

Assignment 6 Homework Report

1.
 - a. 2
 - b. 8
 - c. The resulting policy in some states given by the case in Question 1 is illegal and they represent the highest expected value if they were to follow along the path of relatively high values. The special case in the parameters set in Question 1 shows that all states reach a value of 100 and they all have the ability to reach the end state with a reward of 100. Since it doesn't matter which state you are in and the cost of living and decay factor doesn't affect the reward, being in an illegal state doesn't matter as long as you eventually converge onto the exit state and not spend too long exploring unnecessary states.
2.
 - a. 1
 - b. Yes, all states point in the general direction of a high-value end state.
 - c. ~ 19 Value Iteration Steps before all states converge at a value of 100.
 - d. The policies have changed from pointing East to other states that crudely follow the golden state path.
3.
 - a. The start state has a value of 0.82 and it seems that after 4 iterations, it has already found a policy to reach the new exit. The policies aren't lined up with the golden path (optimal solution). The policies are mainly lined up with the reward = 10 exit probably because of cost of living and decay expense.
 - b. The policy now follows the golden path (optimal solution) after reaching convergence with a start state value of 36.9.
4.
 - a. In all of the simulations/runs, the agent managed to deviate from the plan 100% of the time. This is probably because even though the noise is 20%, the number of states it has to traverse compounds the percentage several times until it eventually fails once. Since the golden path is length 7, even 1 unplanned change of states can cause the agent to unsuccessfully reach the endstate as it takes 1 action to make a mistake and another to recover.
 - b. Out of 10 simulations, the agent arrived at the goal state in only 6 simulations.
 - c. In the simulations that the agent didn't make it to the goal state, the average states it was away from the goal was around two states.
 - d. The agent barely or never traverses the states in the upper half of the triangular region.
- 5.

- a. It is important that the values have converged. This is because the general prediction of traveling between states to eventually reach the goal state has been adequately “thought” out and have been calculated through several revolutions to determine the optimal path to reach the goal state. Converging on a set of values and executing the policy that follows the actions set by the convergence would allow the agent to get to a goal state with the max reward set by the predetermined parameters of the simulation.
- b. If the accuracy of values depends on several traversals of states, then it is of paramount importance to revisit states several times to generate an accurate policy to follow to reach a high reward end state.