Reinforcement Learning Assignment 3

Maida Cao (TA), Prof. Zou

2021-4-8

1 Introduction

The goal of this assignment is to do experiment with model-free control, including on-policy learning (Sarsa) and off-policy learning (Q-learning). For deep understanding of the principles of these two iterative approaches and the differences between them, you will implement *Sarsa* and *Q-learning* at the application of the Cliff Walking Example, respectively.

2 Cliff Walking

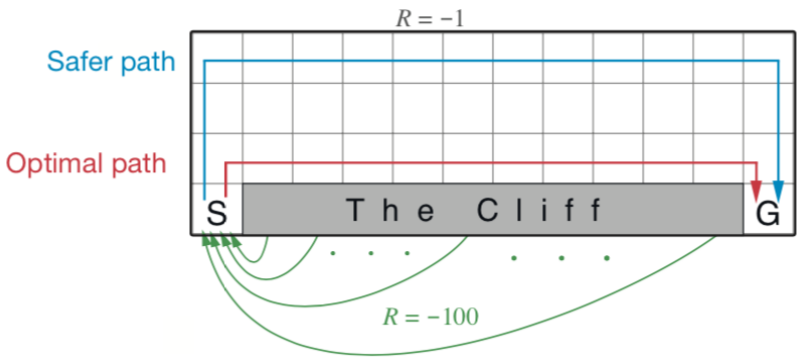

Figure 1: Cliff Walking

Consider the gridworld shown in the Figure 1. This is a standard undiscounted, episodic task, with start state (S), goal state (G), and the usual actions causing movement up, down, right, and left. Reward is -1 on all transitions except those into the region marked "The Cliff". Stepping into this region incurs a reward of -100 and sends the agent instantly back to the start.

3 Experiment Requirments

- Programming language: python3
- You should build the Cliff Walking environment and search the optimal travel path by Sara and Q-learning, respectively.
- Different settings for ϵ can bring different exploration on policy update. Try several ϵ (e.g. $\epsilon = 0.1$ and $\epsilon = 0$) to investigate their impacts on performances.

4 Report and Submission

- Your reports and source code should be compressed and named after "studentID+name".
- The files should be submitted on Canvas before Apr. 16, 2021.