|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Roll No |  |  |  |  |  |  |  |  |  |  |  |  |

 ****

**Presidency University**

**Bengaluru**

 **SCHOOL OF ENGINEERING**

**MID TERM EXAMINATION SET A**

**Summer Semester**: 2023 - 24

**Course Code**: CSE 3011

**Course Name**: REINFORCEMENT LEARNING

**Program & Sem**: B. Tech & VII Sem (7CAI1)

**Date**: 31.10.23

**Time: 9.15 – 11.15am**

**Max Marks**: 60

**Weightage**: 30%

 **Instructions:**

1. *Read the all questions carefully and answer accordingly.*
2. *All questions are compulsory.*

**Part A [Memory Recall Questions]**

**Answer all the Questions. Each question carries 2 marks. (5Qx 2M= 10M)**

Q.NO. 1. Define the terms state space, action space, goal and reward. (C.O.1) [L1]

Q.NO. 2. Define the Bellman equation for V(s) in a stochastic environment with a deterministic policy. (C.O.1) [L1]

Q.NO. 3. Define the return of an episode in an episodic task with an example. (C.O.1) [L1]

Q.NO. 4. Differentiate finite and infinite horizon. (C.O.1) [L2]

Q.NO. 5. Differentiate on policy and off policy in Monte-Carlo control techniques. (C.O.2) [L2]

 **Part B [Thought Provoking Questions]**

**Answer all the Questions. Each question carries 10 marks. (3Qx10M=30M)**

Q.NO. 6. Find the value of all states in the environment given below using Bellman equation, where s2 and s3 are the final states. (C.O.1) [L3]

 A0

P(s1|s0,A0) = 1

 A1

P(s2|s1,A1) = 0.7

$$R\_{s1s2}^{A1}=3$$

 A1

P(s3|s1,A1) = 0.3

$$R\_{s1s3}^{A1}=+1$$

$$R\_{s0s1}^{a0}=2$$

Q.NO. 7. The RL agent has to find an optimal policy in an environment whose model dynamics is known. Name the algorithm(s) for the same. Explain one of them in detail. (C.O.1) [L2]

Q.NO. 8 For the environment given: S : {s0,s1,s2} where s2 is the goal state A : {a0,a1,a2,a3}.

Input policy : $ π\left(s0\right)=0.8$ $π\left(s0\right)=0.1$ $π\left(s0\right)=0.1$ $ π\left(s1\right)=1$

Reward functions: $R\left(s0,a0,s1\right)=5$ $R\left(s0,a1,s1\right)=2$ $R\left(s0,a2,s1\right)=2$ $R\left(s1,a3,s2\right)=1$

1. identify the type of the policy used in each state
2. Using Monte-Carlo prediction, evaluate the given policy by finding V(s).

. (C.O.2) [L2]

 **Part C [Problem Solving Questions]**

**Answer all the Questions. Each question carries 20 marks. (1Qx20M=20M)**

Q.NO.9. Implement the Frozen Lake Environment using a random policy for the agent. Show the output of the following:

a. create and render the environment b. print the State Space and action Space

c. Find P(action = right, state = A)

d. Generate 30 episodes and find the total return of episodes, which are multiples of 3. Print the output in the following format

|  |  |
| --- | --- |
| Episode # | Return |
|  |  |

 (C.O.1) [L3]