David Taylor - djtaylr2@illinois.edu
Homework 1 Part 2
Artificial Intelligence

1.

a.

My reward is 0.0 for all episodes, it seems that the robot isn't interested in delivering packages. When viewing the policy simulator it look like the robot just moves up. This would mean that it's picking the first direction each time and never doing anything random, and learning from it. The robot gets stuck at the top of the screen.

b.

I get a broad range of vales with no indication of an convergence, the ranges are from 20– 200 and bounce back and forth between each episode. Running the program multiple times gives me similar results each time in that there is no consistency. When viewing the policy simulator the robot will sometimes get stuck between two values. There is definitively learning it's just not necessarily smooth and gets capped around 200.

c.

When epsilon is higher I get a range of negative values, with a few very low positive numbers. There is no relationship between each episode other than they are likely to be negative values. When running the robot seems to stick to the same path even when a quicker path is close, and not really trying to avoid the slippery spots. It seems that it finds one path then continues.

2.

a.

My results are mostly negative in a range from about -20 to -5, I get the same results for each time I run the scripts. Basically, negative rewards which seems to be based on the robots obliviousness to the thief

b.

I get incredible improvements with the best reward value ranges so far. I get a mean of about 280 and a range from 270 – 300. The improvements seem to be solely on the robots awareness of the thief, allowing it to avoid it.

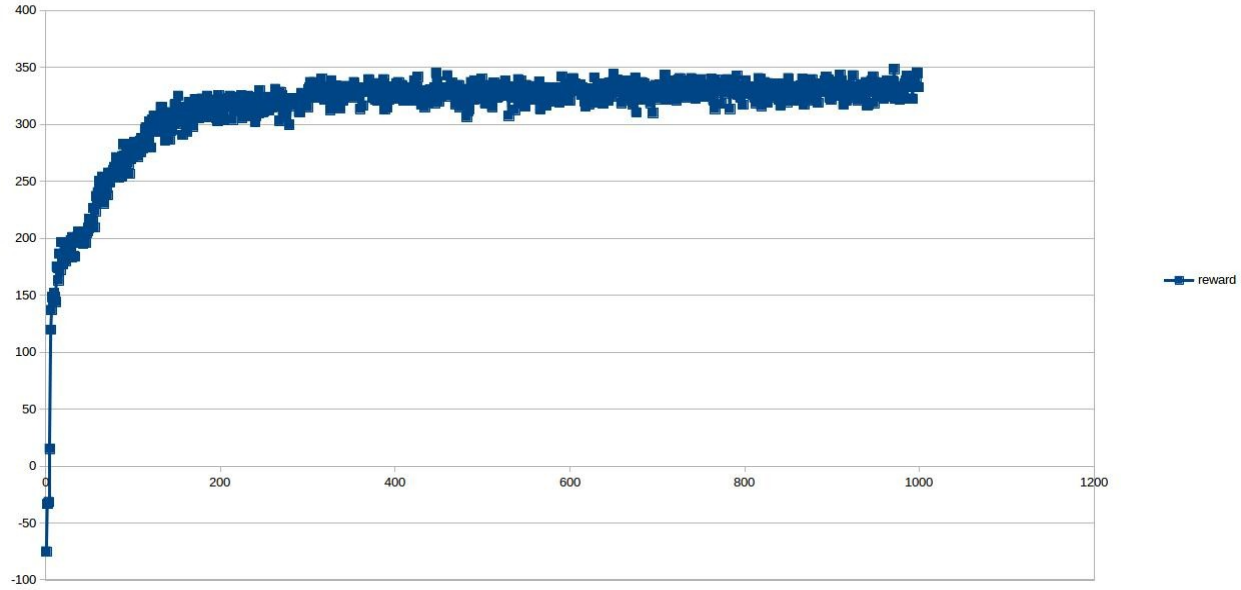
c.

My best parameters we're epsilon 0.05 and rate .01, it seemed that the closer my learning rate got to 1, the slower the increase in reward happened meaning the rate of learning was slower. Anything greater than .01 for my epsilon value resulted in a lower total reward as it seems if the value grew the reward shrunk.

