Reinforcement Learning Assignment 3

Those are the results of my implementation.

Sarsa, $\epsilon = 0.1$

Sarsa, $\epsilon = 0.01$

Sarsa, $\epsilon = 0$

Q-learning, $\epsilon = 0.1, 0.01, 0$

From those above screenshots, we can tell that when ϵ is large, Sarsa intends to choose a safer path. That is because when Sarsa calculates values of grids near the cliff, it is more likely to take the cliff reward into account, so those grids will have lower rewards.

Q-learning always gets the optimal path, no matter what value ϵ is. Moreover, when ϵ is 0, Sarsa gets the same result as Q-learning.