Roll No.											

PRESIDENCY UNIVERSITY

BENGALURU

End - Term Examinations – MAY 2025

Date: 27-05-2025 **Time:** 09:30 am – 12:30 pm

School: SOCSE	**Program:** B. Tech	
Course Code: CSE3011	**Course Name:** REINFORCEMENT LEARNING	
Semester: VI	**Max Marks**: 100	**Weightage**: 50%

CO – Levels	**CO1**	**CO2**	**CO3**	**CO4**	**CO5**
Marks	**26**	**26**	**24**	**24**	

Instructions:

(i) Read all questions carefully and answer accordingly.

(ii) Do not write anything on the question paper other than roll number.

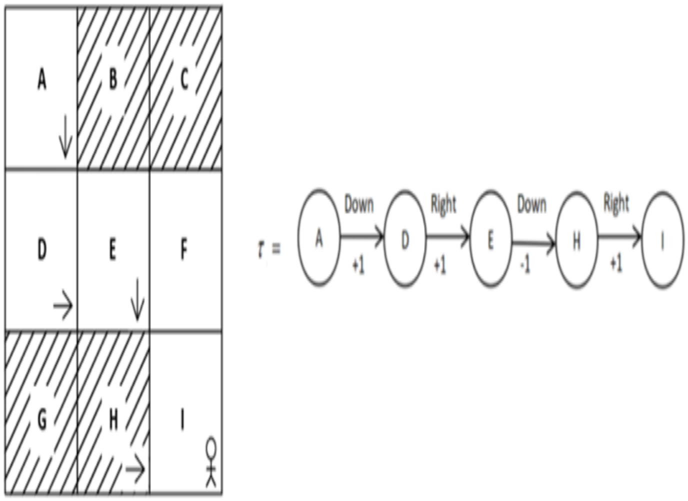
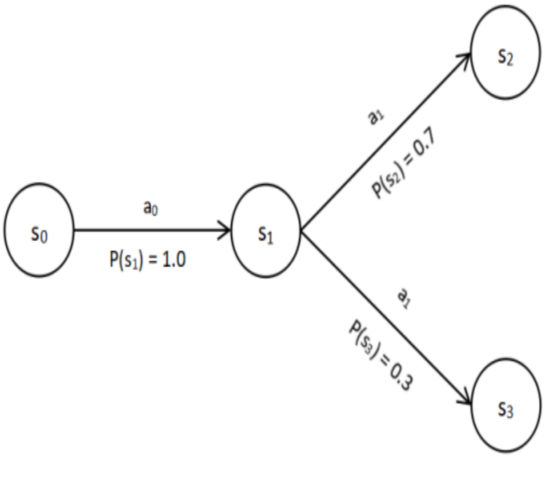
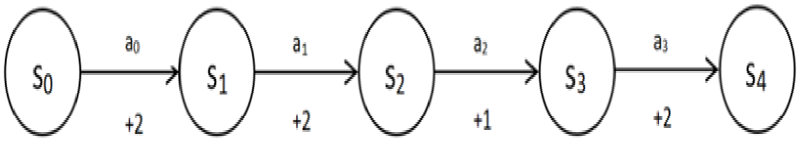
Part A

Answer ALL the Questions. Each question carries 2marks. 10Q x 2M=20M

1.	Explain the elements of RL	**2 Marks**	**L2**	**CO1**
2.	Define 'reward' and 'return' for an episodic task with an example for each	**2 Marks**	**L1**	**CO1**
3.	Define a) Episode b) Optimal policy	**2 Marks**	**L1**	**CO1**
4.	Define Q function	**2 Marks**	**L1**	**CO2**
5.	Write any two differences between on-policy and off policy TD Control algorithms	**2 Marks**	**L2**	**CO3**
6.	Define Value function	**2 Marks**	**L1**	**CO2**
7.	What is the value of the cards J,4,Q and 'Ace' in the blackjack game?	**2 Marks**	**L1**	**CO3**
8.	What is the significance of T in softmax exploration?	**2 Marks**	**L1**	**CO4**
9.	Write the equation to find V(S) in the Monte Carlo method	**2 Marks**	**L2**	**CO2**
10.	Define Thompson sampling of an arm in MAB Problem	**2 Marks**	**L1**	**CO4**

