

Assignment No. 7

EECS 658

Introduction to Machine Learning

Due: 11:59 PM, Tuesday, November 23, 2021

Submit deliverables in a single zip file to BlackBoard

Name of the zip file: FirstnameLastname\_Assignment7 (with your first and last name)

Name of the Assignment folder within the zip file: FirstnameLastname\_Assignment7

Deliverables:

1. Copy of Rubric7.docx with your name and ID filled out (do not submit a PDF)
2. Python source code.
3. Screen print showing the successful execution of your Python code. (Copy and paste the output from the Python console screen to a Word document and PDF it).
4. Answer to Part 1, Question 1.
5. Answer to Part 2, Question 2.

Assignment:

- For both parts, we are going to use a modified version of the Gridtask World described in the lectures. The only difference is that the grid is a 5-by-5 grid instead of a 4-by-4 grid:

	1	2	3	4
5	6	7	8	9
10	11	12	13	14
15	16	17	18	19
20	21	22	23	

- Everything else is the same:
  - The goal of the Gridworld Task is for a robot, starting at any square in the grid, to move through the grid and end up in a termination state (grey squares) which ends the game.
  - Each grid square is a state (s).
  - The actions (a) that can be taken are up, down, left, or right.
  - The rewards (r) are:
    - -1 on all transitions
    - 0 terminal state
    - -1 when it hits a wall (no transition, but still a reward)
  - Assume:
    - $p(r_{t+1}|s_t, a_t) = 0.25$
    - $p(s_{t+1}|s_t, a_t) = 0.25$
    - $\gamma = 1$

### Part 1: RL Policy Iteration Algorithm

- Write a Python program that uses the RL Policy Iteration algorithm to develop an optimal policy ( $\pi^*$ ).
- The program should display the optimal policy ( $\pi^*$ ) as a Python array similar to this:

```
[[0.0 -0.9 -0.8 -0.7 -0.6]
 [-0.9 -0.9 -0.8 -0.7 -0.6]
 [-0.9 -0.9 -0.8 -0.7 -0.6]
 [-0.9 -0.9 -0.8 -0.7 -0.6]
 [-0.9 -0.9 -0.8 -0.7 0.0]]
```

- The policy the robot follows, no matter what square it is in, is to go to the square next to it with the highest value.
- If it follows this policy, it will end up in one of the termination squares in the least amount of moves.
- Note: The values in this array are NOT the ones you will get.
- The program should print out the policy array with the iteration number for iteration 0 (the initial values), iteration 1, iteration 10, and the final iteration.
- Determine a method for deciding when the Policy Iteration algorithm has converged.
- Question 1: Explain the convergence method and why you picked it. There is no wrong answer. You will get credit for any method you pick as long as it converges and you provide a reasonable explanation of why you picked it.
- For help, consult: <https://towardsdatascience.com/reinforcement-learning-rl-101-with-python-e1aa0d37d43b>

### Part 2: RL Value Iteration Algorithm

- Write a Python program that uses the RL Value Iteration algorithm to develop an optimal policy ( $\pi^*$ ).
- The program should display the optimal policy ( $\pi^*$ ) as a Python array similar to this:

```
[[ 0 -1 -2 -3 -4]
 [-5 -6 -7 -8 -9]
 [-5 -6 -7 -8 -9]
 [-5 -6 -7 -8 -9]
 [-5 -6 -7 -8 0]]
```

- The policy the robot follows, no matter what square it is in, is to go to the square next to it with the highest value.
- If it follows this policy, it will end up in one of the termination squares in the least amount of moves.
- Note: The values in this array are NOT the ones you will get.
- The program should print out the policy array with the iteration number for iteration 0 (the initial values), iteration 1, iteration 2, and the final iteration.
- Determine a method for deciding when the Value Iteration algorithm has converged.

- Question 2: Explain the convergence method and why you picked it. There is no wrong answer. You will get credit for any method you pick as long as it converges and you provide a reasonable explanation of why you picked it.

Remember:

- Your Programming Assignments are individual-effort.
- You can brainstorm with other students and help them work through problems in their programs, but everyone should have their own unique assignment programs.