Total No. of Questions: 6 *Total No. of Printed Pages:3*

Enrollment No.....

Faculty of Engineering
End Sem Examination Dec-2023
RA3EL05 Reinforcement Learning

Programme: B.Tech. Branch/Specialisation: RA

Duration: 3 Hrs. **Maximum Marks: 60**

Note: All questions are compulsory. Internal choices, if any, are indicated. Answers of Q.1 (MCQs) should be written in full instead of only a, b, c or d. Assume suitable data if necessary. Notations and symbols have their usual meaning.

Q.1 i. Which technique in reinforcement learning involves solving problems by breaking them into smaller subproblems and solving each subproblem iteratively? **1**

(a) Deep learning (b) Q-learning
(c) Dynamic programming (d) Markov decision processes

ii. Which of the following best describes the concept of a Markov Decision Process (MDP)? **1**

(a) A sequence of non-Markov states
(b) A deterministic environment
(c) A formal framework for modeling sequential decision-making
(d) An algorithm for solving reinforcement learning problems

iii. What does the term "temporal difference" refer to in reinforcement learning? **1**

(a) The difference between the current state and the previous state
(b) The time it takes for an agent to make a decision
(c) The difference between predicted and actual rewards at each time step
(d) The time it takes for an agent to explore the entire state space

iv. Which reinforcement learning algorithm is specifically designed for on-policy learning and is used for estimating action values in a model-free setting? **1**

(a) Monte Carlo methods (b) Temporal difference learning
(c) Q-learning (d) SARSA

v. Which reinforcement learning algorithm is known for using deep neural networks to approximate the action-value function and has been successful in solving complex tasks, such as playing video games? **1**
(a) Policy gradient methods
(b) Proximal Policy Optimization (PPO)
(c) Trust Region Policy Optimization (TRPO)
(d) Deep Q-networks (DQN)

vi. Which reinforcement learning algorithm is designed to ensure that policy updates are limited to a "trust region," preventing large policy changes that could result in catastrophic failures during training? **1**
(a) Q-learning (b) PPO (c) TRPO (d) DQN

vii. What is a common challenge in multi-agent learning that does not typically occur in single-agent settings? **1**
(a) Lack of diversity in actions
(b) Coordination and competition conflicts
(c) Exploration and exploitation balance
(d) Transfer of learned policies

viii. What type of data does imitation learning primarily rely on? **1**
(a) Randomly generated data
(b) Reward signals
(c) Demonstrator's actions and states
(d) Test data from the target task

ix. In which application area does reinforcement learning find extensive use in optimizing processes, such as resource allocation and traffic management? **1**
(a) Game playing (b) Control systems
(c) Robotics (d) Healthcare

x. AlphaGo, a milestone in AI, demonstrated the power of reinforcement learning in which domain? **1**
(a) Game playing (b) Healthcare
(c) Finance (d) Defence Sector

Q.2 i. What do you mean by value functions? **2**

ii. Explain the concept of a Markov Decision Process (MDP) and how it is used in reinforcement learning. Provide an example to illustrate your explanation. **8**

OR iii. Discuss the role of dynamic programming in reinforcement learning. Provide a step-by-step explanation of the dynamic programming process and how it can be applied to solve reinforcement learning problems. **8**

Q.3 i. Define Eligibility traces. **2**

ii. Write a Python code that demonstrates how to implement the SARSA algorithm using the OpenAI's gym module to load the environment. **8**

OR iii. Describe the concept of value function approximation in reinforcement learning and explain when it is useful. Provide an example of a situation where value function approximation is applied. **8**

Q.4 i. What are disadvantages of Policy gradient methods? **2**

ii. Write short notes on **8**
(a) PPO (b) TRPO

OR iii. Explain the concept of policy gradient methods in reinforcement learning and discuss the trade-offs associated with using them. Provide an example to illustrate your explanation. **8**

Q.5 i. What do you mean by Thompson Sampling? **2**

ii. Discuss the concept of hierarchical reinforcement learning and its applications. **8**

OR iii. Discuss the concept of transfer learning in reinforcement learning and its applications. **8**

Q.6 Write short notes on any two:

i. Reinforcement learning for control systems. **5**

ii. Reinforcement learning in healthcare. **5**

iii. Reinforcement learning in Gaming. **5**

******

Marking Scheme
Reinforcement Learning (T) - RA3EL05 (T)

Q.1	i)	c) Dynamic programming		**1**
	ii)	c) A formal framework for modelling sequential decision-making		**1**
	iii)	c) The difference between predicted and actual rewards at each time step		**1**
	iv)	d) SARSA		**1**
	v)	d) Deep Q-networks (DQN)		**1**
	vi)	c) TRPO		**1**
	vii)	b) Coordination and competition conflicts		**1**
	viii)	c) Demonstrator's actions and states		**1**
	ix)	b) Control systems		**1**
	x)	a) Game playing		**1**
Q.2	i.	Numerical …….. Action	(As per explanation)	**2**
	ii.	Explanation	4 Marks	**4**
		Uses	2 Marks	**2**
		Example	2 Marks	**2**
OR	iii.	Explanation	4 Marks	**4**
		Uses	2 Marks	**2**
		Example	2 Marks	**2**
Q.3	i.	Definition	(As per explanation)	**2**
	ii.	import numpy as np…… policy.	(As per explanation)	**8**
OR	iii.	Description	6 Marks	**6**
		Example	2 Marks	**2**
Q.4	i.	Frequently, ……. problem.	2 Marks	**2**
	ii.	Short note on PPO	4 Marks	**4**
		Short note on TRPO	4 Marks	**4**
OR	iii.	Explanations	3 Marks	**3**
		Trade-offs	3 Marks	**3**
		Examples	2 Marks	**2**
Q.5	i.	Probabilistic …… rewards.	(As per explanation)	**2**
	ii.	Discussion	5 Marks	**5**
		Applications	3 Marks	**3**
OR	iii.	Discussion	5 Marks	**5**
		Applications	3 Mark	**3**
Q.6		Any Two		
	i.	Short note	(As per explanation)	**5**
	ii.	Short note	(As per explanation)	**5**
	iii.	Short note	(As per explanation)	**5**

******