Programming Assignment 3

EE932: Introduction to Reinforcement Learning

May 23, 2024

Instructions

- **Assignment Deadline - 5th Jun 2024**
- Kindly name your submission files as ‘`RollNo_Name_PA3.ipynb`‘.
- 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 iPearl 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.

Cliff Walking (Deadline - 5th Jun 2024)

`[10 Marks]` Through this grid world exercise we will compare SARSA and Q-learning algorithms, highlighting the difference between them. Consider the grid world shown in the Figure below. This is a standard undiscounted, episodic task, with start and goal states, and the usual actions causing movement up, down, right, and left. Reward is -1 on all transitions except those into the the region marked “The Cliff.” Stepping into this region incurs a reward of -100 and sends the agent instantly back to the start. The episode ends when the agent reaches the goal state.

Given the template notebook skeleton code, Implement and answer the following questions in the notebook:

1. Implement and compare the SARSA and Q-Learning methods `[3.5+3.5 Marks]`

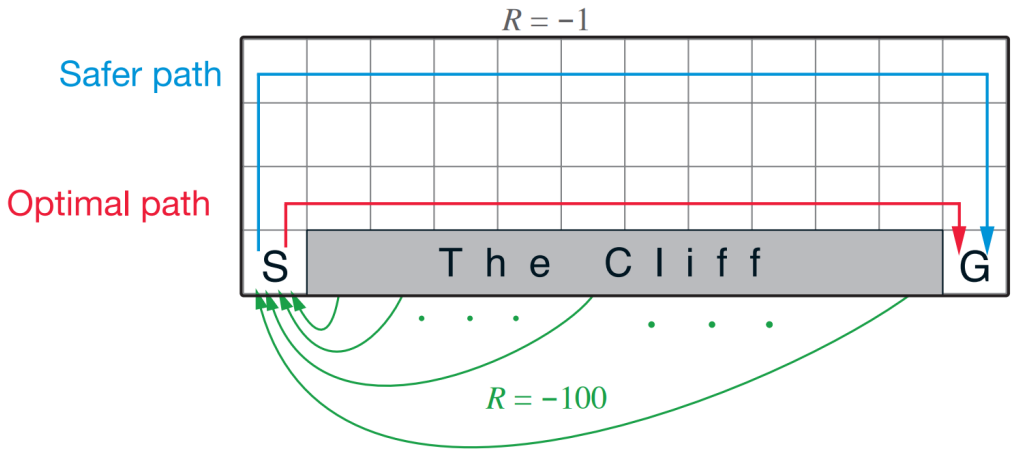

Figure 1: Cliff Walking Environment

2. Why is Q-learning considered an off-policy control method? `[2 Marks]`
3. Which algorithm takes a safer path? Why? `[1 Marks]`
4. (**Bonus**) Suppose while learning the action selection is greedy. Is Q-learning then exactly the same algorithm as SARSA? Will they make exactly the same action selections and weight updates? `[2 Marks]`