Multi-Agent Systems

Homework Assignment 5

MSc AI, VU

E.J. Pauwels

Version: December 6, 2023— Wednesday, December 13, 2022 (23h59)

NB: Unless otherwise indicated, the problems below can be solved using pen and paper.

1 Bellman equations

Rewrite the Bellman equations for v_π and q_π for the following special cases:

1. Deterministic policy π: each state is mapped to a single action (say a_s);

$$\pi(a \mid s) = \begin{cases} 1 & \text{if } a = a_s \\ 0 & \text{otherwise} \end{cases}$$

2. Combination of deterministic policy and deterministic transition $p(s' \mid s, a)$. The latter is characterized by the fact that applying an action a to a state s results each time in the same successor state s_a;

$$p(s' \mid s, a) = \begin{cases} 1 & \text{if } s' = s_a \\ 0 & \text{otherwise} \end{cases}$$

2 MDP 1

Consider an MDP with a circular state space with an odd number of nodes (i.e. the nodes are positioned along a circle and labeled 0 through n, with n even). Assume that the 0-node is an absorbing terminal state and arriving at this state yields a one-time reward of 10. In the other nodes, one can go in either one of the two circle directions, resulting in reward of 0 (unless you transition to the terminal state). Assume an equiprobable policy π (i.e. going in either direction with prob $1/2$) and no discounting (i.e. $\gamma = 1$).

1. What would be the corresponding values functions v_π and q_π?

2. What would be an optimal policy? Is this unique? What are the corresponding value functions v^* and q^*?

3. How would your answer for (2) change if each non-terminal step accrued a reward of $r_{NT} = -1$?

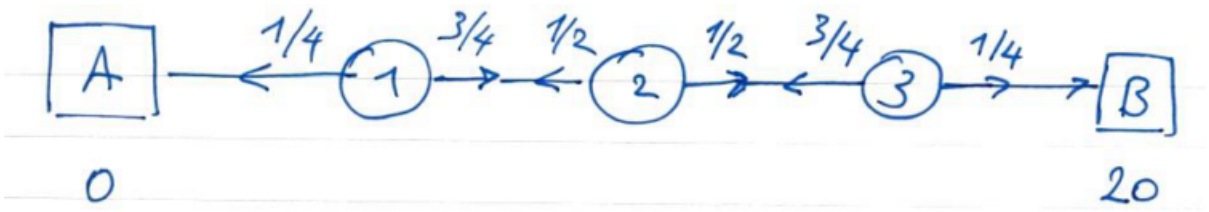
4. How would your answer for (2) change if $\gamma < 1$? (Assume $r_{NT} = 0$).

5. How would your answer for (2) change if the number of non-terminal states was odd? (Assume $r_{NT} = -1$ and $\gamma = 1$)

3 MDP 2

Consider the following MDP (see table and figure below). It has two absorbing states (A and B) that yield final rewards 0 and 20, respectively. In each non-terminal state, there are two actions (L(eft) or R(ight)) and the corresponding probabilities (determined by the policy π) are tabulated below. Non-terminal transitions incur a (negative) reward of -2. Furthermore, we assume throughout this question that there is **no discounting**, i.e. $\gamma = 1$.

$state(s)$	$action(a)$	$\pi(a \mid s)$	$reward(r)$
1	L	$1/4$	0
1	R	$3/4$	-2
2	L	$1/2$	-2
2	R	$1/2$	-2
3	L	$3/4$	-2
3	R	$1/4$	20



1. Compute the state value function $v_\pi(s)$ under the policy π for all three states $s = 1, 2, 3$.

2. Compute the state-action values $q_\pi(2, R)$ and $q_\pi(3, L)$.

3. What would be an optimal policy π^* for this MDP? Is it unique?

4 GT: Shapley value for apex game (25%)

In this game there are five players. Player 1 is the big player and all the others are small players. The big player together with one or more small players can earn value 1. If the four small players cooperate, they can also generate value 1. Hence, a coalition S has value 1, i.e. $v(S) = 1$, if

- it comprises the big player and at least one small player, i.e. $1 \in S$ and $\#S \geq 2$;
- if all small players are part of it, i.e. $2, 3, 4, 5 \in S$ (possibly in addition to 1).

Compute the Shapley value for each of the players.