Uka Tarsadia University

B. Tech.

CSE / CSE (AI&ML) / CSE (CC) / CSE (CS) / CE / CE (SE) / IT

Semester VI

Program Elective -III

REINFORCEMENT LEARNING

AI5020

EFFECTIVE FROM December-2024

Syllabus version: 1.00

Subject Code	Subject Title
AI5020	Reinforcement Learning

Teaching Scheme				Examination Scheme			
Hours		Credits		Theory Marks		Practical Marks	Total Marks
Theory	Practical	Theory	Practical	Internal	External	CIE	
3	2	3	1	40	60	50	150

Objectives of the course:

- Learn how to define RL tasks and the core principals behind the RL, including policies, value functions, deriving Bellman equations
- Understand and work with tabular methods to solve classical control problems
- Recognize current advanced techniques and applications in RL

Course outcomes:

Upon completion of the course, the student shall be able to

CO1: Describe the salient characteristics of reinforcement learning that set it apart from AI and passive machine learning.
CO2: Model a control task in the framework of MDPs.
CO3: Identify the model based from the model free methods.
CO4: Identify stability/convergence and approximation properties of RL algorithms.
CO5: To integrate planning with learning.
CO5: Describe the exploration vs exploitation challenge.

Sr. No.	Topics	Hours
	Unit – I	
1	**Introduction to RL and Markov Decision Process (MDP):** The RL Problem, Markov Process, Markov Reward Process, Markov Decision Process and Bellman Equations, Partially Observable MDPs.	4
	Unit – II	
2	**Planning by Dynamic Programming (DP):** Policy Evaluation, Value Iteration, Policy Iteration, DP Extensions and Convergence using Contraction Mapping.	8
	Unit – III	
3	**Model-free Prediction:** Monte-Carlo (MC) Learning, Temporal-Difference (TD) Learning, TD-Lambda and Eligibility Traces. **Model-free Control:** On-Policy MC Control, On-Policy TD Learning and Off-Policy Learning	10

Unit – IV		
4	**Value Function Approximation:** Incremental Methods and Batch Methods, Deep Q-Learning, Deep Q-Networks and Experience Replay **Policy Gradient Methods:** Finite-Difference, Monte-Carlo and Actor-Critic Methods	10
Unit – V		
5	**Integrating Planning with Learning:** Model-based RL, Integrated Architecture and Simulation-based Search	8
Unit – VI		
6	**Exploration and Exploitation (Bandits):** Multi-arm Bandits, Contextual Bandits and MDP Extensions **Integrating AI Search and Learning:** Classical Games: Combining Minimax Search and RL	5

Sr. No.	**Reinforcement Learning (Practicals)**	**Hours**
1	To implement Multi-Armed Bandits (Epsilon-Greedy and UCB) by introducing the simplest RL setup where an agent chooses from multiple slot machines to maximize reward.	4
2	To implement Tabular Q-Learning in Gridworld. Implement Q-Learning in a discrete grid environment where the agent must reach a goal state.	4
3	To implement SARSA in a Maze Navigation Task. Compare SARSA with Q-Learning in a grid-based maze.	4
4	To study and implement Deep Q-Network (DQN) with CartPole.	4
5	Extend the DQN approach implemented in practical 4 to overcome overestimation of Q-values and speed up learning.	4
6	To study and implement Policy Gradients (REINFORCE) on a Simple Continuous Task.	4
7	To study Actor-Critic (A2C) in LunarLander. Combine policy- and value-based methods to train an agent with Actor-Critic methods.	6

Text book:

2. Richard S. Sutton and Andrew G. Barto – "Reinforcement Learning: An Introduction", 2nd Edition, MIT Press, 2020.

Reference books:

2. Csaba Szepesvári – "Algorithms of Reinforcement Learning; Synthesis Lectures on Artificial Intelligence and Machine Learning", Morgan & Claypool Publishers, 2010.
3. Dimitri P. Bertsekas – "Reinforcement Learning and Optimal Control", 1st Edition, Athena Scientific, 2019.

4. Dimitri P. Bertsekas – "Dynamic Programming and Optimal Control", 4th Edition, Athena Scientific, 2017.

Course objectives and Course outcomes mapping:

- Learn how to define RL tasks and the core principals behind the RL, including policies, value functions, deriving Bellman equations: CO1, CO2.
- Understand and work with tabular methods to solve classical control problems: CO3, CO4.
- Recognize current advanced techniques and applications in RL: CO5, CO6.

Course units and Course outcomes mapping:

Unit No.	Unit Name	Course Outcomes					
		CO1	**CO2**	**CO3**	**CO4**	**CO5**	**CO6**
1	Introduction to RL and Markov Decision Process (MDP)	✓					
2	Planning by Dynamic Programming (DP)		✓				
3	Model-free Prediction and Model-free Control			✓			
4	Value Function Approximation and Policy Gradient Methods				✓		
5	Integrating Planning with Learning					✓	
6	Exploration and Exploitation (Bandits) and Integrating AI Search and Learning						✓

Programme outcomes:

PO 1: Engineering knowledge: An ability to apply knowledge of mathematics, science, and engineering.

PO 2: Problem analysis: An ability to identify, formulates, and solves engineering problems.

PO 3: Design/development of solutions: An ability to design a system, component, or process to meet desired needs within realistic constraints.

PO 4: Conduct investigations of complex problems: An ability to use the techniques, skills, and modern engineering tools necessary for solving engineering problems.

PO 5: Modern tool usage: The broad education and understanding of new engineering techniques necessary to solve engineering problems.

PO 6: The engineer and society: Achieve professional success with an understanding and appreciation of ethical behaviour, social responsibility, and diversity, both as individuals and in team environments.

PO 7: Environment and sustainability: Articulate a comprehensive world view that integrates diverse approaches to sustainability.

PO 8: Ethics: Identify and demonstrate knowledge of ethical values in non-classroom activities, such as service learning, internships, and field work.

PO 9: Individual and team work: An ability to function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO 10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give/receive clear instructions.

PO 11: Project management and finance: An ability to demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO 12: Life-long learning: A recognition of the need for, and an ability to engage in life-long learning.

Programme outcomes and Course outcomes mapping:

Programme Outcomes	Course Outcomes					
	CO1	CO2	CO3	CO4	CO5	CO6
PO1						
PO2		✓	✓		✓	✓
PO3	✓	✓			✓	✓
PO4	✓	✓		✓		✓
PO5						✓
PO6						
PO7						
PO8						
PO9						
PO10						
PO11						
PO12						