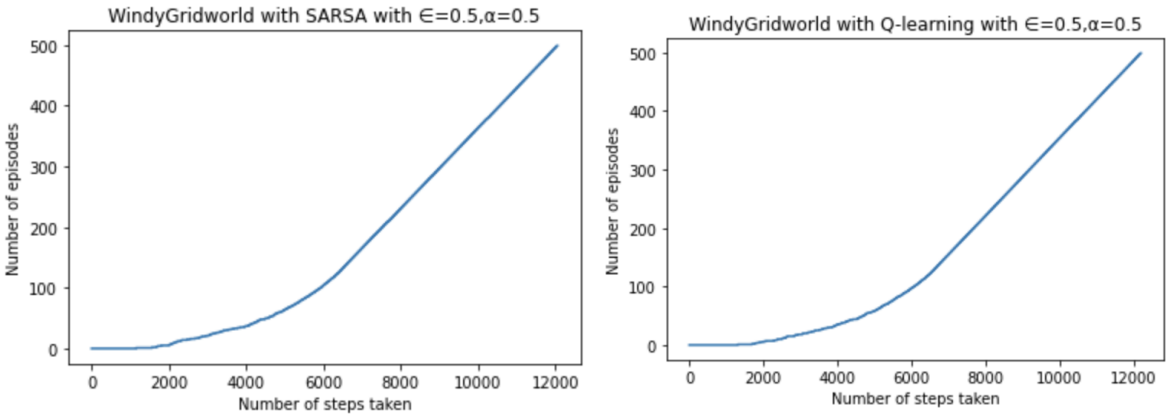
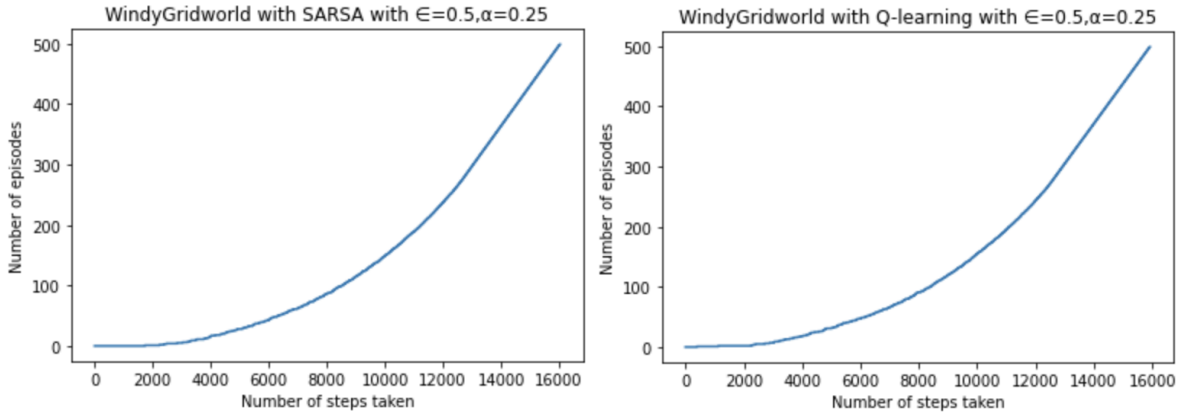
Part 1- Windy Grid World with SARSA and Q-Learning

