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

 ****

**Presidency University**

**Bengaluru**

 **SCHOOL OF ENGINEERING**

**MID TERM EXAMINATION SET B**

**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 environments in reinforcement learning with an example.

 (C.O.1) [L2]

Q.NO. 3. What is discount factor? What happens when it is 0.2 and 0.9? (C.O.1) [L1]

Q.NO. 4. Define the function Q(s, a) in an episodic task, with an example. (C.O.1) [L1]

Q.NO. 5. Define the prediction and control tasks in reinforcement learning. (C.O.2) [L1]

 **Part B [Thought Provoking Questions]**

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

Q.NO. 6. For the environment given below (C.O.2) [L3]

S : {s0,s1,s2} where s2 is the goal state A : {a0,a1,a2,a3}.

Input policy : $ π\left(s0\right)=0.7$ $π\left(s0\right)=0.2$ $π\left(s0\right)=0.1$ $ $

 $ π\left(s1\right)=1$

Reward functions: $R\left(s0,a0,s1\right)=5$ $R\left(s0,a1,s1\right)=3$ $R\left(s0,a2,s1\right)=1$ $R\left(s1,a3,s2\right)=2$

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

Q.NO. 7. Write the Bellman equation of the value function of a state in a deterministic and stochastic environment. Find the value of all the states in the trajectory given below using Bellman equation. Assume γ = 1.

a2

a1

a3

a0

S

-1

+3

+2

+1

 (C.O.1) [L3]

 Q.NO. 8. Assuming a random initial policy for state A as action 0, using policy iteration algorithm and the model dynamics of state A given in the table below, find V(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 |

 **Part C [Problem Solving Questions]**

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

Q.NO.9. Implement the Cart-Pole balancing 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. generate 20 episodes using a random policy. d. Find the total return of the first 10 episodes.

 (C.O.1) [L3]