SARSA vs Q-Learning

Exercise 5.5: SARSA vs QLearning

In this exercise, you will compare the performance of your SARSA agent with your Q-Learning agent.

Make sure that you have:

1. Completed the setup requirements as described in the Set Up Lab Environments section
2. Completed the previous exercises in this lab

By now, you should have implemented the Sarsa and the QLearning agent. As you've probably noticed, both the agents perform comparably within the SimpleRoomsEnv, CliffWalkingEnv, and WindyGridworldEnv environments.

1. Let's compare both agents in more details using the CliffWalkingEnv environment.
2. Set up an experiment with your **SARSAAgent** and the CliffWalkingEnv environment.
3. Use the default values for alpha, epsilon, and gamma for your SARSAAgent, and run the the experiment say for **1000** episodes with the interactive set to **False**.
4. Run this experiment several times and observe the results.

5. Now, set up an experiment with your **QLearningAgent** and the CliffWalkingEnv environment.
6. Use the default values for alpha, epsilon, and gamma for your QLearningAgent, and run the the experiment say for **1000** episodes with the interactive set to **False**.
7. Run this experiment several times and observe the results.

Lab Question

1/1 point (graded)

Let's jog our memory a little bit. What is the minimum steps required to reach the goal in this environment?

○ 1

○ 12

● 13 ✓

○ 15

○ 48

Submit

You have used 1 of 2 attempts

✓ Correct (1/1 point)

Lab Question

1/1 point (graded)

Based on your observation of the above experiments, on average, which agent reach the goal within the minimum steps required more times that the other?

○ Sarsa

● QLearning ✓

Submit

You have used 1 of 2 attempts

✓ Correct (1/1 point)

Lab Question

1/1 point (graded)

Based on your observation of the above experiments, on average, which agent fails more times than the other after 100 episodes?

○ Sarsa

● QLearning ✓

Submit

You have used 1 of 2 attempts

✓ Correct (1/1 point)

"After an initial transient, Q-learning learns values for the optimal policy, that which travels right along the edge of the cliff. Unfortunately, this results in its occasionally falling off the cliff because of the epsilon-greedy action selection. Sarsa, on the other hand, takes the action selection into account and learns the longer but safer path through the upper part of the grid. Although Q-learning actually learns the values of the optimal policy, its on-line performance is worse than that of Sarsa, which learns the roundabout policy. If epsilon were gradually reduced, then both methods would asymptotically converge to the optimal policy."

-- Sutton & Barto