**Assignment 1**

1. (a)

Thus, we have the value function for :

(b)

No, when , since the optimal policy is this case is always taking action , the corresponding optimal value of any state is dependent on discount factor . However, when , the value for all states except goal state is 0, so no matter which policy we choose, the value stays the same, which means the optimal policy going-right is not unique.

(c)

Thus, we have the value function for :

When adding a constant , the optimal policy is unchanged, since the advantage reward of adding the constant for each state is a constant:

(d)

Thus, we have the value function for :

If , the optimal policy is unchanged, since the reward model is just scaling by a constant factor(positive); if , the value for any policy is a constant of 0, thus the optimal policy is not unique; if , the value of previous optimal policy in the new reward model is now smallest, which is definitely not the optimal policy in our new reward model.

2. (a)

(b)

The optimal action is .

(c)

Since is the number of times the optimal policy converges to taking action , the value function for state is:

The state-action value for state action is:

The state-action value for state action is:

When , is always greater than or equal to :

3. (a)

(b)

After solving the linear equation, we have:

(c)

(d)

Since , we can assume the approximation function for each action is specified as , . Thus we have the tight bound

4. (c)

Stochastic policy will increase the number of iterations required, except in minor cases, the number of iterations remains unchanged, since the agent explores the environment at the beginning, which may increase the number of times the agent reaches the reward. The optimal policy may not be guaranteed to converge, because during exploration, the value iteration or policy iteration may terminate when the difference between current value function and previous value function is smaller than some tolerate constant .