1. World without thief
   (a) Discount factor = 0.9, learning rate = 0.1, e = 0(does not perform random actions) Simulate 1000 episodes of 10000 steps. Explain observation.

   This epsilon is 0 means the robot will avoid the slippery area and go to the grid with best utility. The first episode is -8. The remaining episodes are all zeroes. However, the second column was blocked by slippery area. The robot slipped and never wants to go through second column in later episodes. That's why the remaining 999 episodes are all zeros.

   (b) e = 0.1 , keep others, Explain observation.
   With epsilon 0.1, the robot will walk randomly. The first 16 episodes are all negative numbers. Then, the remaining are all positive numbers. That means the robot slipped at first then it learns from these experiences well.

   (c) e = 0.5,
   With epsilon 0.5, the robot has a great probability to walk randomly. From the simulation, we can see the robot usually slipped because the huge randomness.

2. World with thief
   (a) e = 0.05, knows_thief is n , how is performance of agent?

   There are lots of negative rewards and few positive rewards. The reason is that robot can not sense the thief. So there are a huge probability that the robot comes cross the thief and receive a penalty. Moreover, the slippery floor also contribute the penalty.

   (b) Set knows_thief as y. Explain performance.

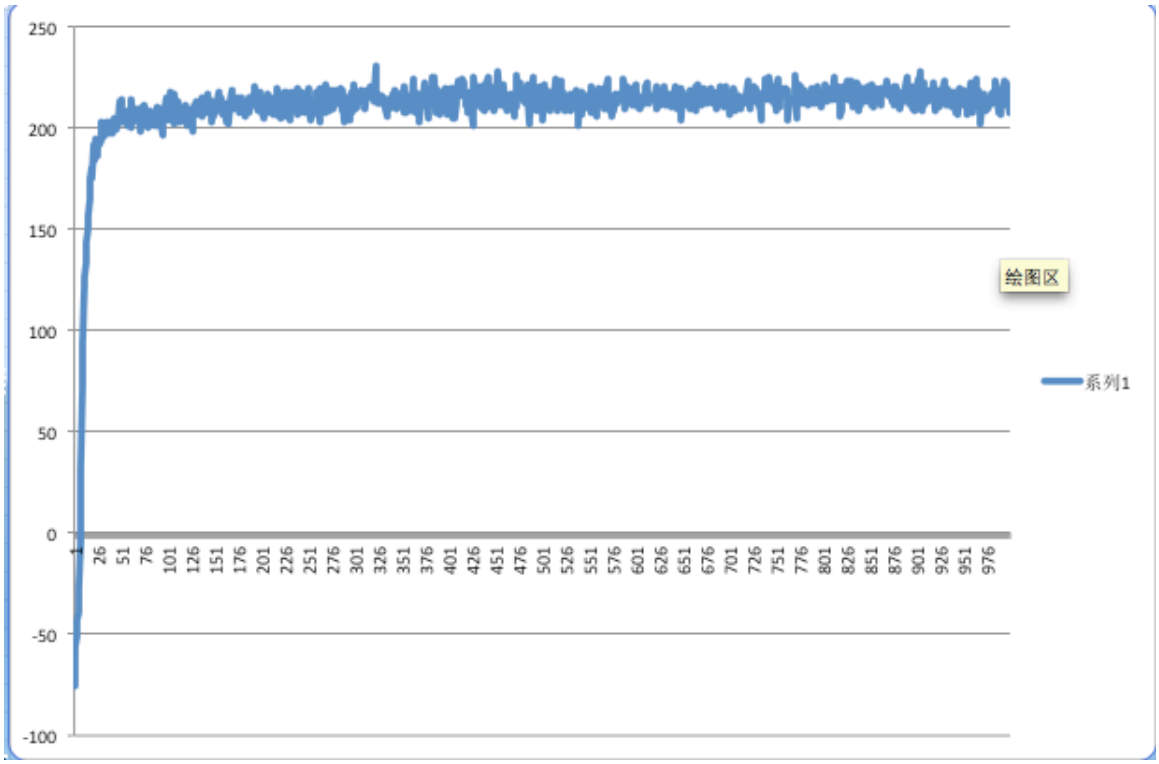
   It performed better than question 2(a) because the robot can sense the thief. After several (5 negative in my test) attempts with negative rewards, robot learned how to avoid slippery grid and thief. The remaining are all big positive number.

(c) Search for the best learning rate and ε. Describe your investigation and conclusions.

Rate = epsilon =0.1

I fixed one of variables. Change one of them, find the best reward then stop change the variable.

3.



The graph is very smooth and making sense. The robot will always deliver the package successfully with high reward. The reason is that robot can sense the thief thus it can dodge the thief. It oscillates because the thief appears randomly so the total rewards oscillate.