Assignment 2

EE675: Introduction to Reinforcement Learning

March 4, 2025

Instructions

- Kindly name your submission files as '`RollNo_Name_A2.ipynb`', based on the part you are submitting. Marks will be deducted for all submissions that do not follow the naming guidelines.
- You are required to work out your answers and submit only the iPython Notebook. The code should be well commented and easy to understand as there are marks for this.
- You may use the notebook given along with the assignment as a template. You are free to use parts of the given base code but may also choose to write the whole thing on your own.
- Submissions are to be made through HelloIITK portal. Submissions made through mail will not be graded.
- Answers to the theory questions, if any, should be included in the notebook itself. While using special symbols use the $\LaTeX$ mode
- Make sure your plots are clear and have title, legends and clear lines, etc.
- Plagiarism of any form will not be tolerated. If your solutions are found to match with other students or from other uncited sources, there will be heavy penalties and the incident will be reported to the disciplinary authorities.
- In case you have any doubts, feel free to reach out to TAs for help.

Policy Iteration and Value Iteration

(Stochastic Maze) `[15 Marks]` The stochastic maze environment is shown in the figure . There are a total of 12 states in the environment represented by the indices $\{0, 1, \cdots, 11\}$. The agent starts in the initial state 0. Four actions are possible in each state: left, right, up, and down. The environment is stochastic and we take the intended action only with a probability $p = 0.8$. We take an orthogonal action with a probability $p = 0.2$. If we collide with the edge of the environment or the wall present in state 5, the agent comes back to the same state. The transition into the goal state 3 has a +1 reward and transition into the hole state 7 has a -1 reward. All other transitions have a -0.01 reward associated with them.

Given an implementation of the environment in the template notebook, answer the following questions:

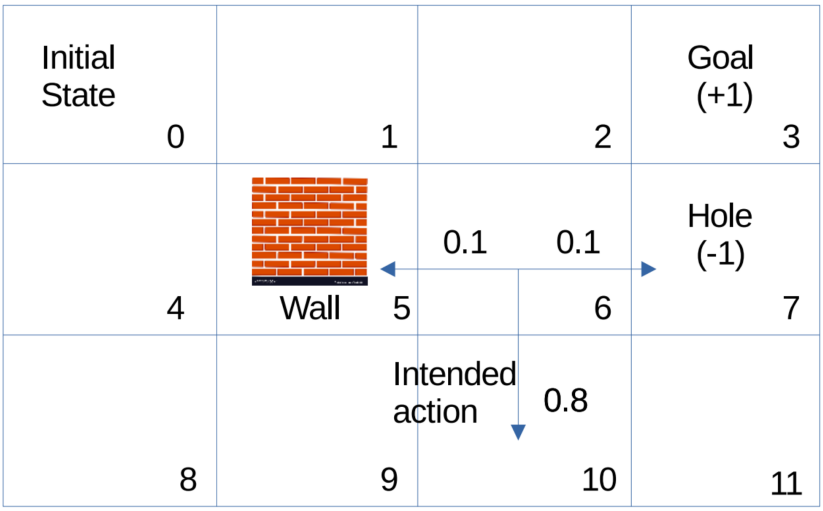

Figure 1: Stochastic Maze Environment

1. Find an optimal policy to navigate the given environment using Policy Iteration (PI) `[10 Marks]`
2. Find an optimal policy to navigate the given environment using Value Iteration (VI) `[3 Marks]`
3. Compare PI and VI in terms of convergence. Is the policy obtained by both same? `[2 Marks]`