Indian Institute Of Technology Patna
Department of Computer Science and Engineering

CS-603 Reinforcement Learning
Mid Semester Examination(2023)

26th September 2023

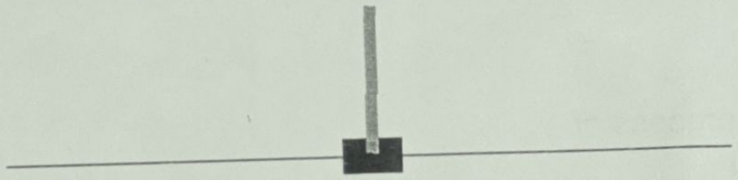
Full Marks: 60
Duration: 2 hours

Instruction
Make reasonable assumptions as and whenever necessary. You can answer the questions in any sequence. However, answers of all the parts to any particular question should appear together. Markings will be based on the correctness and soundness of the outputs. Proper indentation and appropriate comments (if necessary) are mandatory.

1.
   a. What are the different learning paradigms? How is reinforcement learning different from other learning paradigms?
   b. Can Rock-Paper-Scissor be designed as a RL problem ?

(3 + 2 + 2)

2.
   a. Define Markov Decision Process.
   b. A pole is attached by a free joint to a cart, which moves along a frictionless track. The pendulum is placed upright on the cart and the goal is to balance the pole by applying forces in the left and right direction on the cart.



Design an episodic game environment and a Markov Decision Problem for the above game.

(3 + 5)

3.
   a. Define dialogue system. What are the different components of a dialogue system?
   b. How can reinforcement learning be used to improve a dialogue system?

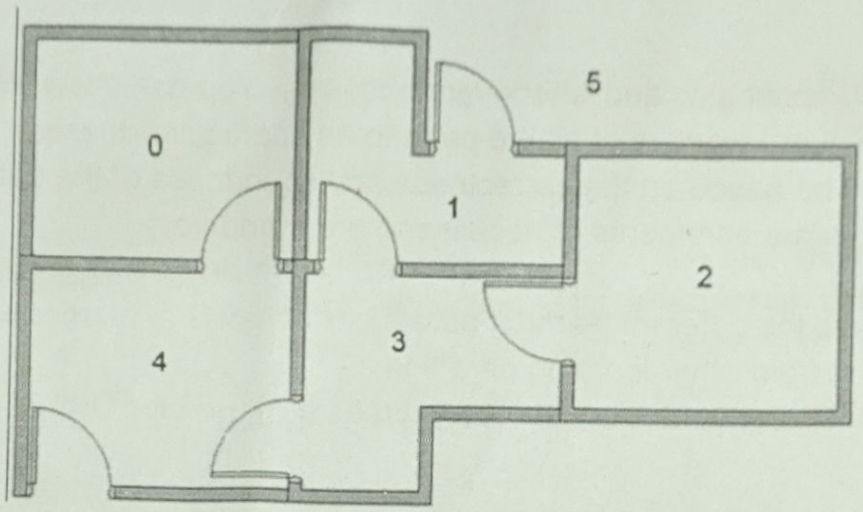
(2 + 3 + 3)

4.
   a. Define Bellman equation.
   b. What is cumulative reward? Why is cumulative reward more important than immediate reward ?
   c. Explain the effect of adding a constant c to all the rewards in (i)Episodic environment and (ii) Continual environment

(3 + 3 + 3)

5.

a. Consider five rooms in the apartments connected by **doors**. The goal is to leave the apartment. The outside of the apartment can be considered as one big room (marked as room no. 5) for ease of computation. Reaching outside the apartment should give an award of 100. Take epsilon = 0.5. Use tabular Q-Learning for three episodes.



(3)

b. Give an intuitive explanation of why tabular Q-learning tends to converge to a solution.

