Rahul Bhatia
(rbhatia5)

CS 440 MP1 Report

1) a) When I run the Simulator with ε = 0.0, the output shows the each episode after the first has an overall reward of 0. This is due to the fact that with an exploration value of zero, the agent will never explore. This means after it gains enough knowledge about one decision, it never explores any other. This is what causes only the first episode to have any differing reward.

   b) When I set the agent to having an exploratory value of 0.1 and run the Simulator I see several episodes with lots of varying rewards. This, in contrast to the trial we ran before, shows that the randomness of choice between equivalent options allows us to explore more freely and ensure that we don't ignore paths.

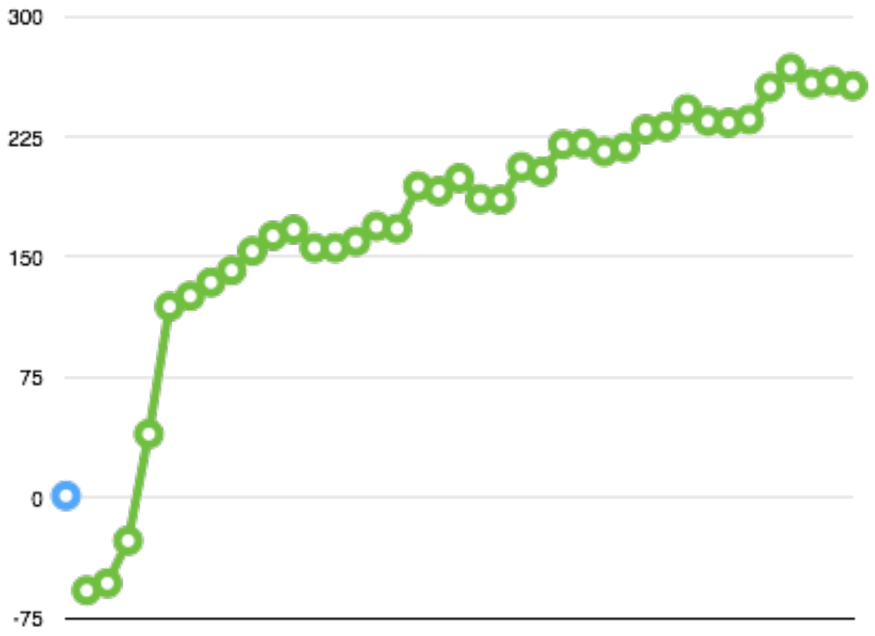
   c) When you increase the exploratory value to 0.5, then most of the rewards become negative. This seems to be a result of the additional exploration. Since the path is random half of the time, there's not as much of the learning being used in decision making. Given that, it seems natural for the overall reward to drop.

2) a) The agent performs very poorly when you set the knows_thief property to be false. This is because it cannot learn based upon the negative rewards. This causes a lot of negative rewards to build up.

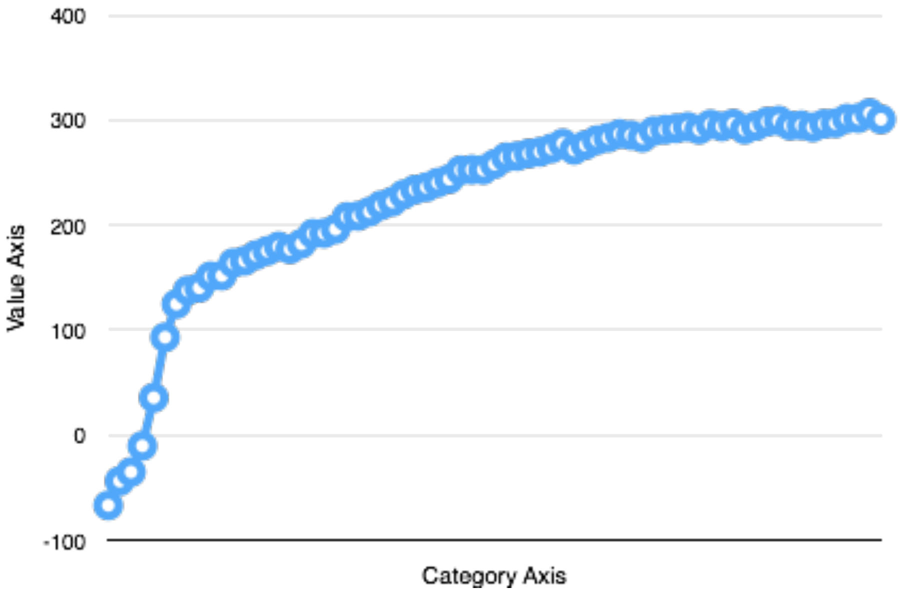
   b) When we set knows_thief to be yes, the resulting reward spread increases dramatically. Our learning function increases in value dramatically now that it takes into account the rewards.

   c) I found that using a learning rate of 0.25 and an epsilon of 0.01 the resulting reward values were in the mid 300s, much higher than most other trials.

3) QLearning over 1000 episodes with:
(x = episode, y = reward)



QLearning Averaged over 10 trials with 1000 episodes each:
(x = episode, y = reward)


Value Axis
Category Axis