Multi-Agent Systems

Authors: Chi Him Ng (2748786), Coen Nusse (2623380)

Homework Assignment 5 MScAl, VU

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 $p\left(s' \mid s, a\right)$. 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;

Both answers are in the attached image for exercise 1.

Bellman equations:

Basic $\to$ $v_\pi(s) = \sum_a \pi(a|s) \sum_{s'} p(s'|s,a)[r(s,a,s') + \gamma v_\pi(s')]$

$q_\pi(s,a) = \sum_{s'} p(s'|s,a)[r(s,a,s') + \gamma \sum_{a'} \pi(a'|s') q_\pi(s',a')]$

① Deterministic Policy $\to$ under policy $\pi: S \xrightarrow{\text{mapped}} a_S$, thus summation over action collapses

$v_\pi(s) = \sum_{s'} p(s'|s,a_s)[r(s,a_s,s') + \gamma v_\pi(s')]$

$q_\pi(s,a) = \sum_{s'} p(s'|s,a)[r(s,a,s') + \gamma q(s', a_{s'})]$

② Deterministic Policy + transition $\to$ $p(s'|s,a) = 1$ if $s' = s_a$

Using previous $\wedge$ $p(s'|s,a) = 1$

$\hookrightarrow$ $v_\pi(s) = r(s, a_s, s_{a_s}) + \gamma v_\pi(s_{a_s})$

$q_\pi(s,a) = r(s,a,s_a) + \gamma q(s_a, a_{s_a})$

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

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^* ?

   Any policy that will eventually lead to state 0.
3. How would your answer for (2) change if each non-terminal step accrued a reward of r_{NT} = -1 ?

See attached image.

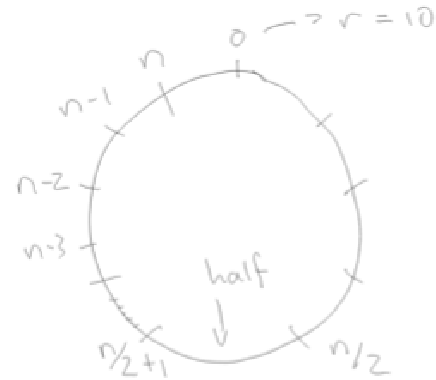
4. How would your answer for (2) change if $\gamma < 1$? (Assume $r_{NT} = 0$).

   The answer is the same as in (3).

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

   See attached image. It is almost the same, however there is a new policy with regards to n/2.

(2)



(1) π = equal → $v(s) = 10 \ \forall s$

$q(s,a) = 10 \ \forall s,a$

(2) Any policy that will lead to arrival to state 0

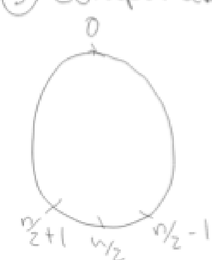
(3) Then you want to get there as soon as possible
So left if you are left and right if on right side.

if $s \leq n/2 \to s-1$ with prob. 1

if $s > n/2 \to s+1$ with prob. 1

(4) The same as (3), you still want to get 0 as quick as possible.

(5) Comparable, however there are now 3 policy's



if $s \leq n/2 - 1 \to s-1$ with prob. 1

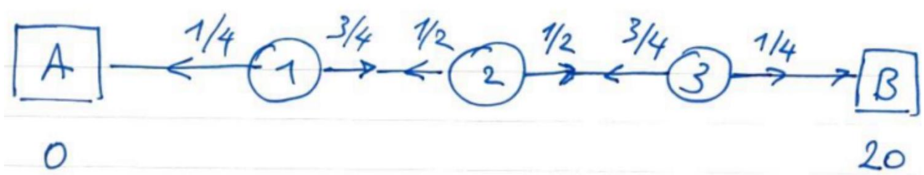
if $s \geq n/2 + 1 \to s+1$ with prob. 1

if $s = n/2 \to$ Do whatever you want

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 $s = 1, 2, 3$.

   See attached image below.

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

   See attached image below.

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

   Always go right. It is unique, since it will always go right, meaning you basically cannot return.

③ ① $V(s) = \sum_a \pi(a|s) \sum_{s'} p(s'|s,a)\left[r(s,a,s') + \gamma v(s')\right]$

$V_\pi(1) = \frac14(0+0) + \frac34(-2+V_\pi(2)) = -\frac32 + \frac34 V_\pi(2)$

$V_\pi(2) = \frac12(-2+V_\pi(1)) + \frac12(-2+V_\pi(3)) = \frac{V_\pi(1)+V_\pi(3)}{2} - 2$

$V_\pi(3) = \frac34(-2+V_\pi(2)) + \frac14(20+0) = \frac34 V_\pi(2) + 3\frac12$

↳ $V_\pi(1) + V_\pi(3) = -\frac32 + \frac34 V_\pi(2) + \frac34 V_\pi(2) + 3\frac12 = \frac32 V_\pi(2) + 2$

↳ $V_\pi(2) = \frac{-\frac32 V_\pi(2)+2}{2} - 2 = \frac34 V_\pi(2) - 1 \rightarrow V_\pi(2) = -4$

↳ $V_\pi(1) = -\frac92$, $V_\pi(3) = \frac12$

② $q_\pi(2,R) = -2 + \frac12 = -1\frac12$

$q_\pi(3,L) = -2 + -4\frac12 = -6\frac12$

③ Go right no matter the state -> Unique, since it will always go right it can never return.

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).

See attached image below.
Compute the Shapley value for each of the players.

(4) 2345 2345 $B = 1$, $S = 2, 3, 4, 5$

$S=0 \to 0$

$S=1 \to 1 \cdot 4$

$S=2 \to 1 \cdot 6$

$S=3 \to 1 \cdot 4$

$S=4 \to 1 \cdot 1$

$\varphi_B = \frac{1}{5}\left[\frac{1}{2}\cdot 4 + 6 + 4 \cdot 1\frac{1}{2} + 2\right] = 3\frac{1}{5}$

$\varphi_S = \frac{1}{5}\left[\frac{1}{2} + 6 + 4\frac{1}{2}\right] = 2\frac{1}{5}$

$\hookrightarrow \binom{2}{0} = 1$

$\binom{2}{1} = 2 \to \frac{1}{2}$

$\binom{2}{2} = 1$

$\binom{2}{3} = \frac{2}{3} \to 1\frac{1}{2}$

$\binom{2}{4} = \frac{1}{2} \to 2$

$S=0 \to 0$

$S=1 \to 0 + 1$

$S=2 \to 0 + 3 \cdot 1$

$S=3 \to 0 + 3 \cdot 1$

$S=4 \to 1 \cdot 1$