Assignment 6 Report for Part A

Tianyi Zhang

Q1
1a. 4 iterations

1b. 8 iterations

1c. No. Most of the actions are not pointing to states close to the goal state's directions.

Q2
2a. 8 iterations

2b. Yes. This policy follows a route that leads to the goal state.

2c. 56 iterations

2d. The policy has changed into a bad policy. The difference are relatively large between Q values in low iteration counts, which leads to a clear policy. When iterations count reaches the convergence point, Q values are all the same. The policy is now choosen with random choice.

Q3
3a. The system converges after 23 iterations. The starting state has the value of 0.82. The policy indicates a route that leads to two goal state. States tend to be more likely to move to the goal that is closer than the other one.

3b. The starting state has value of 36.9. Now the states does not all move to the closer goal states. More states try to follow a long route to reach the goal with larger reward. The low discount value passes the reward back to the statess that are far from the goal state.

Q4
4a. 5 simulations

4b. 9 simulations

4c. 1 step away

4d. The agent never visit the top side of the state graph.

Q5
5a. No. As result shown in question 2d, after convergence the policy are arbitrary at states that have the same Q value for all actions. It is very importance to have strong difference among Q values in a state to form a good policy.

5b. If there is no noise, for the states close to the golden policy route, it is importance for the agent to obtain accurate values. For the states that are nearly never visited by the agent, the re-visit is not necessary. If noise occurs, the agent might misled by the noise and get lost in the states that the agent does not visit frequently.