Reg. No.										

MANIPAL INSTITUTE OF TECHNOLOGY
MANIPAL
(A constituent unit of MAHE, Manipal)

DEPARTMENT OF MECHATRONICS
II SEMESTER M.TECH. (INDUSTRIAL AUTOMATION & ROBOTICS)
END SEMESTER EXAMINATIONS, MAY 2024
SUBJECT: ARITIFICAL INTELLGENCE AND EXPERT SYSTEMS [MTE 5002]
Date: 05 May 2024

Time: 3 Hours **MAX. MARKS: 50**

Instructions to Candidates:
- Answer **ALL** the questions.
- Missing data can be assumed and suitably justified.

Q. No	Question	M	CO	PO	LO	BL
1A.	In a credit card fraud detection system, out of 80 fraudulent transactions, the system correctly identifies 70 as fraud and misses 5. It correctly classifies 900 non-fraudulent transactions as not fraud, but incorrectly flags 10 as fraud. Illustrate the confusion matrix and calculate the precision, recall, and F1 score of the fraud detection system.	5	**1**	**3**	**2**	**3**
1B.	A man is known to speak truth 3 out of 4 times. He throws a die and reports that it is a six. Examine the probability that it is actually a six.	3	**1**	**3**	**2**	**4**
1C.	Explain how Bayesian networks enhance predictive modeling in machine learning by leveraging conditional independence statements and probabilistic reasoning to represent complex relationships among variables?	2	**1**	**3**	**1**	**4**
2A.	A Bayesian network is used to model the relationship between a patient's symptoms, medical history, and the likelihood of having a particular disease. The network has three variables: A: The patient's medical history (positive or negative) B: The patient's symptoms (present or not present) C: The likelihood of having the disease (high, medium, or low) The conditional probability table for the network is as follows: P(B=present|A=positive, C=high) = 0.9 P(B=present|A=positive, C=medium) = 0.6 P(B=present|A=positive, C=low) = 0.3 P(B=present|A=negative, C=high) = 0.6 P(B=present|A=negative, C=medium) = 0.3 P(B=present|A=negative, C=low) = 0.1 P(C=high|A=positive) = 0.5 P(C=medium|A=positive) = 0.4 P(C=low|A=positive) = 0.1 P(C=high|A=negative) = 0.2 P(C=medium|A=negative) = 0.5 P(C=low|A=negative) = 0.3 Given this information, Estimate the probability of a patient having a high likelihood of having the disease (C=high) if they have symptoms (B=present).	5	**1**	**1**	**2**	**5**
2B.	You have performed k-means clustering on a dataset containing customer spending data. The dataset consists of features such as annual income and spending score. After clustering the data into k clusters, you want to estimate the silhouette coefficient for cluster 2, point B2 by evaluating cohesion and separation distance between the clusters.	3	**1**	**3**	**2**	**5**

