Assignment - 2

Due date: 26th February, 2024

Answer the following questions:

1. What does $P(s_1 \mid s_0, a_2)$ mean in plain english.
2. What is the difference between a policy and an optimal policy?
3. $v_\pi(s) > v_*(s)$. True/False
4. Given below are the values of each state in a **gridworld**(with 9 states) under a particular policy,

2.0	2.5	3.0
3.5	4.0	4.5
4.0	4.5	10.0

It is given that,

State space = { (0,0), (0,1), … (2,2)}

Action space = {up, left, right, down}

Terminal state = (2,2)

Rewards:

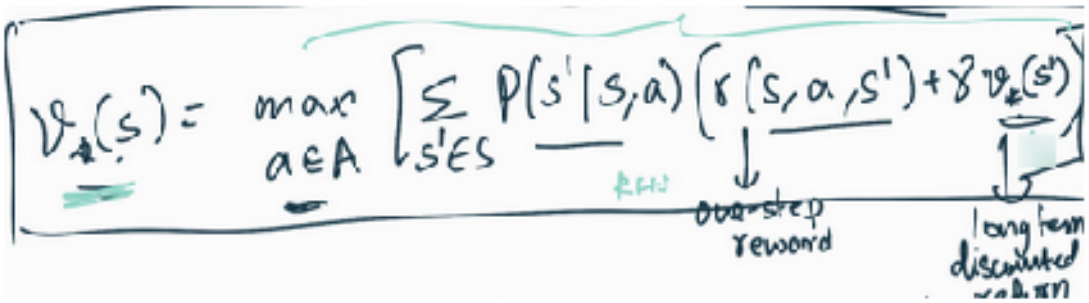
+ 10 for transition to the terminal state.
- 1 for every other transition.

From the above we can infer the one-step rewards as follows

r((0,0), right, (0,1)) = -1
r((2,1), right, (2,2)) = 10
…
…

Consider that the state transitions are deterministic given the action. Meaning for example, if the agent takes the action right, it moves to the block on the right side (if it exists) with a probability 1.

bellman's optimality equation



Assume discount factor, gamma = 1.

Use **bellman's optimality equation** and decide which action (left, right, up, down) should be taken from each state.

Hint: Choose the one step greedy action that is expected to give the maximum return. Use RHS of bellman's optimality equation for state value function. You will have to fill the arrow marks representing the best action to be taken from each state the agent can be in. For example: