|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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

**Date**: 30.10.23

**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 with respect to the grid-world environment. (C.O.1) [L1]

Q.NO. 2. Differentiate deterministic and stochastic policy in reinforcement learning.

 (C.O.1) [L2]

Q.NO. 3. What is discount factor? What happens when it is 0 and 1? (C.O.1) [L1]

Q.NO. 4. Define an ‘episode’ and the ‘return’ of an episode with an example. (C.O.1) [L1]

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. For the environment given below

S : {s0,s1,s2} where s2 is the goal state A : {a0,a1}.

Input policy : $ π\left(s0\right)=0.5$ $π\left(s0\right)=0.5$ $ π\left(s1\right)=1$

Reward functions: $R\left(s0,a0,s1\right)=1$ $R\left(s0,a1,s1\right)=3$ $R\left(s1,a1,s2\right)=1$

1. identify the type of the policy used in each state
2. Using Monte-Carlo prediction, evaluate the given policy

 (C.O.2) [L3]

Q.NO. 7. Using the value iteration algorithm and the model dynamics of state A given in the table below, find the optimal value of state A, after the first iteration. The state and action space of the environment are S : {A,B,C} and A : {0,1} (C.O.1) [L3]

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  State (s) | Action (a) |  Next state (s’) | Transition probability $P\_{ss'}^{a}$ | Reward function$$R\_{ss'}^{a}$$ |
| A | 0 | A | 0.1 | 0 |
| A | 0 | B | 0.8 | -1 |
| A | 0 | C | 0.1 | 1 |
| A | 1 | A | 0.1 | 0 |
| A | 1 | B | 0.0 | -1 |
| A | 1 | C | 0.9 | 1 |

Q.NO. 8. In the grid world environment given below, the following policy is used by the agent in each state. 𝜋(down|𝐴) = 0.8 𝜋(right|𝐴) = 0.20 𝜋(right|D) = 1 𝜋(right|E) = 1

 𝜋(down|B) = 1 𝜋(down|F) = 1

Find V(A) using Bellman equation.



 (C.O.1) [L3]

 **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. State Space and action Space

c. generate 20 episodes using a random policy. d. Find the total return of the first 10 episodes.

 (C.O.1) [L3]