Register Number:……..…

M.E/M.TECH/MCA DEGREE EXAMINATIONS: NOV / DEC 2024

(Regulation 2018)

Third Semester

DATA SCIENCE

P18ITE0013: Reinforcement Learning

COURSE OUTCOMES

CO1: Explain the reward function and Markov Decision Process

CO2: Solve problems using Dynamic Programming

CO3: Apply temporal difference (TD) learning method for reinforcement learning problem

CO4: Apply Gradient methods for Reinforcement Learning

CO5: Solve case studies like Elevator dispatching, Samuel's checker player, TDgammon using RL algorithms

CO6: Apply Hierarchical Reinforcement Learning Algorithms

Time: Three Hours **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 1 = 10 Marks)

1. **Assertion:** In reinforcement learning, the "explore-exploit dilemma" is essential for learning the optimal policy. CO1 [K_2]

   Reason: Exploring too much will prevent the agent from exploiting the learned policy.

   a) Both A and R are true and R is the correct explanation of A.

   b) Both A and R are true but R is not the correct explanation of A.

   c) A is true but R is false.

   d) A is false but R is true.

2. Which of the following is the learner and decision maker? CO1 [K_2]

   a) Reward

   b) Environment

   c) Agent

   d) State

3. Why value Prediction (or Policy Evaluation) with Function Approximation can be viewed as supervised learning mainly? CO2 [K_2]

   a) A) We can learn the value function by training with batches of data obtained from the agent's

   b) We use stochastic gradient descent to learn the value function.

interaction with the world.

c) Each state and its behaviour forms an input-output training example which we can use to train our approximation to the value function

d) Its labels are pre-defined

4. Assertion: Monte Carlo methods rely on complete episodes of interaction with the environment. CO2 [K_2]
Reason: Monte Carlo methods calculate state values using the expected return after each state-action pair is visited.

a) Both A and R are true and R is the correct explanation of A.

b) Both A and R are true but R is not the correct explanation of A.

c) A is true but R is false.

d) A is false but R is true.

5. Match the following terms with their descriptions: CO3 [K_3]

a) Importance sampling	1. Learning the policy that is bein followed.
b) Policy iteration	2. Evaluating the current policy thr episodes and improving it.
c) Rollout	3. Estimating return by sampling p future trajectories.
d) On-policy learning	4. Sampling technique for correct in off-policy learning.

a) a-1, b-2, c-3, d-4.

b) a-1, b-3, c-2, d-4.

c) a-4, b-3, c-2, d-1.

d) a-4, b-2, c-3, d-1.

6. Which of the following methods belong to off-policy learning? CO3 [K_3]

a) SARSA
b) Q-learning
c) R-learning
d) TD(0)

a) a and b

b) b and c

c) c and d

d) b only

7. **Assertion:** Eligibility traces help in associating rewards to states that were visited some steps before the reward was received. CO4 [K_2]

Reason: Eligibility traces allow the algorithm to handle delayed rewards more effectively by spreading credit backwards through states.

a) Both A and R are true and R is the correct explanation of A.
b) Both A and R are true but R is not the correct explanation of A.
c) A is true but R is false.
d) A is false but R is true.

8. Which of the following is true about gradient estimation in policy gradient methods? CO4 [K_3]
a) It always provides an exact gradient of the loss function.
b) It uses stochastic sampling to approximate the gradient.
c) It relies on a deterministic policy for gradient updates.
d) It updates the value function, not the policy.
a) a and b
b) c only
c) b only
d) b and d

9. Assertion: In hierarchical reinforcement learning, an "option" can represent both primitive actions and higher-level strategies. CO6 [K_3]
Reason: Options are fixed and cannot be learned during training.
a) Both A and R are true and R is the correct explanation of A.
b) Both A and R are true but R is not the correct explanation of A.
c) A is true but R is false.
d) A is false but R is true.

10. What was the significance of Samuel's checker player in the context of reinforcement learning? CO5 [K_2]
a) It used hierarchical reinforcement learning to solve board games
b) It was the first program to use temporal difference learning.
c) It implemented the HAM framework.
d) It was designed for robotic control tasks.

PART B (10 x 2 = 20 Marks)

11. Give an example of a reward in the context of reinforcement learning. CO1 [K_3]
12. When considering an innovation, explain the role of explore-exploit tradeoff. CO1 [K_2]
13. Differentiate horizontal scalability with vertical scalability. CO2 [K_2]
14. Bring out the importance of rollouts in Monte Carlo through an example. CO2 [K_2]
15. Justify, which of the two replacing traces and accumulating traces, gives significant improvement in learning rate. CO3 [K_3]
16. Compare and contrast between on-policy (SARSA) and off-policy (Q-learning) methods. CO3 [K_2]

17.	Distinguish between the exact and stochastic policy gradient methods.	CO4	[K_2]
18.	Which of the control algorithms in Reinforcement Learning is most stable? Give your reason.	CO5	[K_3]
19.	In the Acrobot case study, how does hierarchical reinforcement learning address the main challenges?	CO5	[K_3]
20.	Differentiate between primitive actions and hierarchical options in Reinforcement Learning.	CO6	[K_2]

PART C (6 x 5 = 30 Marks)

21.	Explain the Markov Decision Process (MDP) concept and its key components.	CO1	[K_2]
22.	Reason out in detail, how asynchronous Dynamic Programming is advantageous over synchronous Dynamic Programming.	CO2	[K_3]
23.	Discuss Q-learning and how it finds the optimal policy.	CO2	[K_2]
24.	Justify how actor-critic methods combine value-based and policy-based approaches for better control in reinforcement learning.	CO4	[K_3]
25.	How do options framework enable temporal abstraction, and what are the key components of an option? Provide examples.	CO6	[K_3]
26.	Explain the MAXQ framework in detail. How does it help in decomposing complex tasks into smaller sub-tasks, and how are value functions used within this framework?	CO6	[K_2]

Answer any FOUR Questions
PART D (4 x 10 = 40 Marks)

27.	Compare and contrast different exploration schemes in reinforcement learning and their impact on the explore-exploit dilemma.	CO1	[K_3]
28.	Discuss how games and afterstates are handled in reinforcement learning, and provide an example of how afterstates can simplify value function learning in board games like chess or backgammon.	CO3	[K_3]
29.	Elaborate on the function approximation methods in reinforcement learning, including gradient descent, linear function approximation, and fitted iterative methods. Provide examples of how they are used for both value prediction and control.	CO4	[K_3]
30.	Analyze the hierarchical frameworks applicable for Elevator Dispatching, Samuel's Checker Player, and TD-Gammon.	CO5	[K_4]
31.	How is hierarchical RL applied to helicopter piloting, and what role do sub-tasks play in managing the complexity of flight control?	CO6	[K_3]

****