Assignment 2 (*10 points*)

Markov Decision Process (MDP) and Value-iteration

Consider a 4x4 grid-world. Here, the agent starts in state s1,1 and must reach the goal in state s4,4. There are two danger-states in s3,2 and s3,4. The reward for the transition into danger-states is -1 and the reward for transitioning into goal-state is +10 and let the living reward be -0.1 for every other state. Consider possible actions as . Define as 0 for all non-terminal states and +10 for the goal state and -1 for the two danger-states. Let the discount factor be . Consider the probability of transitioning to the intended state as 70% and every other state as 10%.

**NOTE**: Please use “Equation” in Word for math notations

Table 1: Given 4x4 grid world

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

**Tasks**:

1. Write down the equation for value-iteration. Define each term and explain in your own words what each term is (*1 point*)

|  |
| --- |
| * **=** The value of state at iteration * **=** The maximization over all possible actions 𝑎 to find the action that yields the maximum value. * **=** The probability of transitioning to state from state given action . This represents the uncertainty in the environment and is part of the Markov Decision Process. * **=** The reward received after transitioning from state 𝑠 to state given action . * **=** The discount factor which reflects the importance of future rewards. * **=** The next state. * **=** The value of the next state 𝑠′ from the previous iteration (𝑘−1) |

1. Convert the given problem statement into math notations and define States, Actions, Reward Function and Transition Function (*1 point*)

|  |
| --- |
|  |

1. Perform value-iteration for 5 iterations and calculate and for all 16 states (*8 points*)
2. Please enter values for (*2 points*)

|  |  |  |  |
| --- | --- | --- | --- |
| -0.10 | -0.10 | -0.10 | -0.10 |
| -0.10 | -0.28 | -0.10 | -0.28 |
| -0.28 | -1.90 | -0.46 | -1.90 |
| -0.10 | -0.28 | 13.27 | 19.00 |

1. Please enter values for (*2 points*)

|  |  |  |  |
| --- | --- | --- | --- |
| -0.28 | -0.30 | -0.28 | -0.30 |
| -0.32 | -0.61 | 4.67 | -0.63 |
| -0.63 | -3.44 | 11.85 | -3.44 |
| 4.83 | 12.91 | 27.26 | 34.39 |

1. Please enter values for (*2 points*)

|  |  |  |  |
| --- | --- | --- | --- |
| -0.43 | 1.79 | 4.63 | 1.78 |
| 1.59 | 3.92 | 11.09 | 4.17 |
| 5.20 | -4.69 | 20.62 | -4.69 |
| 13.22 | 22.87 | 38.91 | 46.85 |

1. Please enter values for (*2 points*)

|  |  |  |  |
| --- | --- | --- | --- |
| Action Right: -0.44  Action Left: -0.43  Action Up: -0.43  Action Down: -0.46 | Action Right: 1.79  Action Left: 0.10  Action Up: 0.09  Action Down: 1.59 | Action Right: 0.52  Action Left: 0.52  Action Up: 2.22  Action Down: 4.63 | Action Right: 0.08  Action Left: 1.78  Action Up: 0.08  Action Down: 1.56 |
| Action Right: 1.56  Action Left: 0.04  Action Up: 0.07  Action Down: 1.59 | Action Right: 3.92  Action Left: -0.21  Action Up: -0.19  Action Down: -2.73 | Action Right: 3.35  Action Left: 3.37  Action Up: 3.57  Action Down: 11.09 | Action Right: 1.54  Action Left: 4.17  Action Up: 0.06  Action Down: -2.48 |
| Action Right: -2.33  Action Left: 1.75  Action Up: 0.19  Action Down: 5.20 | Action Right: -4.69  Action Left: -4.69  Action Up: -4.69  Action Down: -4.69 | Action Right: -0.12  Action Left: -0.12  Action Up: 6.54  Action Down: 20.62 | Action Right: -4.69  Action Left: -4.69  Action Up: -4.69  Action Down: -4.69 |
| Action Right: 13.22  Action Left: 8.14  Action Up: 4.69  Action Down: 8.14 | Action Right: 22.87  Action Left: 9.66  Action Up: 2.13  Action Down: 14.75 | Action Right: 38.91  Action Left: 20.61  Action Up: 19.75  Action Down: 28.73 | Action Right: 46.85  Action Left: 46.85  Action Up: 46.85  Action Down: 46.85 |