Indian Indian Institute of Technology Patna

Department of Computer Science & Engineering

Reinforcement Learning CS6131

Mid-Semester Examination

Date: 25.09.2024

Full Marks: 60
Duration: 2 hours

Instruction

Make reasonable assumptions as and whenever necessary. You can answer the questions in any sequence. However, answers to all the parts of 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. Explain the concept of "No Free Lunch" theorem in the context of training and testing in Machine Learning. **2 marks**

B. Consider the Atari game "Breakout," where the player controls a paddle at the bottom of the screen to bounce a ball upward to break bricks arranged at the top. The objective is to eliminate all the bricks without losing the ball below the paddle.

A reinforcement learning agent is being trained to play "Breakout" using two different reward models:

- Reward Model 1:

- If the agent successfully eliminates all the bricks (completes the level): +1 reward
- Else:0 reward

- Reward Model 2:

- If the agent successfully eliminates all the bricks (completes the level): +1 reward
- If the agent loses the ball and has no remaining lives (fail): 0 reward
- Else: Reward is proportional to the negative number of remaining bricks, i.e., Reward = (-d(S', G)) where:

(S') = Next state (number of bricks remaining)
(G) = Goal state (zero bricks remaining)
(d(S', G) =) Number of bricks remaining in (S')

i. How do the two reward models influence the agent's learning process differently?

ii. Which reward model is likely to enable the agent to learn the game more efficiently, and why? **6 marks**

C. **Briefly Describe the below topics:** **2*2=4 marks**

i. Hold-Out-set Evaluation Method

ii. Leave-One-Out Cross Validation

D. "Accuracy is not suitable for some applications"- Justy your answer **2 marks**

2. A. What is a policy in reinforcement learning, and what role does it play in the agent's behaviour? **3 marks**

B. Describe the exploration-exploitation trade-off in reinforcement learning with an example. **3 marks**

3. A. Explain the Markov assumption in the context of Markov Decision Processes (MDPs) and its Components **3 marks**

B. What is a discount factor? How does it affect the implementation of cumulative reward? **3 marks**

C. Suppose 15 men out of 300 men and 25 women out of 1000 are good orators. An orator is chosen at random. Find the probability that a male person is selected. Assume that there are equal numbers of men and women. **3 marks**

D. You toss a fair coin two times and let the random variable (X) represent the number of heads obtained. Determine the range of (X), denoted as (R_X), and find its probability mass function (P_X). **3 marks**

4. How does a double-deep Q-Network differ from a deep Q-Network? Why are two networks used in a double-deep Q network? Describe the algorithm of the Double Deep Q Network. **2+2+5= 9 marks**
5. What is Dueling DQN? Write down the training process of Dueling DQN. **2+5=7 marks**
6. What is the role of experience replay memory? What is Prioritized Experience Replay? **3+3= 6 marks**

7. Describe the algorithm of tabular Q-Learning; Describe the differences between reinforcement learning, supervised learning and unsupervised learning. **3+3=6 marks**