NIRMA UNIVERSITY

Institute:	Institute of Technology
Name of Programme:	Integrated BTech (CSE)-MBA
Course Code:	3CS108ME24
Course Title:	Reinforcement Learning
Course Type:	Department Elective-V
Year of Introduction:	2024-25

L	T	Practical Component				C
		LPW	PW	W	S	
3	0	2	-	-	-	4

Course Learning Outcomes (CLO):

At the end of the course, the students will be able to –

1. summarise the fundamental concepts and principles of reinforcement learning (BL2)
2. make use of tabular methods to solve classical control problems (BL3)
3. choose suitable approximation solutions for reinforcement learning (BL3)
4. recommend suitable techniques and applications of reinforcement learning. (BL5)

Unit	Contents	Teaching Hours (Total 45)
Unit-I	**Foundations:** Introduction and Basics of RL, Defining RL Framework, Markov decision process (MDP), state and action value functions, Bellman equations, optimality of value functions and policies, Bellman optimality equations.	07
Unit-II	**Prediction and Control by Dynamic Programming:** Overview of dynamic programming for MDP, definition and formulation of planning in MDPs, principle of optimality, iterative policy evaluation, policy iteration, value iteration.	07
Unit-III	**Monte Carlo Methods for Model Free Prediction and Control:** Overview of Monte Carlo methods for model-free RL, Monte Carlo control, on-policy and off-policy learning, Importance sampling, Incremental Monte Carlo Methods for Model Free Prediction.	07
Unit-IV	**TD Methods:** Overview TD (0), TD (1), and TD(λ), k-step estimators, unified view of DP, MC, and TD evaluation methods, TD Control methods - SARSA, Q-Learning, and their variants.	07
Unit-V	**Function Approximation Methods**: Overview of function approximation methods, gradient descent from Machine Learning, Gradient MC and Semi-gradient TD (0) algorithms, Eligibility trace for function approximation, Control with function approximation, least squares, Experience replays in deep Q-Networks.	10
Unit-VI	**Recent Advances and Applications:** Meta-learning, Multi-Agent Reinforcement Learning, Partially Observable Markov Decision Process, Applying RL for real-world problems	07

Self-Study:

The self-study contents will be declared at the commencement of the semester. Around 10% of the questions will be asked from self-study contents

Suggested Readings/ References:

1. Richard S. Sutton and Andrew G. Barto, Reinforcement learning: An introduction, MIT Press
2. Wiering, Marco, and Martijn Van Otterlo, Reinforcement Learning-Adaptation, learning, and optimization, Springer
3. Dimitri P. Bertsekas, Reinforcement Learning and Optimal Control, Athena Scientific.
4. Warren B. Powell, Reinforcement Learning and Stochastic Optimization: A Unified Framework for Sequential Decisions, Wiley
5. Csaba Szepesvári, Algorithms for Reinforcement Learning, Springer

Suggested List of Experiments:

Sr. No.	Title	Hours
1	Write a program to develop an agent that takes random actions in a grid world environment.	04
2	Write a program that constructs an agent with a Q-learning algorithm.	02
3	Create a program that trains an agent using SARSA and Q-learning.	02
4	Write a program to create a multi-armed bandit problem with multiple arms or actions, using different exploration strategies such as epsilon-greedy and UCB.	04
5	Write a program to design a Markov Decision Process (MDP) and employ the value iteration algorithm to calculate optimal values.	02
6	Write a program to design a Markov Decision Process (MDP) and employ the policy iteration algorithm to calculate optimal policy.	02
7	Write a program to simulate the CartPole environment in OpenAI Gym and implement a Deep Q Network.	04
8	Write a program to design an environment with a continuous action space and implement an actor-critic architecture with a neural network.	02
9	Develop a DQN-based reinforcement learning model to tackle a real-world application.	04
10	Develop an A2C-based reinforcement learning model to tackle a real-world application	04