# 20AD41E8 - REINFORCEMENT LEARNING

|  |  |  |  |
| --- | --- | --- | --- |
| Course Category: | Professional Elective | Credits: | 3 |
| Course Type: | Theory | Lecture-Tutorial-Practical: | 3-0-0 |
| Prerequisite: | Linear algebra and calculus, Machine learning and Knowledge in programming languages. | Sessional Evaluation:Univ. Exam Evaluation:Total Marks: | 4060100 |
| Objectives: | * To pick the best-known action for any given state, which means the actions have to be ranked, and assigned values relative to one another.
* To gain knowledge of basic and advanced reinforcement learning techniques.
* To understand and work with approximate solutions (deep Q network-based algorithms)
 |

|  |  |
| --- | --- |
| Course Outcomes | Upon successful completion of the course, the students will be able to: |
| CO1 | Illustrates various elements of reinforcement techniques. |
| CO2 | Describes the finite markov decision processes and its main ideas. |
| CO3 | Define the key features of reinforcement learning and distinguishes it from AI and non-interactive machine learning. |
| CO4 | Analyze any given application; decide if it is formulated as reinforcement learning problem. |
| CO5 | Apply Monte Carlo method for prediction. |
| CO6 | Adapt Temporal-Difference (TD) learning for prediction. |
| Course Content | UNIT-I**The Reinforcement Learning Problem:** Reinforcement Learning, Examples, Elements of Reinforcement Learning, Limitations and Scope, An Extended Example: Tic-Tac-Toe, History of Reinforcement Learning.UNIT-II**Multi-arm Bandits:** An n-Armed Bandit Problem, Action-Value Methods, Incremental Implementation,Tracking a Nonstationary Problem, Optimistic Initial Values, Upper-Confidence-Bound Action Selection, Gradient Bandits, Associative Search. UNIT-III**Finite Markov Decision Processes:** 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:** Policy Evaluation, Policy Improvement, Policy Iteration, Value Iteration, Asynchronous Dynamic Programming, Generalized Policy Iteration, Efficiency of Dynamic Programming.UNIT-V**Monte Carlo Methods:** 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 (TD) Learning:**  TD Prediction, Advantages of TD Prediction Methods, Optimality of TD(0), Sarsa: On-Policy TD Control, Q-Learning: Off-Policy TD Control, Games, Afterstates, and Other Special Cases. |
| Text Books &ReferenceBooks | **TEXT BOOKS:**1. Richard S. Sutton and Andrew G. Barto, "Reinforcement Learning: An Introduction", 2nd Edition.

**REFERENCE BOOKS:**1. Kyriakos G. Vamvoudakis, Yan Wan, Frank L. Lewis, Derya Cansever,"Handbook of Reinforcement Learning and Control (Studies in Systems, Decision and Control, 325)", 1st Edition.
2. Nimish Sanghi,"Deep Reinforcement Learning with Python: With PyTorch, TensorFlow and OpenAI Gym", 1st Edition.
3. Boris Belousov, Hany Abdulsamad, Pascal Klink, Simone Parisi, Jan Peters."Reinforcement Learning Algorithms: Analysis and Applications", 1st Edition.
 |
| E-Resources | 1. <https://nptel.ac.in/courses/106106143>
2. <https://www.coursera.org/specializations/reinforcement-learning>
 |

**CO-PO Mapping:** 3-High Mapping, 2-Moderate Mapping, 1-Low Mapping

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   | **PO1** | **PO2** | **PO3** | **PO4** | **PO5** | **PO6** | **PO7** | **PO8** | **PO9** | **PO10** | **PO11** | **PO12** |
| **CO1** |  | 2 | 2 |  |  |  |  |  |  |  |  |  |
| **CO2** | 2 | 1 | 2 |  |  | 1 | 1 |  |  |  |  |  |
| **CO3** | 1 | 1 | 1 | 2 | 2 | 1 | 1 |  |  |  |  |  |
| **CO4** | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| **CO5** | 2 |  2 | 2  | 2 | 1 | 1 | 1 | 1 | 1  | 1 | 1 | 1 |
| **CO6** | 2  | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 |