



PRESIDENCY UNIVERSITY

BENGALURU

Roll No.																			
----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--

Mid - Term Examinations - MARCH 2026

Date: 10-03-2026

Time: 09:30am - 11:00am

School: SOCSE	Program: B.Tech	
Course Code: CAI2507	Course Name: Reinforcement Learning	
Semester: VI	Max Marks: 50	Weightage: 25%

CO - Levels	CO1	CO2
Marks	24	26

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 2 marks.

5Q x 2M=10M

1	How does RL differ from other ML paradigms?	2 Marks	L2	CO1
2	What is the difference between the Bellman expectation and Bellman optimality equations?	2 Marks	L2	CO1
3	What is the Monte Carlo method?	2 Marks	L1	CO2
4	How do prediction tasks differ from control tasks?	2 Marks	L2	CO2
5	Why do we use incremental mean updates?	2 Marks	L1	CO2

Part B

Answer the Questions.

Total Marks 40M

6.	a.	Explain the concepts of value function in Reinforcement Learning. Discuss their mathematical foundations of value function with suitable example.	10 Marks	L2	CO1
-----------	-----------	---	-----------------	-----------	------------

Or

7.	a.	Explain the applications of Reinforcement Learning in real-world systems. Discuss its role in Dynamic pricing, Finance, Natural Language Processing and Recommendation system.	10 Marks	L2	CO1
-----------	-----------	--	-----------------	-----------	------------

8.	a.	Using the Model dynamics table of State A find the optimal policy using Value Iteration.	10 Marks	L3	CO1																																			
<table border="1" style="margin: auto; border-collapse: collapse;"> <thead> <tr> <th style="padding: 5px;">State (<i>s</i>)</th> <th style="padding: 5px;">Action (<i>a</i>)</th> <th style="padding: 5px;">Next State (<i>s'</i>)</th> <th style="padding: 5px;">Transition Probability $P(s' s,a)$ or $P_{ss'}^a$</th> <th style="padding: 5px;">Reward Function $R(s,a,s')$ or $R_{ss'}^a$</th> </tr> </thead> <tbody> <tr><td style="text-align: center;">A</td><td style="text-align: center;">0</td><td style="text-align: center;">A</td><td style="text-align: center;">0.1</td><td style="text-align: center;">0</td></tr> <tr><td style="text-align: center;">A</td><td style="text-align: center;">0</td><td style="text-align: center;">B</td><td style="text-align: center;">0.8</td><td style="text-align: center;">-1</td></tr> <tr><td style="text-align: center;">A</td><td style="text-align: center;">0</td><td style="text-align: center;">C</td><td style="text-align: center;">0.1</td><td style="text-align: center;">1</td></tr> <tr><td style="text-align: center;">A</td><td style="text-align: center;">1</td><td style="text-align: center;">A</td><td style="text-align: center;">0.1</td><td style="text-align: center;">0</td></tr> <tr><td style="text-align: center;">A</td><td style="text-align: center;">1</td><td style="text-align: center;">B</td><td style="text-align: center;">0.0</td><td style="text-align: center;">-1</td></tr> <tr><td style="text-align: center;">A</td><td style="text-align: center;">1</td><td style="text-align: center;">C</td><td style="text-align: center;">0.9</td><td style="text-align: center;">1</td></tr> </tbody> </table> <p style="text-align: center; margin-top: 5px;">Model dynamics of state A</p>						State (<i>s</i>)	Action (<i>a</i>)	Next State (<i>s'</i>)	Transition Probability $P(s' s,a)$ or $P_{ss'}^a$	Reward Function $R(s,a,s')$ or $R_{ss'}^a$	A	0	A	0.1	0	A	0	B	0.8	-1	A	0	C	0.1	1	A	1	A	0.1	0	A	1	B	0.0	-1	A	1	C	0.9	1
State (<i>s</i>)	Action (<i>a</i>)	Next State (<i>s'</i>)	Transition Probability $P(s' s,a)$ or $P_{ss'}^a$	Reward Function $R(s,a,s')$ or $R_{ss'}^a$																																				
A	0	A	0.1	0																																				
A	0	B	0.8	-1																																				
A	0	C	0.1	1																																				
A	1	A	0.1	0																																				
A	1	B	0.0	-1																																				
A	1	C	0.9	1																																				

Or

9.	a.	Discuss Bellman optimality equation and explore how it is useful for finding the optimal Bellman value and Q functions.	10 Marks	L2	CO1
-----------	-----------	---	-----------------	-----------	------------

10.	a.	Describe the step-by-step algorithm of Monte Carlo prediction (Value function), and illustrate with a suitable example.	10 Marks	L2	CO2
------------	-----------	---	-----------------	-----------	------------

Or

11.	a.	Explain Every-Visit Monte Carlo prediction. Compare it with First-Visit MC prediction using an illustrative example.	10 Marks	L2	CO2
------------	-----------	--	-----------------	-----------	------------

12.	a.	Explain the Monte Carlo control algorithm used for finding an optimal policy in Reinforcement Learning. Describe the steps involved and the role of policy improvement.	10 Marks	L2	CO2
------------	-----------	---	-----------------	-----------	------------

Or

13.	a.	Describe the MC control using exploring starts algorithm in detail. Illustrate how the policy is improved after each episode.	10 Marks	L2	CO2
------------	-----------	---	-----------------	-----------	------------