Graphic Era HILL UNIVERSITY
Established by an Act of the State Legislature of Uttarakhand (Adhiniyam Sankhya 12 of 2011)
University under section 2(f) of UGC Act, 1956

End Term (Even) Semester Examination May-June 2025

Roll no.....

Name of the Program and semester: M.Tech CSE Part-Time II semester
Name of the Course: Reinforcement Learning
Course Code: MCS 253

Time: 3 hour

Maximum Marks: 100

Note:

(i) All the questions are compulsory.
(ii) Answer any two sub questions from a, b and c in each main question.
(iii) Total marks for each question is 20 (twenty).
(iv) Each sub-question carries 10 marks.

Q1. (2X10=20 Marks)

a. Explain the working of the Perceptron Learning Algorithm. Derive the update rule and illustrate its convergence with a simple 2D dataset. (CO1)

b. Analyze how activation functions (Sigmoid, Tanh, ReLU) impact learning in MLPs. Compare their gradient behavior and suitability in deep architectures. (CO1)

c. Analyze the exploration-exploitation trade-off. How can techniques like ε-decay or Boltzmann exploration help achieve a balance? (CO1)

Q2. (2X10=20 Marks)

a. Explain how the Deep Q-Network (DQN) differs from standard Q-learning. What role does the target network play in stabilizing learning? (CO2)

b. Simulate Q-Learning for a simple maze with 4 states and 2 actions each. Provide the Q-table updates for two episodes. (CO2)

c. Evaluate the advantages of using Dueling DQN and Prioritized Experience Replay. How do they improve upon the original DQN? (CO2)

Q3. (2X10=20 Marks)

a. Describe the Vanilla Policy Gradient method. Why does it use log-likelihood gradients? Explain with the mathematical formula. (CO3)

b. Implement the REINFORCE algorithm on a basic continuous action environment. Explain how the rewards affect the policy updates. (CO3)

c. Compare PPO, TRPO, and DDPG in terms of stability, convergence speed, and real-world applicability. Which would you choose for robotics, and why? (CO3)

Graphic Era
HILL UNIVERSITY
Established by an Act of the State Legislature of Uttarakhand (Adhiniyam Sankhya 12 of 2011)
University under section 2(f) of UGC Act, 1956

End Term (Even) Semester Examination May-June 2025

Q4. (2X10=20 Marks)

a. What is Multi-Agent Reinforcement Learning (MARL)? Describe its challenges and potential applications. (CO4)

b. Design a cooperative MARL setup where agents must share information to achieve a common goal (e.g., warehouse robots). Describe their communication strategy. (CO4)

c. Analyze how competition and cooperation affect policy learning in MARL. Support your answer with examples or simulations. (CO4)

Q5. (2X10=20 Marks)

a. Describe how PPO improves upon traditional policy gradient methods. What problem does the clipped objective solve? (CO5)

b. Train a PPO agent using the CleanRL library on a continuous control task (e.g., LunarLander-v2). Track the learning curve and interpret results. (CO5)

c. Evaluate the use of Reinforcement Learning from Human Feedback (RLHF) in sensitive applications like education or healthcare. What safeguards are necessary? (CO5)