|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 (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 ‘reward’ and ‘return’ with a suitable example for both. (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. Write the Bellman equation for Q(s, a) in a stochastic environment. (C.O.1) [L1]

Q.NO. 5. Differentiate the prediction and control tasks in reinforcement learning. (C.O.2) [L2]

 **Part B [Thought Provoking Questions]**

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

Q.NO. 6. Write the Bellman equation of the value function of a state in

i) a deterministic environment with a deterministic policy

ii) a stochastic environment with a deterministic policy (C.O.1) [L3]

 Find the value of all the states in the trajectory given below using Bellman’s recursive definition of V(s).

a0

+2

a1

a2

a3

+2

-1

+3

Q.NO. 7. Using MC prediction evaluate the given policy in the environment whose details are given below. (C.O.2) [L3]

State space : {s0,s1,s2} where s2 is the final state. Action space : {a0,a1}

Policy of the states : π(a0|s0) = 0.5 π(a1|s0) = 0.5 π(a1|s1) = 1.0

Reward function: R(s0,a0,s1) = 1 R(s0,a1,s1) = 3 R(s1,a1,s2) = 1

Q.NO. 8. 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]

 **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. State Space and action Space

c. generate 20 episodes using a random policy.

d. Find the total return of the episodes which are multiples of 5. Print the output in the following format.

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

 (C.O.1) [L3]