Number of printed pages - 2

Semester VII (B.Tech.)

Er. No. 20.1B30.8

Academic Year: 2023-24

Jaypee University of Engineering & Technology, Guna

T-3 (Odd Semester 2023)

18B14CI853 - REINFORCEMENT LEARNING

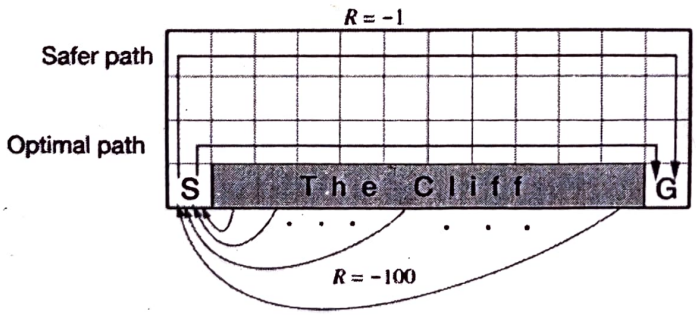
Maximum Duration: 2 Hours

Maximum Marks: 35

Notes:

1. This question paper has 7 questions.
2. Write relevant answers only.
3. Do not write anything on question paper.

		Marks	CO No.
Q1.	A gambler has the opportunity to make bets on the outcomes of a sequence of coin flips. If the coin comes up heads, he wins as many dollars as he has staked on that flip; if it is tails, he loses his stake. The game ends when the gambler wins by reaching his goal of $100, or loses by running out of money. Provide an optimal policy by using the concept of dynamic programming for the above Gambler's Problem.	[05]	CO4
Q2.	Suppose a wire frame forming a closed loop is dunked in soapy water to form a soap surface or bubble conforming at its edges to the wire frame. If the geometry of the wire frame is irregular but known, how can you compute the shape of the surface by using the concept of Monte Carlo method?	*[05]*	*CO5*
Q3.	Consider the grid-world shown below:	[05]	CO4



This is a standard undiscounted, episodic task, with start and goal states, and the usual actions causing movement up, down, right, and left. Compare the performance of Sarsa and Q-learning on the above grid and highlight the difference between on-policy (Sarsa) and off policy (Q-learning) methods.

Q4.	Show and explain the processes of Dyna-Q algorithm. Also show the architecture of Dyna agents using these processes of Dyna-Q algorithm.	**[05]**	**CO2**
Q5.	In instrumental conditioning experiments learning depends on the consequences of behavior: the delivery of a reinforcing stimulus is contingent on what the animal does. Explain the experiment performed by the American psychologist Edward Thorndike in instrumental conditioning experiments learning.	**[05]**	**CO3**
Q6.	Show an implementation of an actor–critic algorithm as an ANN with component networks implementing the actor and the critic.	**[05]**	**CO4**
Q7.	*Personalizing* web services such as the delivery of news articles or *advertisements* is one approach to increasing users' satisfaction with a website or to increase the yield of a marketing campaign. How we can use the concept of reinforcement learning in increase the yield of a marketing campaign?	**[05]**	**CO5**