

HW 3 - CSE 473 Spring 2021

Due Date: Friday, May 7th, 2021 at 11:59 pm PDT

Total Points: 20 points

Problem 1: Value Iteration [10 Points]

Consider the 101x3 world below. In the start state, the agent has a choice of two deterministic actions, *Up* or *Down*, but in the other states the agent always takes the deterministic action, *Right*.

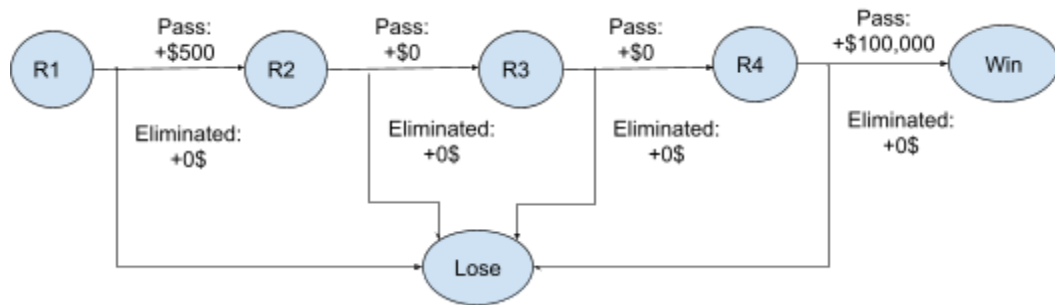
-50	1	1	1	1	..	1	1	1	1(TERMINAL)
START									
50	-1	-1	-1	-1	..	-1	-1	-1	-1(TERMINAL)

1. (6 pts) Compute the utility of each action as a function of γ .
2. (4 pts) Assuming a discounted reward function, for what values of the discount factor γ should the agent choose Up as the initial action?

Problem 2. MDP [10 Points]

You are controlling a robot walking across a narrow bridge. The goal is to walk straight at every step to get to the reward at the end of the bridge. However, the controller is noisy, so the robot does not always follow your instructions. You control the robot in rounds, one step at a time. At each step, the robot will either step forward, or fall. The robot receives a small \$500 reward after the first successful step (R1), to entice it onto the bridge. In each subsequent step, there is no reward, except at the end of the bridge after the final round (R4), where you receive a \$100,000 reward.

The probability of the robot following your instructions, thereby passing each round, is as follows: $P(R1=pass)=4/5$, $P(R2=pass)=3/4$, $P(R3=pass)=1/2$, $P(R4=pass)=1/5$.



6.1. Model this problem as an MDP. This will require you to fully specify all the components of the MDP and the transition and reward function for all states and actions. [5 pts]

6.2. What is the value of each state when $\gamma = 1$? [5 pts]