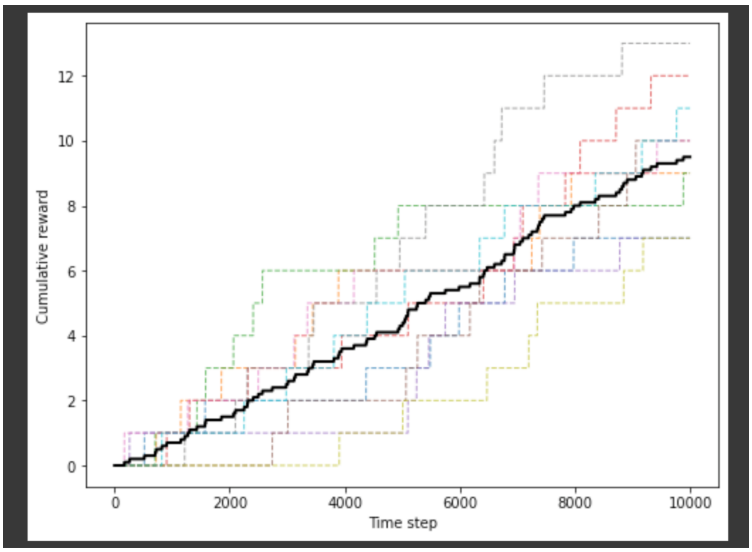
Exercise 0: An Invitation to Reinforcement Learning

Submission by: Aadhar Bansal

Question 3:

The Random policy differs from the Manual as in the Random policy the code randomly selects the Actions from the four probable actions [“UP”, “DOWN”, “LEFT”, “RIGHT”], it does not consider state into the picture. In contrast, the Manual Policy depends on the User Intelligence, If we consider the User is sophisticated and Entering Action based on the form. It will give higher accuracy and higher rewards.

In Random Policy, the Code picks the action randomly from the four options, so the chances are that many times the rewards chances would be missed because of that.



Plot for the Random Policy

Question 4:

The **Worse Policy** which I employed compared to the Random policy is that as the Goal is in the UP-Right corner of the Four Rooms Design, It chooses the random action out of two options that are [“UP”, “RIGHT”] irrespective of the state, the problem with this strategy is this will miss the reward chances much more as by this policy the state would more often go to the extremes of Row = 10 and Column = 10, and the state again would change to the reset. Overall, It will give a lesser cumulative reward compared to the Random policy.

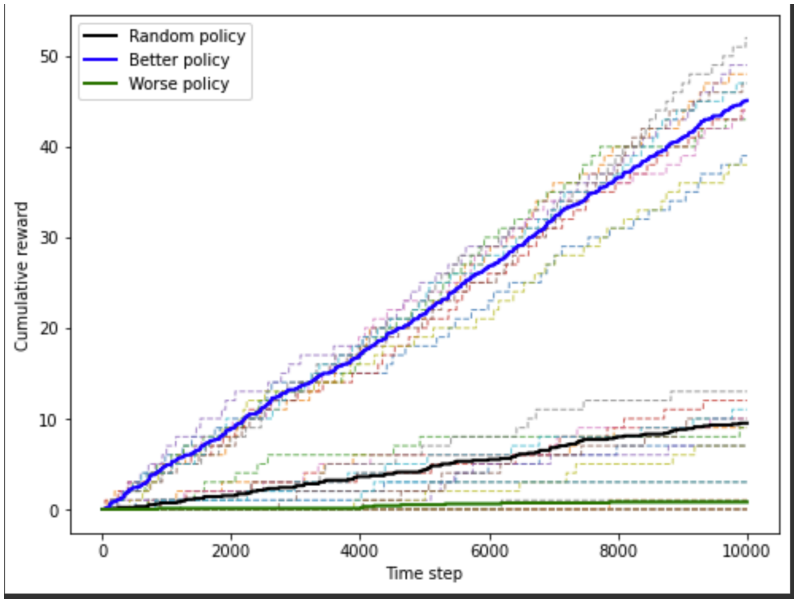
For the **Better Policy**, I have employed the policy which chooses the Action based on the State that is:

If the objecting State is the first Row, then it chooses action from the [“UP”, “RIGHT”, “LEFT”], similarly if in the last row chooses from the [“DOWN”, ”LEFT”, ”RIGHT”],

Same for the Column, if in the first column chooses from the [“UP”, “RIGHT”, “DOWN”], and the last column it chooses from [“UP”, “DOWN”, “LEFT”].

