CS/CE 352/368: Introduction to Reinforcement Learning:
Assignment #1

Dr. M. Shahid Shaikh

Due on February 16, 2025, 11.59pm

Name & ID:

Problem 1

(50 points) **Environment Setup** (may contain spoilers for Spider-man)

Mr. Ditkovitch is hoping to evict all nerds who don't pay rent in his apartment, and has one final nerd to evict: Peter Parker. Unfortunately all his previous attempts to catch the crafty web-slinger have fallen short, and he turns to you, with your knowledge of Markov Decision Processes (MDP's) to help him catch Peter Parker once and for all.
Consider the following MDP environment where the agent is Mr. Ditkovitch:

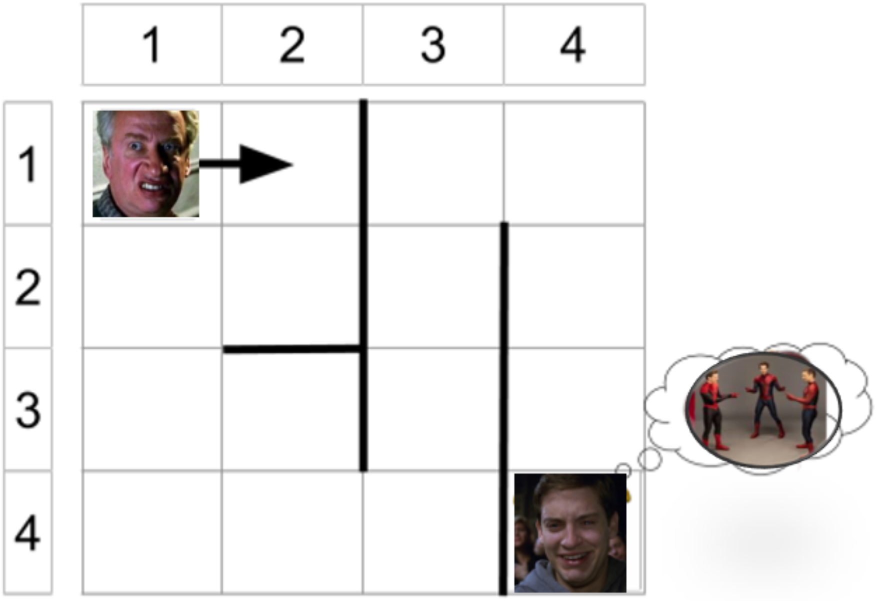

Figure 1: Friendly Spider-man Neighborhood

Here's how we will define this MDP:

- $\mathcal{S}$ **(state space)**: a set of states the agent can be in. In this case, the agent (Mr. Ditkovitch) can be in any location (row, col) and also in any orientation $\in \{N, E, S, W\}$. Therefore, state is represented by a three-tuple (row, col, dir), and S = all possible of such tuples.
- $\mathcal{A}$ **(action space)**: a set of actions that the agent can take. Here, we will have just three actions: turn right, turn left, and move forward (turning does not change row or col, just dir). So our action space is R, L, M. Note that Mr. Ditkovitch is debilitatingly short, so he cannot travel through (or over) the walls. Moving forward when facing a wall results in no change in state (but counts as an action).
- $R(s, a)$ **(reward function)**: In this scenario, Mr. Ditkovitch gets a reward of 5 by moving into the Avengers compound (the cell containing Peter Parker), and a reward of 0 otherwise.
- $p(s'|s, a)$ **(state space)**: We'll use a deterministic environment, so this will be 1 if s' is reachable from s and by taking a, and 0 if not.

(a) (5 points) What are $|\mathcal{S}|$ and $|\mathcal{A}|$ (size of state space and size of action space)?

(b) (5 points) Why is it called a "Markov" decision process?

(c) (5 points) What are the following transition probabilities?

$$p((1,1,N)|(1,1,N),M),$$
$$p((1,1,N)|(1,1,E),L),$$
$$p((2,1,S)|(1,1,S),M),$$
$$p((2,1,E)|(1,1,S),M).$$

(d) (5 points) Given a start position of (1, 1, E) and a discount factor of $\gamma = 0.5$, what is the expected discounted future reward from a = R? For a = L? (Fix $\gamma = 0.5$ for following problems).