	Assume that the dataset has been clustered into 3 clusters (k=3), and the cluster labels for each data point are given under table. Table 1: Cluster labels for the given datapoints **Cluster 1**: Point A1: (2,5); Point A2:(3,4); Point A3:(4,6) **Cluster 2**: Point B1:(8,3); Point B2: (9,2); Point B3: (10,5) **Cluster 3**: Point C1: (6,10); Point C2: (7,8); Point C3: (8,9)					
2C.	Consider two data points x=$\begin{bmatrix}1\\2\end{bmatrix}$; y=[2 3] with C, σ and q = 1, Estimate linear, Non homogenous and homogenous kernel calculation in SVM.	2	**1**	**3**	**2**	**4**
3A.	Apply Genetic algorithm to maximize the value of the function $f(x) = -x^2 + 2x$. Given that the initial population is 11010, 00111, 10110 and 00101 with respective probabilities 0.4, 0.15, 0.7 and 0.9. Select the crossover between first and fifth digits. Determine the best value for a function after two iterations.	5	**2**	**3**	**2**	**5**
3B.	Discuss the applications of Genetic Algorithms (GAs) in the fields of engineering design and computer-aided molecular design. Discuss the challenges and potential benefits of using GAs in these areas, and provide examples of specific problems that have been addressed using this approach.	3	**2**	**1**	**1**	**5**
3C.	Explain the concept of "inertia weight" in Particle Swarm Optimization (PSO) and examine it affects the balance between exploration and exploitation in the algorithm.	2	**2**	**1**	**1**	4
4A.	Consider a neural network with one input layer, one hidden layer, and one output layer. The input layer has 3 neurons, the hidden layer has 4 neurons, and the output layer has 2 neurons. The activation function used in both the hidden and output layers is the sigmoid function. Suppose we have a training example where the input features are [0.5, 0.8, 0.2], and the corresponding target output is [0.4, 0.7]. The weights and biases of the neural network are initialized as follows: 1. Input to Hidden Layer Weights: $\mathbf{W}_{ih} = \begin{bmatrix}0.1 & 0.2 & 0.3 & 0.4\\0.2 & 0.3 & 0.4 & 0.5\\0.3 & 0.4 & 0.5 & 0.6\end{bmatrix}$ 2. Hidden layer biases: $\boldsymbol{b}_{\mathrm{h}} = \begin{bmatrix}0.1\\0.2\\0.3\\0.4\end{bmatrix}$ 3. Hidden to output layer weights: $\boldsymbol{W_{ho}} = \begin{bmatrix}0.5 & 0.6\\0.6 & 0.7\\0.7 & 0.8\\0.8 & 0.9\end{bmatrix}$ 4. Output Layer Weights: $\boldsymbol{b_o} = \begin{bmatrix}0.5\\0.6\end{bmatrix}$ The network uses the mean squared error (MSE) loss function for training. Perform one step of backpropagation to evaluate the weights and biases of the network using stochastic gradient descent (SGD) with a learning rate of 0.1.	**5**	**3**	**1**	**2**	**5**
4B.	A data scientist working for a cybersecurity firm, tasked with developing a machine learning model to detect malicious network traffic and prevent cyberattacks. After considering various algorithms, you decide to use Random Forest due to its ability to handle high-dimensional data and provide robust classification performance. Anticipate the drawbacks occurred as you deploy and evaluate the Random Forest model in a real-world cybersecurity environment.	**3**	**3**	**1**	**2**	**6**
4C.	Imagine, you're developing a neural network for detecting fraudulent transactions in a banking system. The neural network architecture includes one input layer with 50 neurons (representing transaction features), one hidden layer with 30 neurons, and one output neuron indicating whether the transaction is fraudulent or not. Compare and contrast the applicability of backpropagation and reinforcement learning in training this neural network, considering factors such as data availability, feedback mechanism, and computational efficiency.	**2**	**3**	**1**	**2**	**5**

<table>
<tr>
<td>5A.</td>
<td>A robot is navigating through a grid-based environment to reach its destination. The grid consists of obstacles and open spaces, and the robot needs to make decisions at each grid cell to determine its next move. You decide to use a decision tree classifier with the Gini index as the splitting criterion to assist the robot in path planning.
Given a sample dataset containing the following features and labels:
Evaluate the Gini index for each feature and determine the optimal feature and split point for the root node of the decision tree based on the provided dataset: Pictorially represent the final decision tree for the given dataset under table 1.

Table 1: Dataset for grid-based environment
<table>
<tr><th>Grid Cell</th><th>Obstacle</th><th>Terrain Type</th><th>Visibility</th><th>Action</th></tr>
<tr><td>1</td><td>Clear path</td><td>Open Space</td><td>Clear</td><td>Move Forward</td></tr>
<tr><td>2</td><td>Solid Rocks</td><td>Rough Terrain</td><td>Partially Obscured</td><td>Turn Left</td></tr>
<tr><td>3</td><td>Water bodies</td><td>Obstacle</td><td>Obscured</td><td>Stop</td></tr>
<tr><td>4</td><td>Clear path</td><td>Open Space</td><td>Clear</td><td>Move Forward</td></tr>
<tr><td>5</td><td>Solid rock</td><td>Rough Terrain</td><td>Partially Obscured</td><td>Turn Right</td></tr>
</table>
</td>
<td>5</td><td>4</td><td>3</td><td>2</td><td>5</td>
</tr>
<tr>
<td>5B.</td>
<td>In the field of healthcare robotics, you've been tasked with designing a robotic system to assist elderly individuals with daily living activities in nursing homes. The robotic system is intended to provide support with tasks such as medication reminders, mobility assistance, and social interaction. Your goal is to develop a user-friendly and effective robotic solution that improves the quality of life for elderly residents while addressing their unique needs and preferences.
Formulate the design procedures of robotic systems to be incorporated in a healthcare robotics system to enhance the quality of life for elderly individuals in nursing homes, considering factors such as user acceptance, autonomy, safety, and ethical considerations?</td>
<td>3</td><td>5</td><td>5</td><td>7</td><td>6</td>
</tr>
<tr>
<td>5C.</td>
<td>In a swarm robotics simulation, a neural network controller is trained using a genetic algorithm (GA). The genetic algorithm initializes a population of 50 neural network controllers with random weights and biases. Each controller is evaluated based on its performance in completing a specific task, and the top 20 controllers are selected for crossover and mutation. If the crossover rate is 0.7 and the mutation rate is 0.1, determine the number of offspring produced in the next generation.</td>
<td>2</td><td>5</td><td>3</td><td>17</td><td>5</td>
</tr>
</table>