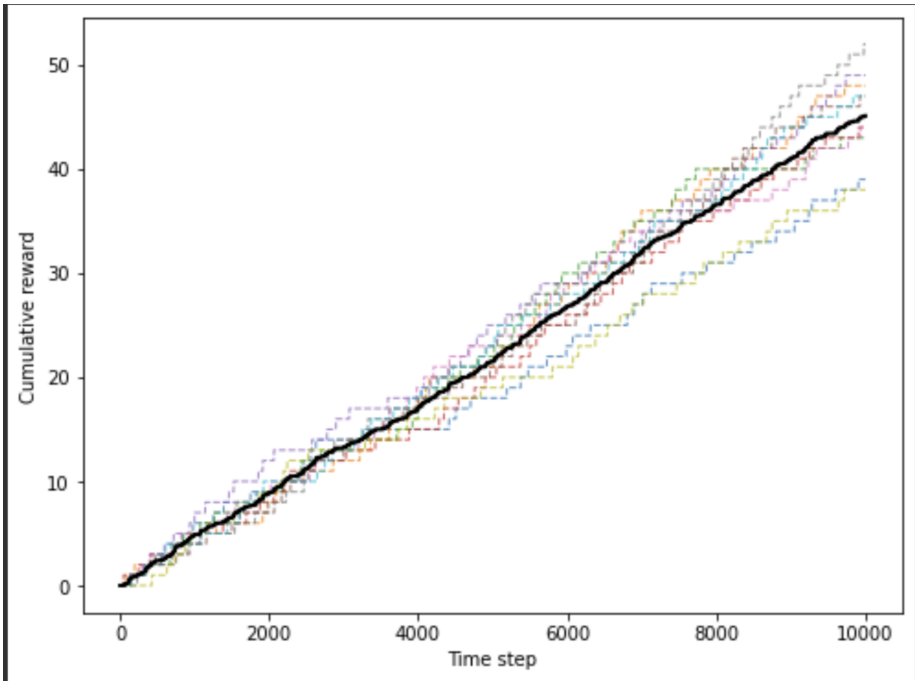
And if the object is in any corner it chooses from the two directions it can Action, like State[0, 0], can only choose from [“UP”, ”RIGHT”].

I have also given a higher probability to the UP and RIGHT action as the GOAL is in the UP-RIGHT corner.

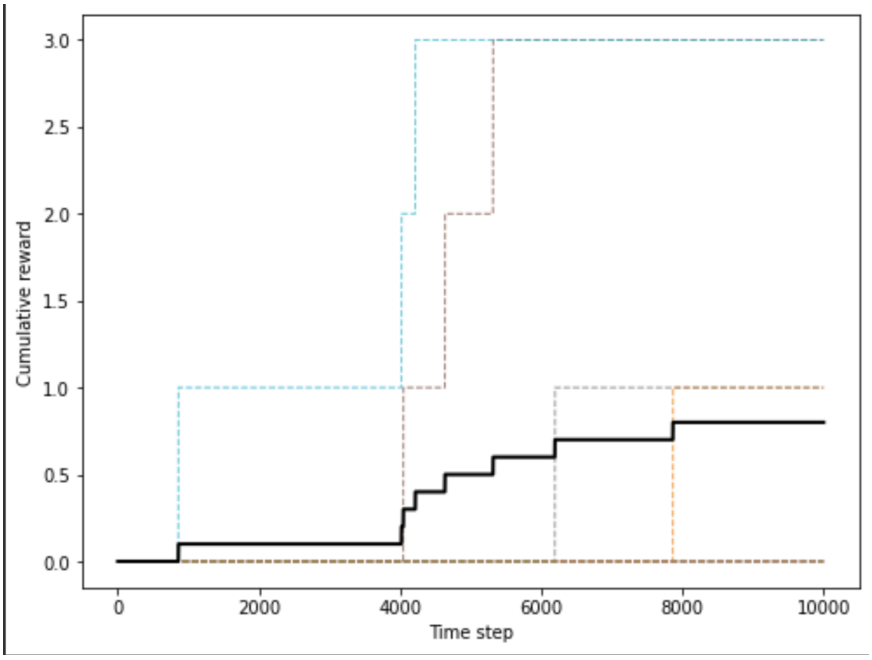


Plot comparing the Better, Random, and Worse Policies

Individual Plots for the Better and Worse Policy



Plot for the Better Policy



Plot for the Worse Policy