q4</td><td>a.</td><td>Write the Nearest-centroid classifier algorithm. Apply the same to predict the class for the given test instance (6,5) using the training dataset given in Table Q.4(a)
<table>
<tr><th>X</th><th>Y</th><th>Class</th></tr>
<tr><td>3</td><td>1</td><td>A</td></tr>
<tr><td>5</td><td>2</td><td>A</td></tr>
</table></td><td>10</td><td>L3</td><td>CO2</td></tr>
</table>

		4 | 3 | A 7 | 6 | B 6 | 7 | B 8 | 5 | B Table Q.4(a)			
	b.	Apply Weighted K-Nearest Neighbor classifier to predict whether the patient is diabetic or not using K = 3 with the given features in Table Q.4(b). Test Example: BMI = 43.6, Age = 40. Predict the Sugar value? **BMI** 33.6 26.6 23.4 43.1 35.3 35.9 36.7 25.7 23.3 31 **Age** 50 30 40 67 23 67 45 46 29 56 **Sugar** 1 0 0 0 1 1 1 0 0 1 Table Q.4(b).	10	L3	CO2
		Module – 3			
Q5	a.	Using the dataset given in Table Q.5(a), compute entropy and information gain and determine the root node using the ID3 algorithm. **Outlook** | **Humidity** | **Play** Sunny | High | No Sunny | Normal | Yes Overcast | High | Yes Rain | High | Yes Rain | Normal | No Table Q.5(a), Construct the first level of the decision tree.	10	L3	CO3
	b.	Apply CART algorithm for the dataset given in Table Q.5(b), and predict whether a student get a job offer or not based on their CGPA **CGPA** >=9 >=8 >=9 <8 >=8 >=9 <8 >=9 >=8 >=8 **Job Offer** Yes Yes No No Yes Yes No Yes Yes Yes Table Q.5(b)	10	L3	CO3
		OR			
Q6	a.	Using Naïve Bayes classifier, determine whether a student will get a job offer for the given details: P(Offer) = 0.6 P(No Offer) = 0.4 P(Good Communication | Offer) = 0.75 P(Good Communication | No Offer) = 0.25 Compute posterior probabilities and justify the classification.	10	L3	CO3
	b.	Apply the Linear Regression for the dataset given in Table Q.6(b) to predict the BMI with the given Age (Target data) =45 using Matric form. (i.e Age=45, BMI=?) **Age** 50 30 40 67 23 **BMI** 33 26 23 43 35 Table Q.6(b)	10	L3	CO3
		Module – 4			
Q7	a.	Apply the Perceptron Learning Algorithm to the training dataset given in Table Q.7(a) with learning rate = 1 and initial weights = 0. Determine the final weights after one complete epoch. **x1** | **x2** | **Target** 0 | 0 | 0 0 | 1 | 0 1 | 0 | 0	10	L3	CO4

		1 | 1 | 1 Table Q.7(a)			
	b.	A neural network consists of two input neurons and one output neuron with a sigmoid activation function. Apply the concepts of forward propagation and backpropagation to perform one complete training iteration using suitable initial weights and a learning rate of 0.5. Compute the network output, determine the error, and update the weights accordingly. Show all necessary calculations.	10	L3	CO4
		OR			
Q8	a.	The hyperplane function for two variables is $b + a_1x_1 + a_2x_2$. If two hyperplanes given are for classifier 1 as is $5+2x_1 + 5x_2$ and is $5+20x_1 + 50x_2$ for classifier 2. i) Apply the distance formula and calculate the distance error function for each classifier. ii) Determine the margin for both classifiers using the hyperplane parameters.	10	L3	CO4
	b.	Apply the concept of Support Vector Machine to draw an optimal hyperplane to classify the given points. Points (4,1), (4, -1) and (6,0) belong to class positive and points (1,0), (0,1) and (0,-1) belong to negative class.	10	L3	CO4
		Module – 5			
Q9	a.	Two objects have coordinates (0, 3) and (5, 8) Apply the appropriate distance metrics to compute the Manhattan distance and Chebyshev distance between the given objects.	10	L3	CO5
	b.	Consider a perceptron to represent the Boolean function and with the initial weights w1= 0.3, w2 = -0.2, learning rate (α) = 0.2 and bias (ø) = 0.4. The activation function used here is the step function f(x) which gives the output value as binary, i.e. 0 or 1. If value of f(x) is greater than or equal to 0, it outputs 1 or else it output 0. Design a perceptron that performs the Boolean function and update the weights until the Boolean function gives the desired output.	10	L3	CO5
		OR			
Q10	a.	Consider the following set of data given in Table Q.10(a). Cluster it using K-means algorithm with initial value of objects 2 and 5 with the coordinate values (4, 6) and (12, 4) as initial seeds. **Objects** | **X-coordinate** | **Y-coordinate** 1 | 2 | 4 2 | 4 | 6 3 | 6 | 8 4 | 10 | 4 5 | 12 | 4 Table Q.10(a).	10	L3	CO5
	b.	Consider the following set of data given in the Table Q.10(b). Apply Hierarchical clustering algorithm and show the clustering steps. **Points** | **A1** | **A2** | **A3** | **A4** | **A5** (x, y) | (1,4) | (2, 6) | (11, 11) | (6, 9) | (6, 4) Table Q.10(b)	10	L3	CO5

-*****-