Part B

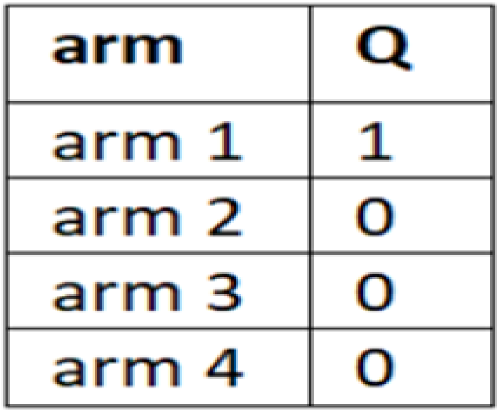
Answer theQuestions. Total Marks 80M

11.	**a.**	For the following grid world Environment Calculate the Value function which follows a deterministic policy  A B C D E F G H I τ = A Down +1 D Right +1 E Down -1 H Right +1 I  s2 a1 P(s2) = 0.7 s0 a0 P(s1) = 1.0 s1 a1 P(s3) = 0.3 s3 Figure 3.2: Transition probability of performing action a_1 in state s_1	**10Marks**	**L3**	**CO1**
	b.	Identify Bellman equation to the value function of a state in a deterministic environment and stochastic environment . Explain each term in it. Find the value of all the states in the trajectory given below using Bellman equation. Assume Y=1  S0 a0 +2 S1 a1 +2 S2 a2 +1 S3 a3 +2 S4	**10 Marks**	**L2**	**CO1**
Or					
12.	**a.**	Implement the reinforcement learning environment namely,	**10**	**L3**	**CO1**

		Frozen Lake Environment using a random policy and show the output of the following: a. Create and render the environment b. Action Space c. State Space d. Generate 20 Episodes and print Return of each episode	**Marks**		
	b.	Discuss stochastic environment and deterministic environment in RL with an example.	**10 Marks**	**L2**	**CO1**

13.	**a.**	Write a python program to find an optimal policy using Q-learning for the frozen lake environment with alpha=0.85 and gamma=0.90 and epsilon =0.8 Create and render the environment Generate policy using 20episodes with 30 timesteps Print the optimal policy	**10 Marks**	**L3**	**CO2**
	b.	Explain different types of RL environments with an example each.	**10Marks**	**L2**	**CO2**
Or					
14.	**a.**	Articulate TD Prediction algorithm in FZLE environment	**10 Marks**	**L3**	**CO2**
	b.	Discuss the appropriate situations that are suitable to apply DP,MC or TD methods to learn optimal policy	**10 Marks**	**L2**	**CO2**

15.	**a.**	Implement SARSA algorithm to learn the optimal policy in Frozen Lake environment using Python.	**10 Marks**	**L3**	**CO3**
	b.	Articulate TD Control algorithm in FZLE environment	10Marks	**L3**	**CO3**
Or					
16.	**a.**	Interpret Thompson sampling strategy to overcome the exploration – exploitation dilemma with the algorithm	**10 Marks**	**L3**	**CO3**
	b.	Compare SARSA and Q Learning exploration strategies	**10 Marks**	**L3**	**CO3**

<table>
<tr><td><b>17.</b></td><td><b>a.</b></td><td>Articulate contextual Bandits and list out the applications of MAB

<table>
<tr><th>arm</th><th>Q</th></tr>
<tr><td>arm 1</td><td>1</td></tr>
<tr><td>arm 2</td><td>0</td></tr>
<tr><td>arm 3</td><td>0</td></tr>
<tr><td>arm 4</td><td>0</td></tr>
</table></td><td><b>10 Marks</b></td><td><b>L3</b></td><td><b>CO4</b></td></tr>
</table>

	b	Briefly Explain the applications of Reinforcement Learning	**10 Marks**	**L2**	**CO4**
Or					
18.	**a.**	Compare advantages and disadvantages of Monte carlo, Dynamic Programming and Temporal Difference in detail	**10 Marks**	**L3**	**CO4**
	b	Demonstrate Upper Confidence Bound to overcome the exploration –exploitation dilemma with the algorithm	**10Marks**	**L2**	**CO4**