(e) (5 points) What is the optimal action from each state, given that orientation is fixed at E? (if there are multiple options, choose any)

(f) (5 points) Mr. Ditkovitch's chief strategist (J. Jonah Jameson from The Daily Bugle) suggests that having $\gamma = 0.9$ will result in a different set of optimal policies. Is he right? Why or why not?

(g) (5 points) J. Jonah Jameson then suggests the following setup: $R(s,a) = 0$ when moving into the Avengers compound, and $R(s,a) = -1$ otherwise. Will this result in a different set of optimal policies? Why or why not?

(h) (5 points) J. Jonah Jameson now suggests the following setup: $R(s,a) = 5$ when moving into the Avengers compound, and $R(s,a) = 0$ otherwise, but with $\gamma = 1$. Could this result in a different optimal policy? Why or why not?

(i) (10 points) Surprise! Frozone from The Incredibles suddenly shows up. J. Jonah Jameson hypnotizes him and forces him to use his powers to turn the ground into ice. Now the environment is now stochastic: since the ground is now slippery, when choosing the action M, with a 0.2 chance, Mr. Ditkovitch will slip and move two squares instead of one. What is the expected future-discounted rewards from s = (2, 4, S)?

Problem 2

(20 points) You won the lottery and they will pay you one million dollars each year for 20 years (starting this year). If the interest rate is 5 percent, how much money do you need to get right away to be indifferent between this amount of money and the annuity (solve in terms of discounts)?

Problem 3

(30 points) Consider the following grid environment. Starting from any unshaded square, you can move up, down, left, or right. Actions are deterministic and always succeed (e.g. going left from state 16 goes to state 15) unless they will cause the agent to run into a wall. The thicker edges indicate walls, and attempting to move in the direction of a wall results in staying in the same square (e.g. going in any direction other than left from state 16 stays in 16). Taking any action from the green target square (no. 12) earns a reward of r_g (so $r(12,a) = r_g\ \forall a$) and ends the episode . Taking any action from the red square of death (no. 5) earns a reward of r_r (so $r(5,a) = r_r\ \forall a$) and ends the episode. Otherwise, from every other square, taking any action is associated with a reward $r_s \epsilon \{-1, 0, +1\}$ (even if the action results in the agent staying in the same square). Assume the discount factor $\gamma = 1$, $r_g = +5$, and $r_r = -5$ unless otherwise specified

Problem 3 continued on next page. . .

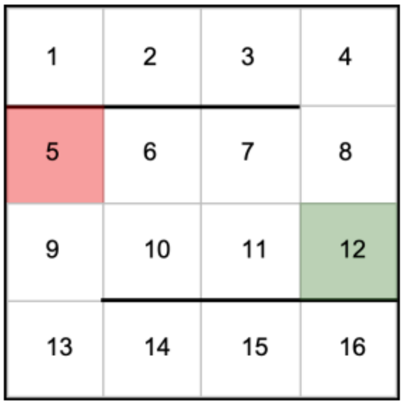

Figure 2: GridWorld

(a) (10 points) Define the value of r_s that would cause the optimal policy to return the shortest path to the green target square (no. 12). Using this r_s, find the optimal value for each square.

(b) (10 points) Consider a general MDP with rewards, and transitions. Consider a discount factor of γ. For this case assume that the horizon is infinite (so there is no termination). A policy π in this MDP induces a value function V_π (lets refer to this as V_π old). Now suppose we have a new MDP where the only difference is that all rewards have a constant c added to them. Can you come up with an expression for the new value function V_π new induced by π in this second MDP in terms of V_π old, c, and γ?

(c) (10 points) Lets go back to our gridworld from (a) with the default values for r_g, r_r, γ and with the value you specified for rs. Suppose we now derived a second gridworld by adding a constant c to all rewards (r_s, r_g, and r_r) such that $r_s = +2$. How does the optimal policy change (Just give a one or two sentence description)? What do the values of the unshaded squares become?

Submission Guidelines:

1. Submit your solutions involving content, proofs, etc. as a latex pdf.