Parameters: epsilon = 0.5, alpha = 0.5, gamma = 1

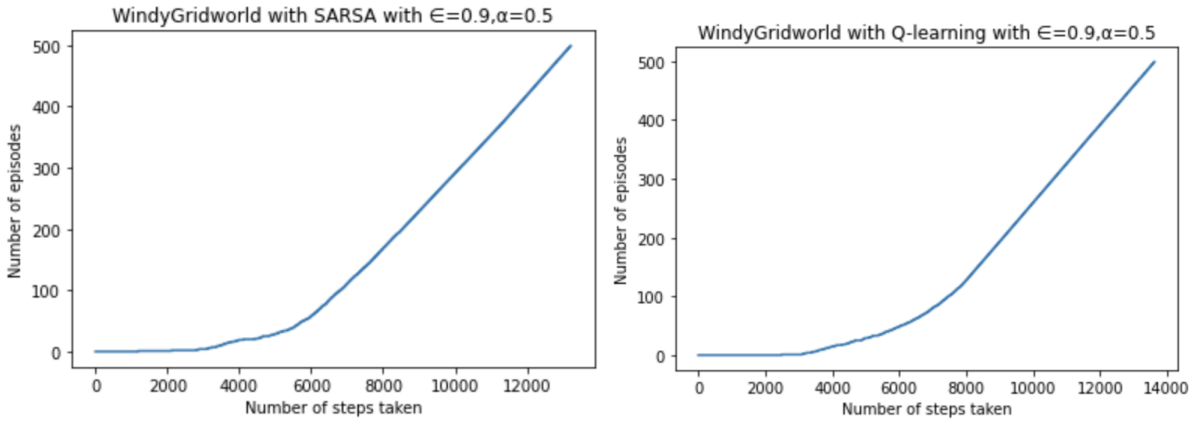
To guarantee convergence, I tried different numbers ascendingly for the episode length parameter for SARSA and Q-Learning. After several trials, I set the length of episodes to 500. Then with the discounted factor gamma = 1, I generate multiple figures of the relative speeds of learning rates in terms of the number of steps taken for 500 episodes with different alpha and epsilon. As shown in the figures below, the one with epsilon = 0.5 and alpha = 0.5 requires the least steps to finish the episodes for both SARSA and Q-Learning. Therefore, epsilon = 0.5 and alpha = 0.5 are chosen for the following tests.



epsilon = 0.5, alpha = 0.5 (approx. 12000 steps)



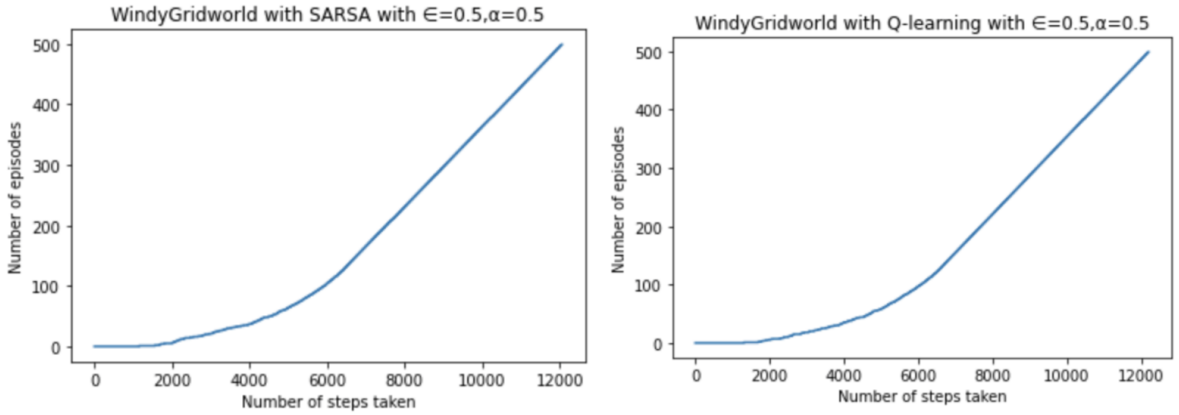
epsilon = 0.5, alpha = 0.25 (approx. 16000 steps)



epsilon = 0.9, alpha = 0.5 (approx. 13000 steps)

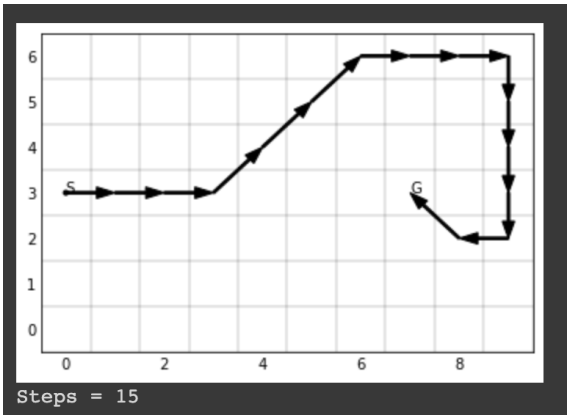
Comparison between SARSA and Q-Learning

To compare between the SARSA and the Q-Learning algorithm, let's take a close look at the following figures. We can tell SARSA takes slightly less steps than Q-learning to finish the episodes. Furthermore, by observing the change of slopes, we can tell that SARSA starts earlier to speed up learning than Q-learning. This reflects that SARSA learns faster than Q-Learning in this environment.



