Innopolis University

Reinforcement Learning DRAFT

Time: TODO min.

Total Marks: TODO.

Name:

Instructions

- This is a pen-and-paper exam, remember to bring a pen, you can use calculator.
- No cheat sheets, notes, or books are allowed.
- Smart devices are strictly prohibited; possessing one during the exam, even if it is turned off, will be considered a cheating attempt.
- Talking is not permitted in the exam hall.
- Keep your focus on your own paper at all times.
- Please write neatly and clearly for better understanding.

DO NOT WRITE HERE, VARIANT IS 1						
Question	**Q1**	**Q2**	**Q3**	**Q4**	**Q5**	**Q6**
Grade						

1. Multiple-choice questions and calculation (points)
    a. Which of the following is an example of a model-free reinforcement learning algorithm? (point)
        i. Dynamic Programming
        ii. Q-Learning
        iii. Monte Carlo Tree Search (MCTS)
        iv. Policy Gradient Methods

    b. Consider a discounted MDP with a discount factor γ=0.8. If the agent receives rewards of +2 at each time step starting from time step t=0 until time step t=3, what is the total discounted future reward at time step t=0? (points)

        Ans = ……..

    c. In a grid-world environment, an agent follows the policy π that moves it left with a probability of 0.4 and right with a probability of 0.6. If the agent is currently in a state with a reward of -1 and moves to a neighboring state with a reward of +2, what is the expected immediate reward for this transition? (points)

        Ans = ……..

    d. Why is the ε-Greedy algorithm preferred over the greedy algorithm in reinforcement learning? (points)
        i. It only chooses the best action without randomness.
        ii. It avoids exploration and focuses solely on exploitation.
        iii. It balances exploration with exploitation, preventing the agent from being stuck on suboptimal actions.
        iv. It stops the agent from exploring at all times.

    e. Which type of problem is suited to reinforcement learning but not effectively addressed by supervised learning? (point)
        i. Problems with sequential decision-making, delayed rewards, and adaptation, like game playing and robotics.
        ii. Tasks with static datasets and labeled data, like image classification. iii. Situations requiring direct mapping from input to output, such as predicting sales data. iv. Cases that do not involve learning from rewards in an interactive environment.

f. Why is supervised learning generally not ideal for tasks like autonomous driving? (points)

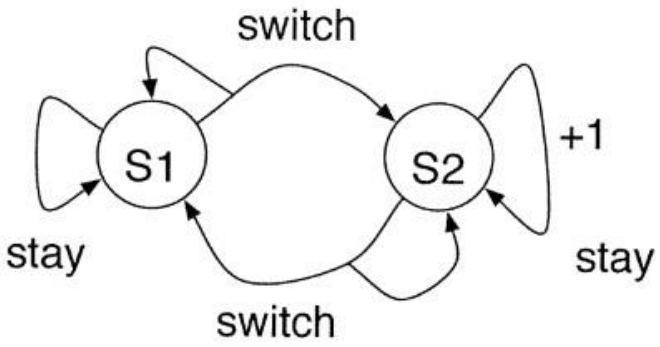
i. It always requires large labeled datasets.

ii. It lacks the ability to handle sequential actions where rewards are received over time.

iii. It does not work well with any static datasets.

iv. It cannot adapt to uncertain environments effectively.

2. Consider the following MDP. What is the optimal policy from both states S1 & S2? There are two states, S1 and S2, and two actions, switch and stay. (points)
Hint: The reward for action stays in state S2 is 1. All other rewards are 0.



3. Briefly explain the policy iteration process. What are the steps involved in achieving it? (points)

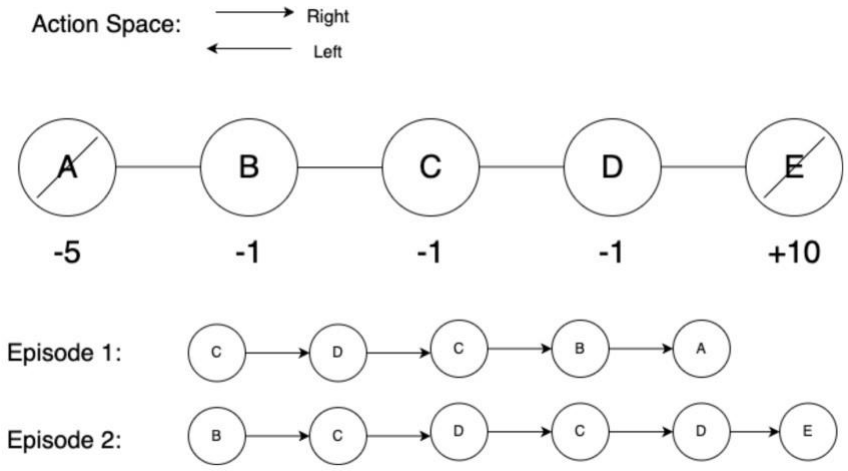
4. Sketch the backup diagrams for the following tabular learning methods: (points)

a. TD(0)

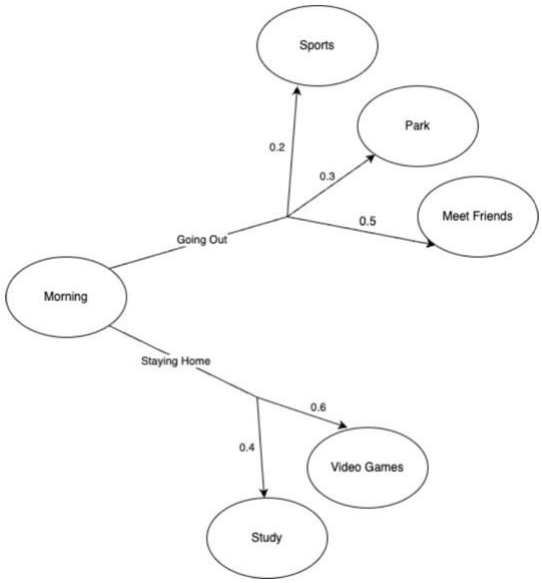
b. single-step full backup (DP)

c. Monte Carlo backup

5. Given the environment and episodes: a) calculate the value function **V(s)** using **TD(0)** error according to sample episodes. State A and E are terminal states. Let Learning-rate alpha is (0.5) and Discount factor gamma is (0.9). The number written underneath each state represents its reward. b) In the case of TD(lambda) which state will be updated the most? Why? (points)



6. Felix is trying to decide what to do with his day of from school. Use bellman equation to calculate action state value **Q(s, a)** for state *Morning* and tell Felix what action to choose. Note that numbers provided on graph represents transition probability. Let the discount factor be (1.0) and assume the reward for both actions is (3.0). (points)



State	V(S)
Sports	5
Park	2
Meet friends	4
Video games	1
Study	8