2311057015

Aditya N Bhatt

AI&ML

RL Assignment -1

Answer the following questions:

1. What does $P(s_1 / s_0, a_2)$ mean in plain English.

   > It represents the probability of transitioning to state s1 given that the current state is s0 and action a2 is taken.

2. What is the difference between a policy and an optimal policy?

   > A policy in reinforcement learning is a strategy that the agent employs to determine the next action based on the current state.

   > An optimal policy is the best policy that results in the maximum expected reward over the long run for each state, compared to all other policies.

3. $v_\pi(s) > v_*(s)$. True/False

   > False
   The value of a state s under an optimal policy v*(s) is always greater than or equal to the value of s under any other policy π, i.e., v*(s) >= vπ(s)

4. Given below are the values of each state in a **gridworld**(with 9 states) under a particular policy,

   It is given that,

   State space = { (0,0), (0,1), … (3,3)}

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

   Terminal state = (3,3)

2.0	2.5	3.0
3.5	4.0	4.5
4.0	4.5	10.0

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

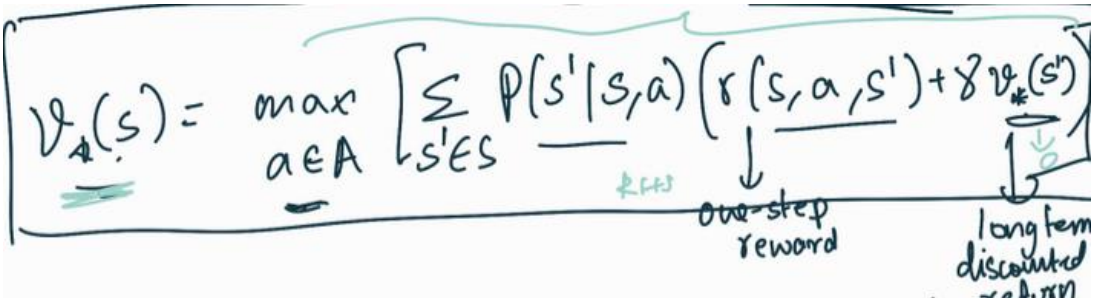
r((0,0),	*right,*	*(0,1))*	*=*	*-1*
r((2,1),	*right,*	*(3,3))*	*=*	*10*

...

...

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.

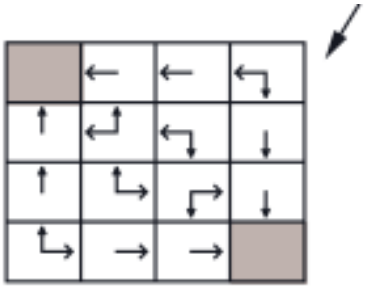
bellman's optimality equation



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:



> The Bellman's optimality equation is given by:

$$v_*(s) = \max_a [r(s, a, s') + v_*(s')]$$

Given the state values and the deterministic nature of the transitions, the optimal actions for each state would be as follows:

(0,0): Right

(0,1): Right

(0,2): Down

(1,0): Up

(1,1): Right

(1,2): Down

(2,0): Up

(2,1): Right

(2,2): Terminal State (No action needed)

(3,0): Up

(3,1): Right

(3,2): Right

(3,3): Terminal State (No action needed)