

राष्ट्रीय प्रौद्योगिकी संस्थान पटना / NATIONAL INSTITUE OF TECHNOLOGY PATNA

संगणक विज्ञान एंव अभियांत्रिकी विभाग / DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING अशोक राजपथ, पटना-८०००५, बिहार / ASHOK RAJPATH, PATNA-800005, BIHAR

L-T-P-Cr: 2-0-2-3

Phone No.: 0612-2372715, 2370419, 2370843, 2371929 Ext- 200, 202 Fax-0612-2670631 Website: www.nitp.ac.in

No:- Date:

CSX4179 Reinforcement Learning

Pre-requisites: Fundamental knowledge Machine Learning

Objectives/Overview:

- 1. To introduce the basics of reinforcement learning
- 2. To introduce the reinforcement learning algorithms
- 3. To introduce dynamic programming and its usage in RL
- 4. To introduce state of the art applications in RL

Course Outcomes:

On completion of this course, the student will be able to:

- CO 1: Knowledge of basic and advanced reinforcement learning techniques.
- CO 2: Identification of suitable learning tasks to which these learning techniques can be applied.
- CO 3: Appreciation of some of the current limitations of reinforcement learning techniques.
- **CO 4:** Formulation of decision problems, set up and run computational experiments, evaluation of results from experiments.

Course Outcomes-Cognitive Levels-Program Outcomes Matrix – [H: High relation (3); M: Moderate relation (2); L: Low relation (1)]

	Program Outcomes											
Dutcome	Engineerir g	m analysis	elopment	PO-4 (Conduct nvestigation ns of complex problems)	(Moder n tool	engineer	and	PO-8 (Ethics)	(Individua	inicatio	manageme nt	PO-12 Life-lor g earning)
CO-1	3	3	3	3	2	3			3	3	1	3
CO-2	3	3	3	3	2	3	·	1	3	3	1	3
CO-3	3	3	3	3	3	3	1	2	3	3	1	3
CO-4	3	3	3	3	2	3	1	1	3	3	1	3

UNIT I: Introduction to Reinforcement Learning Problem:

Reinforcement Learning, Elements of Reinforcement Learning, Limitations and Scope, An Extended Example: Tic-Tac-Toe, History of Reinforcement Learning.

UNIT II: Multi-arm Bandits:

Lectures: 03

Lectures: 02

An n-Armed Bandit Problem, Action-Value Methods, Incremental Implementation, Tracking a Nonstationary Problem, Optimistic Initial Values, Upper-Confidence-Bound

UNIT III: Finite Markov Decision Processes:

Lectures: 05

The Agent–Environment Interface, Goals and Rewards, Returns, Unified Notation for Episodic and Continuing Tasks, The Markov Property, Markov Decision Processes, Value Functions, Optimal Value Functions, Optimality and Approximation.

UNIT IV: Dynamic Programming:

Lectures: 04

Policy Evaluation, Policy Improvement, Policy Iteration, Value Iteration, Asynchronous Dynamic Programming, Generalized Policy Iteration, Efficiency of Dynamic Programming.

UNIT V: Monte Carlo Methods:

Lectures: 05

Monte Carlo Prediction, Monte Carlo Estimation of Action Values, Monte Carlo Control, Monte Carlo Control without Exploring Starts, Off-policy Prediction via Importance Sampling, Incremental Implementation, Off-Policy Monte Carlo Control, Importance Sampling on Truncated Returns.

UNIT VI: Temporal-Difference Learning:

Lectures: 05

TD Prediction, Advantages of TD Prediction Methods, Optimality of TD(0), Sarsa: On-Policy TD Control, Q-Learning: Off-Policy TD Control, Games.

UNIT VII: Policy Approximation:

Lectures: 04

Actor–Critic Methods, Eligibility Traces for Actor–Critic Methods, R-Learning and the Average-Reward Setting, Vanilla policy gradient method,

Text Books:

- 1. RS Sutton Reinforcement Learning: An Introduction Stanford University.
- 2. Enes Bilgin, Mastering Reinforcement Learning with Python, Packt.

Reference Books:

- 3. Sudharsan Ravichandiran, Hands-On Reinforcement Learning with Python: Master Reinforcement and Deep Reinforcement Learning Using OpenAI Gym and TensorFlow.
- 4. Aske Plaat, Deep Reinforcement Learning, Springer Singapore