Assignment 6 Report for Part A

Andrew Garwood

1. 3 disks, no noise, one goal, living reward = 0, discount = 1
   1. 4 steps are needed to get 1/3 of the states green (utility = 100)
   2. 8 steps are needed to get all states to utility = 100
   3. The policy of this set up is to almost always go right toward the goal state. This is because all choices result in the same value. This is not an optimal policy because there are cases where a state cannot go right or going right is not the best choice
2. Repeat of step 1 except with 20% noise
   1. It takes 8 steps for the start state to receive a nonzero value
   2. This policy is less arbitrary than the first. It is better because it results in reaching the goal state faster: the policies of the states tend to always lead toward the fastest route to the bottom right state.
   3. It takes 56 states for VI to converge with this setup
   4. The policy did not change. It did not change because we had already found the optimal policy. As the VI converged, it kept this policy because it was the best.
3. Repeat above set up with 20% noise, 2 goals, discount = 0.5
   1. The policy indicates that the agent will always go for the 10.0 reward IF the agent starts at the default starting point. The lower left third (except for two states) always goes for the 10.0 reward while the top and bottom right third (with the two states excluded from before) always go for the 100.0 reward. The start state has a value of 0.82. It had 23 iterations to get to convergence
   2. Converged after 46 iterations. The policy indicates that the agent will always go for the 100.0 reward except if starting on 10.0 state. Because the discount factor results in less depreciation than a, the agent will value the 100.0 reward a lot more. The starting state has value 36.9
4. 10 action simulation 10 times
   1. [off: IIII I | runs: IIII IIII II ] It went off the plan 5 times out of 10 runs
   2. It reached the goals state 8 times. (+- 1 if I miscounted)
   3. The first time it did not make it, it was 1 step away. The next time it did not make it, it was 6 steps away. The 6 seems rather large, however I believe the odds were just not in the agent’s favor that run.
   4. The agent tends to never visit the top third of the whole triangle. It tends to visit states within the bottom two thirds, trying to stay on path (I don’t think it hit the set of 3 states with the 10.0 reward.
5. Reflections
   1. I believe that it is very important for the values of the states to converge because that is the only way to guarantee an optimal path. When one reaches the start state initially, it is not always the most optimal policy.
   2. In order to produce the most accurate, optimal policy, it would be very important that all states be visited a lot. Perhaps the agent would reach a point that the values of states were not changing a lot- perhaps then it can stop visiting those states. However, the agent must visit a lot of states to get the accurate, optimal policy.