Optimal Poicy

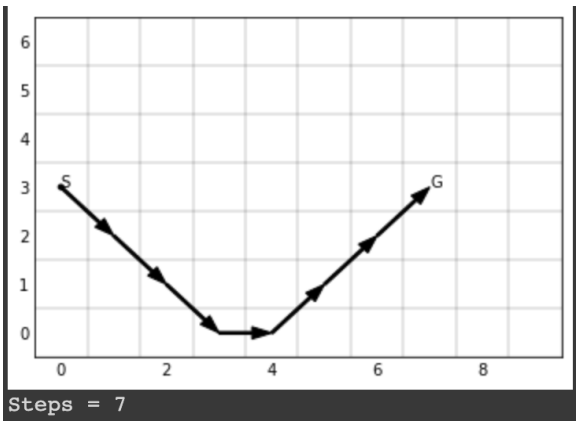
The optimal policy for SARSA and Q-Learning are the same, which both take 15 steps. The trajectory is shown as below.



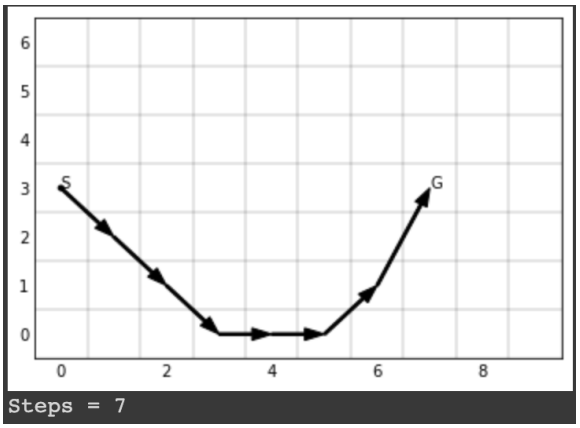
Part 2- Windy Grid World with King's moves and stochastic transition

Comparison between SARSA and Q-Learning

With King's moves, the performances for both SARSA and Q-learning are better than them without. They only take 7 steps to reach the goal state. Trajectories of optimal policies are shown below.



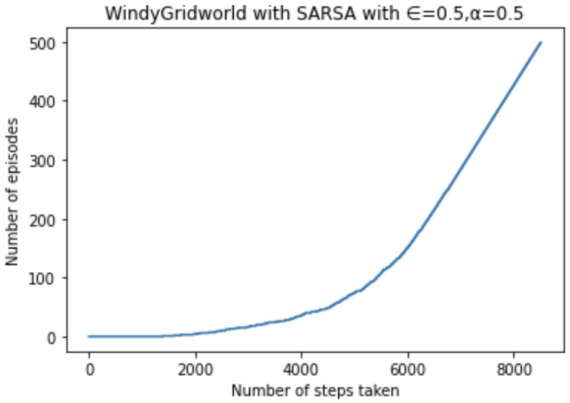
SARSA

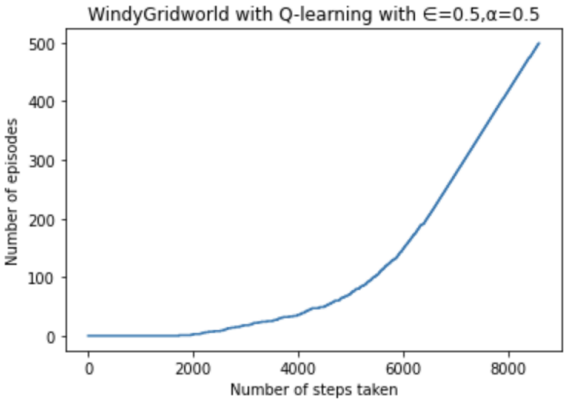


Q-Learning

Optimal Policy

The steps taken for 500 episodes are about 8000, rather than 12000 steps in the previous environment. Between SARSA and Q-Learning, the patterns are the same as in part 1, which SARSA takes slightly shorter steps and it learns faster by having steeper slopes early.


WindyGridworld with SARSA with ∈=0.5,α=0.5
Number of episodes
Number of steps taken
0
100
200
300
400
500
0
2000
4000
6000
8000


WindyGridworld with Q-learning with ∈=0.5,α=0.5
Number of episodes
Number of steps taken
0
100
200
300
400
500
0
2000
4000
6000
8000