KUMARAGURU
college of technology
character is life

B.E/B.TECH DEGREE EXAMINATIONS: NOV/DEC 2022

(Regulation 2018)

Fifth Semester

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

U18AII5203**:** REINFORCEMENT LEARNING

COURSE OUTCOMES

CO1: Understand the underpinnings to structure classical solutions for Reinforcement Learning problem.

CO2: Apply deep learning architectures to train agents navigating from virtual world from sensory data.

CO3: Analyze basic Reinforcement Learning algorithms for simple sequential decision making and control problems in uncertain conditions.

CO4: Build system of agents to demonstrate collaboration or cooperation.

**Time: Three Hours**   **Maximum Marks: 100**

Answer all the Questions:-

PART A (10 x 2 = 20 Marks)

(Answer not more than 40 words)

1. Identify the properties for the task environment Medical Diagnosis.	CO1	[K$_3$]
2. Compare reinforcement learning with the other learning techniques.	CO1	[K$_1$]
3. Examine the reason behind the overestimation of Q-values by regular Q-learning and DQN.	CO2	[K$_4$]
4. Identify the tasks where Mote-Carlo methods can be applied.	CO1	[K$_3$]
5. Show the theory behind the evolutionary algorithms.	CO3	[K$_1$]
6. List few policy based methods in reinforcement learning.	CO3	[K$_1$]
7. Summarize leveraging neural networks to predict machine failures that learns intelligent behaviors from sensory data.	CO2	[K$_2$]
8. Apply min-max algorithm for the following two player game scenario.	CO4	[K$_3$]
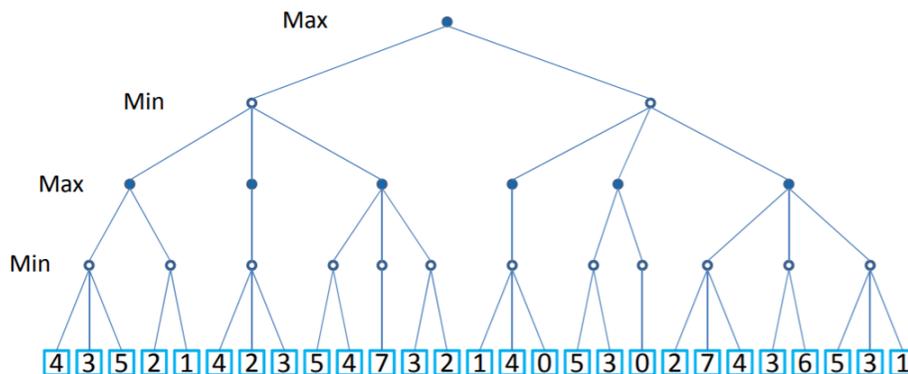
9. What is Hierarchical Reinforcement Learning?	CO4	[K$_1$]

10. Outline Reinforcement Learning in continuous space   CO4 [K₁]

Answer any FIVE Questions:-
PART B (5 x 16 = 80 Marks)
(Answer not more than 400 words)

11. a) Consider the below Frozen Lake problem.   8   CO2 [K₄]

S	F	F	F
F	H	F	H
F	F	F	H
H	F	F	G

(S: starting point, safe), (F: frozen surface, safe), (H: hole, fall to your doom),

(G: goal). The idea is to reach the goal from the starting point by walking only on a frozen surface and avoiding all the holes. Analyze using First-Visit and Every-Visit Monte- Carlo Methods.

   b) Explain Reinforcement Learning, its elements, the components and goal of RL 8   CO1 [K₁]
algorithm associated with it.

12. a) Experiment with Deep Q- Learning algorithm by considering the below CartPole 8   CO2 [K₃]
environment and apply Experience replay concept.
- The CartPole environment is a simple environment where the objective is to move a cart left or right in order to balance an upright pole for as long as possible.
- The state space is described with 4 values representing Cart Position, Cart Velocity, Pole Angle, and Pole Velocity at the Tip.

The action space is described with 2 values (0 or 1) allowing the car to either move left or right at each time step.

   b) Compare and contrast Deep Q-Learning, Deep Q-Network, Double DQN and 8   CO3 [K₁]
Dueling DQN with appropriate illustration.

13. a) Derive Bellman's Equation and solve the following problem using Markov 8   CO1 [K₃]
Decision Process.

3.3	8.8	4.4	5.3	1.5
1.5	3.0	2.3	1.9	0.5
0.1	0.7	0.7	0.4	-0.4
-1.0	-0.4	-0.4	-0.6	-1.2
-1.9	-1.3	-1.2	-1.4	-2.0

b) RL agents incrementally update their parameters while they observe a stream of experiences. In their simplest form, they discard incoming data immediately, after a single update. Analyze and provide solution for the issues associated with the above scenario.   8   CO3   [K4]

14. a) Solve the architectural summary of LeNet and AlexNet Convolutional Neural Networks Architectures.   8   CO2   [K3]

AlexNet (output of 1000 labels)

Layers		Feature Maps	Size	Kernel Size	Stride	Activation
Input/ Image		1	227x227x3	-	-	-
1		96				
		96				
2		256				
		256				
3		384				
4		384				
5		256				
		256				
6		-				
7		-				
8		-				
O/P		-				

LeNet (output of 10 class labels)

Layers		Feature Maps	Size	Kernel Size	Stride	Activation
Input/ Image		1	32x32	-	-	-
1		6				

		6				
2		16				
		16				
3		120				
4		-				
O/P		-				

b)	i)	State the key idea behind REINFORCE algorithms.	8	CO3	[K$_4$]
	ii)	Examine mathematically the need for Trajectory and how gradient ascent is going to affect the expected returns in REINFORCE algorithm.			

15. a) Compare and Contrast the following Reinforcement Learning Terminologies.   8   CO3   [K$_2$]

     i)   Value-based & Policy-based

     ii)   On-Policy & Off-Policy

     iii)   Model based & Model Free

   b) Explain Actor- Critic method and analyze why Optimization is much needed   8   CO4   [K$_4$]
using Deep Deterministic Policy Gradient (DDPG).

16. a) i) Examine how hierarchical reinforcement learning can be used to train   8   CO4   [K$_2$]
smarter agents.

     ii)   Show certain use cases in the field of robotics where various Deep Reinforcement Learning techniques could be applied.

   b) Analyze the Multiplayer environment for Markov games and explain how to train   8   CO4   [K$_4$]
an agent in collaborative and competitive environment

****