(3)

6. Explain any four ML/AI biases with examples. How can you prevent them?

(4 + 2)

7.

a. Explain the process of
   i. DQN
   ii. Double DQN
   iii. Duelling DQN

(5 + 5 + 5)

8. How does Prioritised Experience Replay Memory improve over Conventional Replay Memory?

(3)

Indian Institute Of Technology Patna
Department of Computer Science and Engineering

CS-603 Reinforcement Learning
End Semester Examination (2023)

30th November 2023

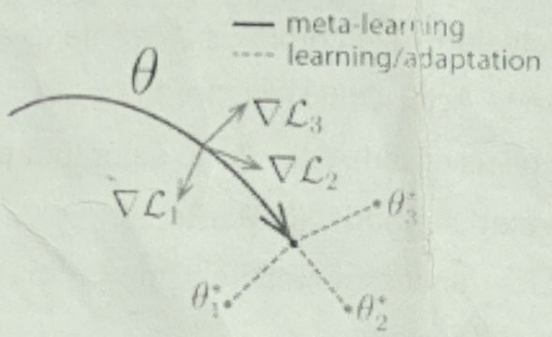
Full Marks: 80
Duration: 3 hours

Instruction
Make reasonable assumptions as and whenever necessary. You can answer the questions in any sequence. **However, answers of all the parts to any particular question should appear together**. Markings will be based on the correctness and soundness of the answers. Proper indentation and appropriate comments (if necessary) are mandatory.

1. POLICY GRADIENT:
    a. What are the disadvantages of Value Based RL? How is Policy Based RL different? Derive the objective function of Policy Gradient. (**6 Marks**)
    b. Write down the steps of the REINFORCE algorithm. What is the intuitive reasoning behind high variance in REINFORCE? (**6 Marks**)
    c. Does adding a baseline introduce bias in REINFORCE. How can we combine baseline and causality? (**4 Marks**)
2. ACTOR-CRITIC:
    a. Explain the objective of actor-critic algorithm with value baseline function in terms of mathematical equations. (**5 Marks**)
    b. What is the difference between Asynchronous Advantage Actor-Critic (A3C) and Synchronous Advantage Actor-Critic (A2C). (**3 Marks**)
    c. Discuss the advantages of using an Actor-Critic algorithm compared to pure policy-based and value-based methods. (**3 Marks**)
    d. How does the actor's policy function differ from the critic's value function in the Actor-Critic algorithm? (**3 Marks**)
3. RLHF AND LLMS
    a. What is RLHF ? Why is it required in the age of instruction tuning? What are the benefits and limitations of RLHF? (**5 Marks**)
    b. What are LLMs? What is the reason behind high performance of LLMs? Why is RLHF important in modern LLMs ? (**6 Marks**)

c. Explain with diagram how modern RLHF work ? (**3 Marks**)

4. HRL
   a. What is the fundamental idea behind Hierarchical Reinforcement Learning (HRL) in comparison to flat (non-hierarchical) RL approaches? (**3 Marks**)
   b. Explain the concept of "options" in the context of HRL. How are options defined, and how do they influence the learning process? (**3 Marks**)
5. RL FRAMEWORK
   a. Explain how Reinforcement learning is different from Supervised and Unsupervised Learning? Explain what is the "EXPLORE - EXPLOIT DILEMMA" in Reinforcement Learning? (**6 Marks**)
6. META LEARNING
   a. "Hyper-parameter optimization is a kind of meta-learning" - Comment (**2 Marks**)
   b. Given two hospitals X and Y. In which of the cases, can you apply meta learning and why? (**6 Marks**)
      i. X and Y have different laboratory tests.
      ii. X and Y have different population demographics.
      iii. X and Y have different specializations.
   c. How is the meta-evaluation different from the conventional evaluation process ? (**2 Marks**)
   d. In the context of meta-learning, explain the following diagram: (**3 Marks**)



   e. MAML requires second-order gradients, explain why? (**3 Marks**)
   f. Design a persona agnostic dialogue agent using meta learning. (**5 Marks**)
7. RL APPLICATION
   a. Why Reinforcement Training is needed for "Explainable Complaint Detection as Text-to-Text Generation Task" ? (**3 Marks**)