David Taylor - djtaylr2@illinois.edu
Homework 1 Part 2
Artificial Intelligence

3.

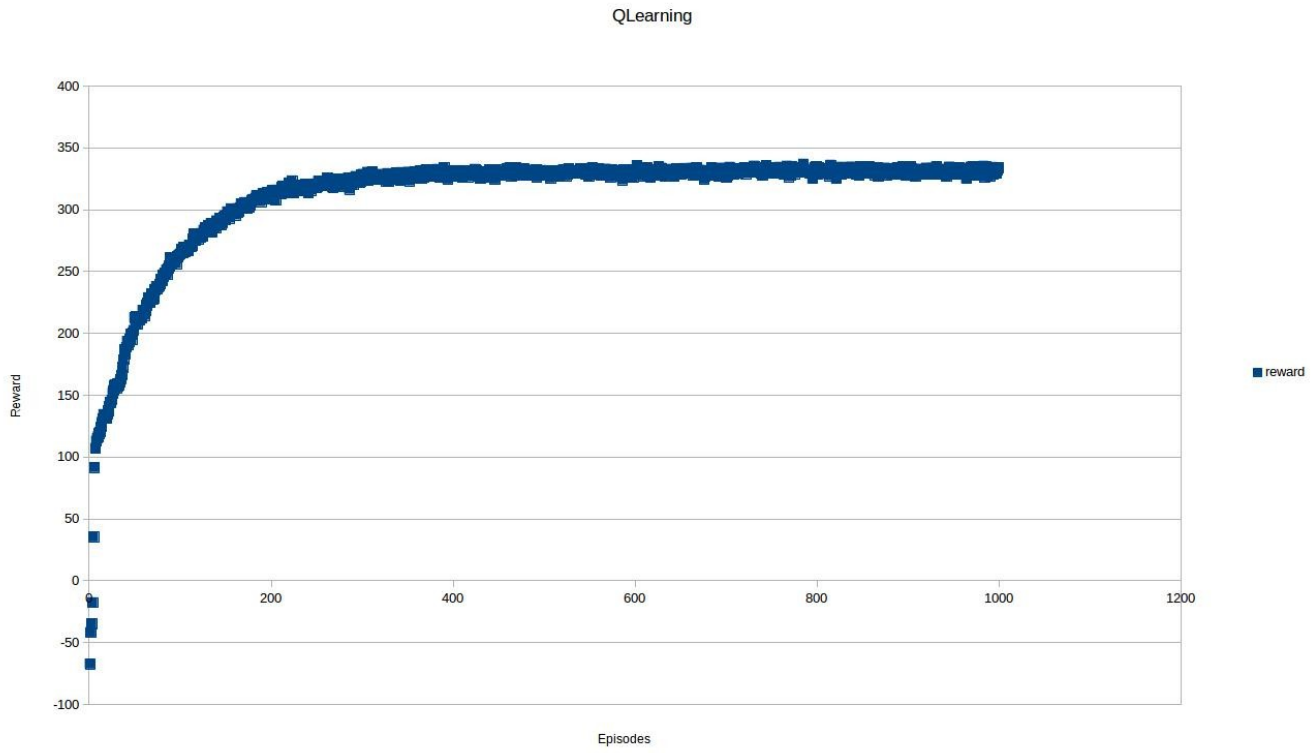
Plot of the reward of 1000 episodes with 10000 steps



There is a steep learning curve with what seems to be a convergence just under 300 around 280. At about the 180th episode the curve seems to stop changing dramatically at all and has subtle variations in change.

David Taylor - djtaylr2@illinois.edu
Homework 1 Part 2
Artificial Intelligence

Average over ten runs



This plot has a less steep learning curve than the first one but it is also seemed to have a smoothing effect. The values seems to be a little less sporadic and make a nicer looking curve with a proper convergence around 280 or just below 300. The curve seems to actually represent a mathematical function charted out, and when compared with the first graph takes a little longer to converge (around the 200th episode) but is really close. Similar, but smoother.