Roll No: | | | | | | | | | | | |

PRESIDENCY UNIVERSITY

GAIN MORE KNOWLEDGE
REACH GREATER HEIGHTS

Presidency University Act, 2013 of the Karnataka Act No. 41 of 2013 | Established under Section 2(f) of UGC Act, 1956
Approved by AICTE, New Delhi | Approved By BCI
Bengaluru

Mid Term Examinations - October 2026

Date: 06/10/2026 **Time:** 02:00 PM - 03:30 PM

Course Code: CBD3413 **Course Name:** Reinforcement Learning for Big Data

Semester: Seventh Semester **Max. Marks:** 50 **Weightage:** 50%

CO - Levels	CO1	CO2
Marks	50	40

PART A: Answer Following Questions. 10 M

Q.N.	Questions	M	CO	BT
1	Differentiate between the Bellman optimality equation and the Bellman expectation equation.	2	CO1	BT2
2	An autonomous car learns to park in a parking space. It observes its position, distance from other cars, steering angle, and available space. It can accelerate, brake, or turn the steering wheel. It receives a positive reward for parking accurately and a negative reward for collisions. Identify the Agent, Environment, State, Action, Reward, and episode in this RL problem.	2	CO1	BT2
3	“Exploitation is always better than exploration in Reinforcement Learning.” Do you agree with this statement? Justify your answer.	2	CO1	BT2
4	What is a Value Function in Reinforcement Learning?	2	CO1	BT1
5	Define MDP trajectory.	2	CO1	BT1

PART B : Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
6	A smart thermostat is in the state. The thermostat can take two actions a1: Turn ON AC a2: Keep AC OFF Compare Q_π(s, ON), Q_π(s, OFF). Assume the discount factor is 0.9 Action | Trial | R_{t+1} | R_{t+2} Turn ON AC | 1 | +8 | +6 Turn ON AC | 2 | +10 | +4 Keep AC OFF | 1 | +3 | −4 Keep AC OFF | 2 | +2 | −5	10	CO1	BT3
7	A smart irrigation system is in the state “Soil is dry.” The system can take two actions, a_1: Turn ON water pump a_2: Keep water pump OFF The discount factor is 0.9. Calculate and compare Q_π(s, ON) and Q_π(s, OFF). Using the results, determine which action is better. The possible experiences and their probabilities are given below. Action | Trial | Probability | R_{t+1} | R_{t+2} Turn ON Pump | 1 | 60% | +8 | +5 Turn ON Pump | 2 | 40% | +5 | +3 Keep Pump OFF | 1 | 70% | +4 | −2 Keep Pump OFF | 2 | 30% | +2 | −5	10	CO1	BT3

PART C: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
8	A smart home energy management system decides how to operate household appliances to reduce energy consumption. The system receives rewards based on its decisions. The discount factor is 0.9. The current state is that electricity demand is high during peak hours. From this state, two possible scenarios can occur depending on the action taken by the system. Time Step | Scenario A – Energy Saving | Scenario B – High Consumption (t+1) | Shifts washing machine usage: +8 | Continues normal usage: +2 (t+2) | Reduces air-conditioner usage: +6 | High AC usage: −4 (t+3) | Successfully reduces peak consumption: +15 | Electricity consumption remains high: +5 a) Calculate the return for Scenarios A and B b) Calculate the state-value function for the policy that is 70% likely to produce the energy-saving scenario	10	CO1	BT3
9	A delivery robot is moving through a shopping mall. At each step, it receives a reward based on its progress toward delivering the package. The discount factor is: γ=0.9 Time Step | Scenario A – Successful Delivery | Scenario B – Delayed Delivery (t+1) | Moves closer to destination: +10 | Moves closer to destination: +5 (t+2) | Avoids an obstacle: +5 | Hits an obstacle: −5 (t+3) | Delivers package: +20 | Delivers package: +10 a) Calculate the state-value function if the policy gives equal probability to both scenarios b) Calculate the state-value function if the policy makes Scenario A more likely: P(A)=0.8, P(B)=0.2 c) Compare the values obtained.	10	CO1	BT3

PART D: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
10	A food-delivery robot moves through a restaurant to deliver an order. The robot follows a fixed policy π throughout both episodes. The discount factor is $\gamma=0.9$ The robot experiences the following episode. State Sequence | Rewards *S0→S1→S2→S1→S3→T* | *2, 3, -1, 4, 6* Using First-Visit Monte Carlo, calculate the value function after Episode 1.	10	CO2	BT3
11	Assume $\gamma=1$ Episode | State Sequence | Rewards Episode 1 | *S1→S2→S3→T* | *5, 2, 1, 4* Episode 2 | *S1→S2→S6→S1→T* | *2, 2, 6, 1, 4* Calculate the value function after Episode 2, using Every-Visit Monte Carlo	10	CO2	BT3

PART E: Answer Any 1 Following Questions. 10 M

Q.N.	Questions	M	CO	BT
12	Given episode: A→B→C→D→T with the rewards 2,3,4,5 respectively. Initial values: $V(A)=4$, $V(B)=5$, $V(C)=6$, $V(D)=7$ Parameters: $\alpha=0.1$, $\gamma=0.9$ Find the updated value of V(A) using 2-step TD and 3-step TD	10	CO2	BT3
13	An agent repeatedly experiences: A→B→T with rewards 2,4 respectively. Initial value estimates: $V(A)=2$, $V(B)=5$. Parameters: $\alpha=0.2$, $\gamma=0.9$. Using TD (0), calculate $V(A)$ and $V(B)$ after 1 iteration.	10	CO2	BT3