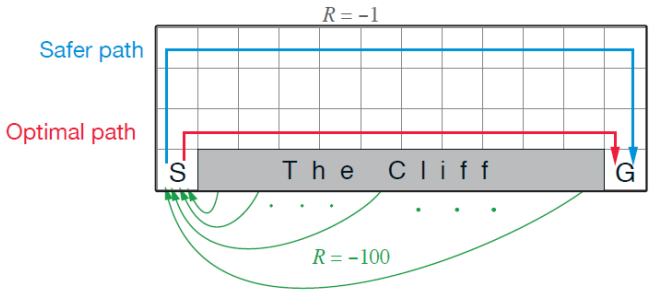
Reinforcement Learning: Programming Exercise #2

In this programing exercise we are going to solve the Cliff Walking problem explained in Example 6.6 page 132 of the textbook.



You need to apply Q-learning and SARA to this problem and show that:

1- SARSA converges to the blue path and Q-learning converges to the red path (shortest path). Explain why that is?

2- Generate the plot of sum of the rewards as a function of episodes. Explain why Q-leaning converges to lower average rewards even though it can find the optimal path?

3- For both methods gradually reduce the ϵ (in epsilon-greedy) and show that both algorithms converge to optimal path and explain why.