'''

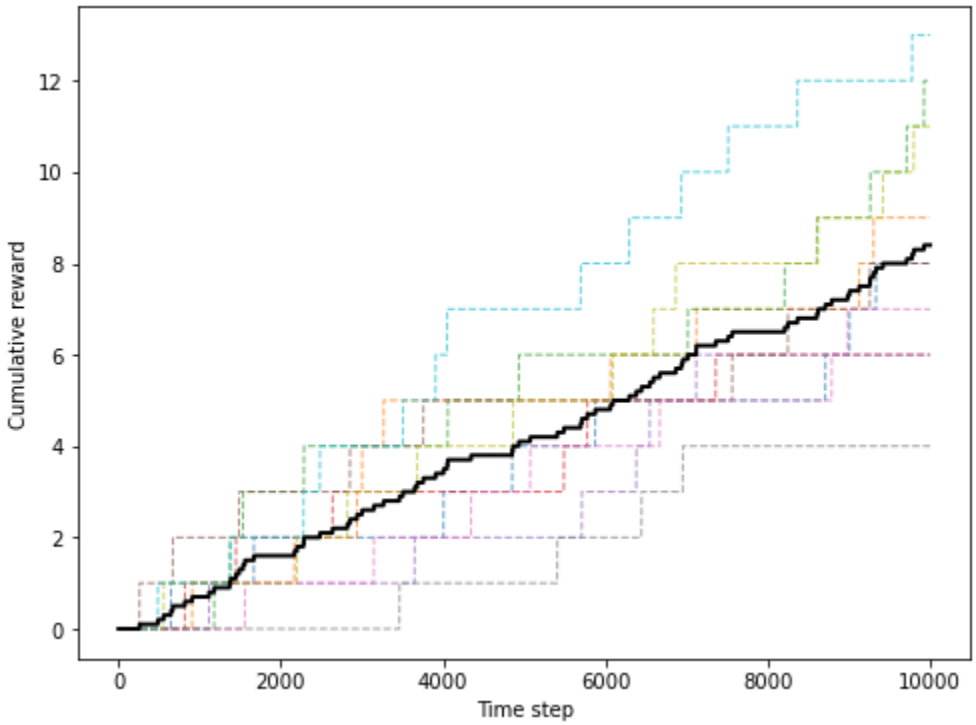
CS 5180 Fall 2022

Exercise 0: An Invitation to Reinforcement Learning

Hongyan Yang

'''

P3 Plot:

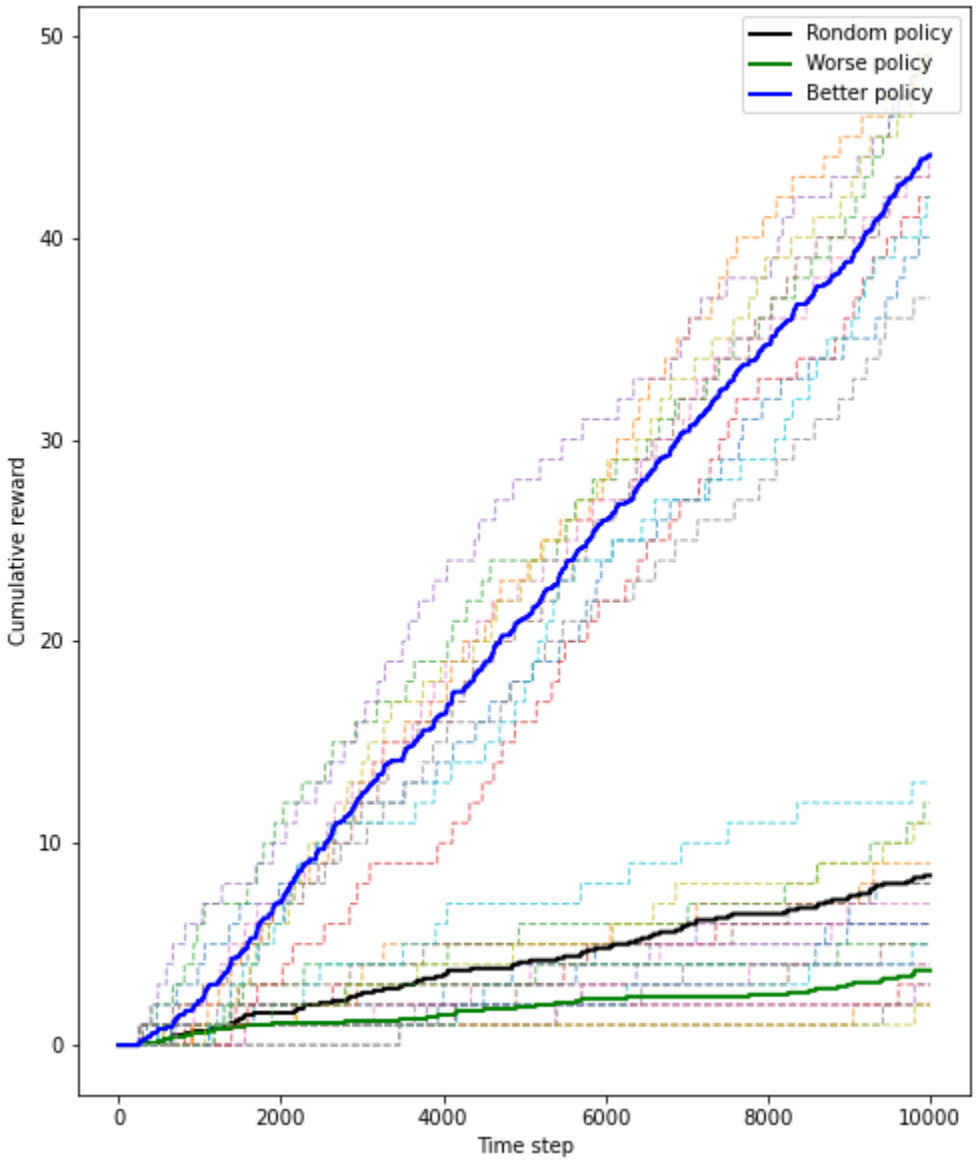


P3 Written:

The random policy performs much worse when compared to the manual policy. Reasons for the difference in performance:

1. The agent follows a stochastic policy to map the state and its action with each direction uniformly at random, making its action inefficient to reach the goal.
2. The agent is affected by a stochastic environment without knowing the occasional noise while human's input can correct the error immediately.
3. The random policy doesn't possess a value function specifies what is good in the long run while human's input applies a value function naturally to pass the walls and reach the goal.
4. The random policy lacks a trial-and-error process and cannot learn from its experience. It also lacks a model of the environment to help it pass the walls.

P4 Plot:



P4 Written:

1. Strategy better policy uses, and why:
   The better policy uses a reinforcement learning method to learn “forbidden acts” that will hit the wall, and it applies a policy to optimize its move towards the goal after the first time it achieved it.
   It’s better because it can learn from its previous mistakes (hit the wall) and avoid such acts in the future to reduce redundant and unnecessary acts and enhance overall performance.
   Also, after the agent meets the goal at the first time. It records the goal’s location and optimize its future acts to move towards the goal. Thus, it can reach the goal more efficiently.

2. Strategy worse policy uses, and why:

The worse policy applies a policy to be biased to move away from the goal after the first time it achieved it.
It's worse because after the agent meets the goal at the first time. It records the goal's location and weaken its future acts by tend to move away from the goal based on their relative locations. Thus, it reaches the goal less efficiently.