SCHEME OF COURSE WORK

Department of Information Technology

Course Details:

COURSE TITLE	Foundations of Reinforcement Learning		
COURSE CODE	20ITH103	L T P C	3 1 0 4
PROGRAM	B.TECH		
SPECIALIZATION	CSE, IT		
SEMESTER	VI		
PRE REQUISITES	Artificial Intelligence		
COURSES TO WHICH IT IS A PREREQUISITE	N/A		

Course Outcomes (COs):

1	Demonstrate various Components of Reinforcement Learning. (L2)
2	Make use of various exploration and exploitation strategies. (L3)
3	Apply Model based and Model Free Prediction techniques. (L3)
4	Make use of different value based Reinforcement Learning Algorithms. (L3)
5	Demonstrate various Policy based Reinforcement Learning Algorithms. (L3)

Course Outcome versus **Program Outcomes**

CO	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	2	1	2		1								2		
CO2	2	1	2		3								2		
CO3	2	1	3		3								2		
CO4	2	1	3		3								2		
CO5	2	1	2		3								2		

S - Strongly correlated, *M* - *Moderately* correlated, *Blank* - *No correlation*

Assessment Methods	Assignment / Quiz / Mid-Test /Seminar/viva

Teaching- Learning & Evaluation

Week	Topic/ Contents	Course Outcomes	Sample questions	Teaching learning strategy	Assessment method & schedule
1	Introduction: Deep Reinforcement Learning, Suitability of RL, Components of Reinforcement Learning - Agent, Environment, Observations, Actions,	CO1	1. List and Explain the components of RL.	Lecture	Assignment-1, Test- 1 Quiz-1
2	Example-The Bandit Walk Environment, Agent-Environment interaction cycle, MDP (Markov Decision Process): The engine of the Environment-States, Actions, Transition Function, Reward Signal.	CO1	2. Define the following a) Agent b) Transition c) Reward d) Policy	Lecture	Assignment-1, Test- 1 Quiz-1
3	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.	CO2	3. Differentiate between various kinds of policies.	Lecture	Assignment-1, Test- 1 Quiz-1
4	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,	CO2	4. Distinguish between exploration and exploitation.	Lecture	Assignment-1, Test- 1 Quiz-1
5	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.	CO2	5. Explain about softmax exploration strategy.	Lecture	Assignment-1, Test- 1 Quiz-1
6	Model Free Reinforcement Learning: Monte Carlo Prediction (MC), First-Visit MC (FVMC),	CO3	6. Illustrate various types of model free reinforcement learning.	Lecture	Assignment-1,2, Quiz-1, Test-1, 2
7	Every-Visit MC (EVMC), Temporal Difference Learning (TD), Learning to estimate	CO3	7. Describe the process of learning to estimate from multiple steps.	Lecture	Assignment-2, Test- 2, Quiz-2

	from multiple steps, N-step TD learning, Forward-view TD(λ), Backward-view TD(λ), Generalized policy iteration(GPI),				
8	TEST 1				
9	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.	CO3	8. Define SARSA. Explain its implementation.	Lecture	Assignment-2, Test- 2, Quiz-2
10	Value Based Reinforcement Learning: Deep reinforcement learning agents with sequential feedback, evaluative feedback, sampled feedback,	CO4	9. Distinguish between various types of feedback in value based reinforcement learning.	Lecture	Assignment-2, Test- 2, Quiz-2
11	Function Approximation for Reinforcement Learning- high-dimensional state and action spaces, continuous state and action spaces,	CO4	10. Explain Function Approximation for Reinforcement Learning	Lecture	Assignment-2, Test- 2, Quiz-2
12	state-value function and action-value function with and without function approximation, Neural Fitted Q (NFQ), Deep Q-Network (DQN)	CO4	11. Distinguish between action –value function and state value function.	Lecture	Assignment-2, Test- 2, Quiz-2
13	Policy Based Reinforcement Learning: Policy Gradient and Actor-Critic Methods—REINFORCE Algorithm and Stochastic Policy Search,	CO4	12. Write a short notes on stochastic policy search.	Lecture	Assignment-2, Test- 2, Quiz-2
14	Vanilla Policy Gradient(VPG), Asynchronous Advantage Actor-Critic (A3C), Generalized Advantage Estimation (GAE), Advantage Actor-Critic(A2C),	CO5	13. Describe the generalized advantage estimation technique.	Lecture	Assignment-2, Test- 2, Quiz-2
15	Deep Deterministic Policy Gradient (DDPG), Twin-Delayed DDPG (TD3), Soft Actor-Critic (SAC).	CO5	14. Summarize the implementation challenges for Deep deterministic policy Gradient.	Lecture	Assignment-2, Test- 2, Quiz-2
16	TEST-2				