Foundations of Reinforcement Learning

Course Code: 20ITH103	L	T	P	C
	3	1	0	4

Pre-requisites: Artificial Intelligence

Course Outcomes: At the end of the course, a student will be able to:

CO1: Demonstrate various Components of Reinforcement Learning. (L2)

CO2: Make use of various exploration and exploitation strategies. (L3)

CO3: Apply Model based and Model Free Prediction techniques. (L3)

CO4: Make use of different value based Reinforcement Learning Algorithms. (L3)

CO5: Demonstrate various Policy based Reinforcement Learning Algorithms. (L3)

UNIT-I: (10 Lectures)

Introduction: Deep Reinforcement Learning, Suitability of RL, Components of Reinforcement Learning -Agent, Environment, Observations, Actions, Example-The Bandit Walk Environment, Agent-Environment interaction cycle, MDP (Markov Decision Process): The engine of the Environment-States, Actions, Transition Function, Reward Signal. (Chapter-1&2)

Learning Outcomes: At the end of the unit, student will be able to

1. List various applications of Reinforcement Learning. (L1)
2. Explain the components of Reinforcement Learning. (L2)
3. Describe the Markov Decision Process. (L2)

UNIT-II: (10 Lectures)

Planning: Objective of a decision making agent-environment, Plan, Optimal policy, Comparison of Policies, Bellman Equation/State-Value Function, Action-Value Function, Action-Advantage Function and Optimality. (Chapter-3)

Exploitation and Exploration of Reinforcement Learning: Bandits- Single-state decision problem(Multi-Armed Bandit(MAB) problem), The cost of exploration, Approaches to solve MAB environments, Greedy Strategy, Random Strategy, Epsilon-Greedy Strategy, Decaying Epsilon-Greedy Strategy, Optimistic Initialization strategy, Strategic exploration, Softmax exploration strategy, Upper confidence bound (UCB) equation strategy, Thompson sampling strategy.(Chapter-4)

Learning Outcomes: At the end of the unit, student will be able to

1. Illustrate best policies of behavior in sequential decision-making problems modeled with MDPs. (L2)
2. List various approaches to solve the MAB environment. (L1)
3. Apply Random and Optimistic Exploration Strategies to make correct decision making. (L3)

UNIT-III: (10 Lectures)

Model Free Reinforcement Learning: Monte Carlo Prediction (MC), First-Visit MC (FVMC), Every-Visit MC (EVMC), Temporal Difference Learning (TD), Learning to estimate from multiple steps, N-step TD learning, Forward-view TD(λ), Backward-view TD(λ), Generalized policy iteration(GPI), Monte Carlo control, SARSA: On-Policy TD control, Q-learning: Off-Policy TD control, Watkins's Q(λ).

Model Based Reinforcement Learning: Dyna-Q, Trajectory sampling. (Chapter-5,6,7)

Learning Outcomes: At the end of the unit, student will be able to

1. Explain different versions of Monte Carlo Prediction. (L2)
2. Differentiate between Model Free and Model based Reinforcement Learning. (L2)
3. Apply different Prediction techniques. (L3)

UNIT-IV: (12 Lectures)

Value Based Reinforcement Learning: Deep reinforcement learning agents with sequential feedback, evaluative feedback, sampled feedback, Function Approximation for Reinforcement Learning- high-dimensional state and action spaces, continuous state and action spaces, state-value function and action-value function with and without function approximation, Neural Fitted Q (NFQ), Deep Q-Network (DQN). (Chapter-8,9,10)

Learning Outcomes: At the end of the unit, student will be able to

1. Demonstrate deep reinforcement learning agents with feedback. (L2)
2. Illustrate function approximation in Reinforcement Learning. (L2)
3. Apply different value based Reinforcement Learning Algorithms. (L3)

UNIT-V: (10 Lectures)

Policy Based Reinforcement Learning: Policy Gradient and Actor-Critic Methods—REINFORCE Algorithm and Stochastic Policy Search, Vanilla Policy Gradient(VPG), Asynchronous Advantage Actor-Critic (A3C), Generalized Advantage Estimation (GAE), Advantage Actor-Critic(A2C), Deep Deterministic Policy Gradient (DDPG), Twin-Delayed DDPG (TD3), Soft Actor-Critic (SAC). (Chapter-11,12)

Learning Outcomes: At the end of the unit, student will be able to

1. Explain various policy Gradient methods. (L2)
2. Describe Acot-Critic methods. (L2)
3. Demonstrate various policy based Reinforcement Learning Algorithms. (L3)

TEXT BOOKS:

1. Miguel Morales, "*Grokking Deep Reinforcement Learning*", Manning Publications, 2020.

REFERENCE BOOKS:

1. Richard S. Sutton and Andrew G. Barto, "*Reinforcement learning: An Introduction*, Second Edition", MIT Press, 2019.
2. Marco Wiering, Martijn van Otterlo(Ed), "*Reinforcement Learning, State-of-the-Art, Adaptation*", Learning, and Optimization book series, ALO, volume 12, Springer, 2012.
3. Keng, Wah Loon, Graesser, Laura, *"Foundations of Deep Reinforcement Learning: Theory and Practice in Python"*, Addison Wesley Data & Analytics Series, 2020.
4. Francois Chollet, "*Deep Learning with Python*", Manning Publications, 2018.

WEB REFERENCES:

1. http://cse.iitkgp.ac.in/~adas/courses/rl_aut2021/syllabus.html