

# CIS 419/519: Homework 6

Jiatong Sun

04/11/2020

Although the solutions are entirely my own, I consulted with the following people and sources while working on this homework:

## 1 Reinforcement Learning I

The reward function is fine because for a situation to have a robot run a maze can break down naturally into episodes and thus, simply using the sum of rewards is reasonable.

However, if the map is large and the episode length is not long enough, the agent will probably fail to discover the exit for many episodes. After running the agent for a while, the agent may never know there can be a total reward better than 0. Besides, since every policy it has already tried gives the same total reward 0, the agent will have no clue how to choose its policy for the following episode. Therefore, the policy is not improved.

There are two easy ways to solve the problem: increase the episode length and/or give non-goal state a value -1. This means the states that have been visited a lot will get worse and worse values so the agent will want to avoid that state. In this way, the agent can eventually reach the goal.

## 2 Reinforcement Learning II

## 3 Random policy for the MountainCar gym environment