Reg. No.															

Question Paper Code	13840

B.E. / B.Tech. - DEGREE EXAMINATIONS, NOV / DEC 2025

Seventh Semester

Artificial Intelligence and Data Science

20AIEL706 - REINFORCEMENT LEARNING

Regulations - 2020

Duration: 3 Hours | Max. Marks: 100

PART - A (MCQ) (10 × 1 = 10 Marks)

Answer ALL Questions

		Marks	K – Level	CO
1.	Which of the following best defines Reinforcement Learning? (a) Supervised learning (b) Unsupervised learning (c) Learning through interaction and reward (d) Feature-based learning	1	K1	CO1
2.	In RL, the component that decides what action to take in a given state is called (a) Reward (b) Environment (c) Policy (d) Agent	1	K1	CO1
3.	The property that the next state depends only on the current state is known as (a) Temporal locality (b) Markov property (c) Bellman property (d) Greedy property	1	K1	CO2
4.	The Bellman equation defines a relationship between (a) State and reward (b) Value of current and next states (c) Agent and policy (d) Discount and state	1	K1	CO2
5.	Which of the following is an example of Dynamic Programming method? (a) SARSA (b) Monte Carlo (c) Value Iteration (d) Q-Learning	1	K1	CO3
6.	Policy Iteration alternates between (a) Value updates and policy evaluation (b) Action updates and learning (c) Random search and optimization (d) Policy evaluation and policy improvement	1	K1	CO3
7.	Which method estimates returns by sampling complete episodes? (a) TD(0) (b) SARSA (c) Monte Carlo (d) Q-Learning	1	K1	CO4
8.	In TD learning, the parameter α refers to (a) Learning rate (b) Discount factor (c) Reward rate (d) Exploration factor	1	K1	CO4
9.	In function approximation, which of the following methods is used to reduce variance in policy gradient estimates? (a) Importance sampling (b) Baseline function (c) Eligibility trace (d) Actor-only method	1	K1	CO5
10.	The REINFORCE algorithm belongs to which category? (a) Value-based (b) Policy-based (c) Model-based (d) Monte Carlo-based	1	K1	CO6

PART - B (12 × 2 = 24 Marks)

Answer ALL Questions

		Marks	K – Level	CO
11.	Define Reinforcement Learning and mention its main elements.	2	K1	CO1
12.	Differentiate Reinforcement Learning from Supervised Learning.	2	K2	CO1
13.	Define Markov Chain and give an example.	2	K1	CO2
14.	Explain the Bellman Expectation Equation in brief.	2	K2	CO2
15.	What is Policy Iteration? List its steps.	2	K1	CO3
16.	Write short notes on Value Iteration.	2	K2	CO3
17.	Differentiate On-policy and Off-policy learning methods.	2	K2	CO4
18.	Explain about TD(λ) method.	2	K2	CO4
19.	Define Function Approximation. Why is it used in RL?	2	K2	CO5
20.	Infer on Experience Replay in Deep Q-Networks.	2	K2	CO5
21.	Explain the concept of Policy Gradient with an example.	2	K2	CO6

22.		What is Actor-Critic architecture?	*2*	*K1*	*CO6*

PART - C (6 × 11 = 66 Marks)

Answer ALL Questions

23.	a)	Describe the origin and evolution of Reinforcement Learning and its connections with Machine Learning and Psychology.	*11*	*K2*	*CO1*
		OR			
	b)	Explain the concepts of random variables, PDFs, and CDFs with examples.	*11*	*K2*	*CO1*
24.	a)	Define Markov Decision Process (MDP) and its components with suitable diagrams.	*11*	*K3*	*CO2*
		OR			
	b)	Derive the Bellman Optimality Equations for MDP.	*11*	*K3*	*CO2*
25	a)	Construct the policy iteration and value iteration algorithms with examples.	*11*	*K3*	*CO3*
		OR			
	b)	Identify the contraction mapping property of Bellman operators using Banach Fixed Point theorem.	*11*	*K3*	*CO3*
26.	a)	Describe the Monte Carlo and Temporal-Difference learning approaches for model-free prediction and control. Compare On-policy and Off-policy methods using SARSA and Q-Learning examples, and explain how these methods unify the principles of DP and MC.	*11*	*K2*	*CO4*
		OR			
	b)	Explain the algorithm for Q-learning and explain.	*11*	*K2*	*CO4*
27.	a)	Identify the importance of function approximation in Reinforcement Learning. Explain Gradient Monte Carlo, Semi-gradient TD(0), and the use of eligibility traces. Analyze how Least Squares and Experience Replay help stabilize learning in Deep Q-Networks.	*11*	*K3*	*CO5*
		OR			
	b)	Build the role of Least Squares and Experience Replay in deep RL.	*11*	*K3*	*CO5*
28.	a)	Analyze the working of Policy Gradient methods using Log-derivative trick.	*11*	*K4*	*CO6*
		OR			
	b)	Describe the REINFORCE algorithm with pseudo-code and discuss how variance is reduced using baselines.	*11*	*K4*	*CO6*