CS 188 Spring 2022 — Introduction to Artificial Intelligence — Written HW 8

Due: Wednesday, April 6, 2022 at 10:59pm (submit via Gradescope).

Policy: Can be solved in groups (acknowledge collaborators) but must be written up individually

Submission: It is recommended that your submission be a PDF that matches this template. You may also fill out this template digitally (e.g. using a tablet). **However, if you do not use this template, you will still need to write down the below four fields on the first page of your submission.**

First name	Qingjing
Last name	zhang
SID	3037581096
Collaborators	none.

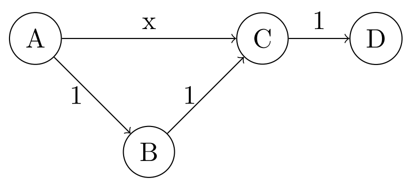
For staff use only:

Q1.	Markov Decision Process	/20
	Total	/20

Q1. [20 pts] Markov Decision Process

Throughout this homework, we use $V(s)$ to denote the value of a state. This is the same as $U(s)$ used in lecture to denote the utility of a state. "Value" and "utility" mean the same thing in a Markov decision process.

(a) [5 pts] Consider the following deterministic MDP with four states A, B, C and D:



The edges designate actions between states, the weights on those edges are the rewards, and the discount factor is $\gamma = 1$. Let k be the **first** iteration of Value Iteration at which the value function converges for some x for a particular state (i.e. $V_k(s) = V^*(s)$). Use the convention from lecture where $V_0(s)$ is the value at initialization, $V_1(s)$ is the value after one iteration, etc. For each state A, B, C, and D, list **all possible** values of k. In the case a value function for a particular state never converges, set $k = \infty$ for that state.

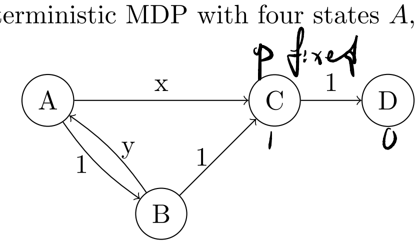
(a) State A, $k = \boxed{2, 3}$

(b) State B, $k = \boxed{2}$

(c) State C, $k = \boxed{1}$

(d) State D, $k = \boxed{0}$

(b) Now consider the following deterministic MDP with four states A, B, C and D:



The edges designate actions between states, the weights on those edges are the rewards, and the discount factor is again $\gamma = 1$. Furthermore assume that $x, y \geq 0$.

(i) [5 pts] Let k be the **first** iteration of Value Iteration for some nonnegative x and y at which the value function converges for a particular state ($V_k(s) = V^*(s)$). For each state A, B, C and D list **all** possible values of k. In case a value for a particular state never converges set $k = \infty$ for that state.

(a) State A, $k = \boxed{\infty}$

(b) State B, $k = \boxed{\infty}$

(c) State C, $k = \boxed{1}$

(d) State D, $k = \boxed{0}$

(ii) [6 pts] Suppose we perform Policy Iteration and that k is the **first** iteration for which the policy is optimal for a particular state (i.e. $\pi_k(s) = \pi^*(s)$). On top of $x, y \geq 0$ also assume that $x + y < 1$ and that tie-breaking during policy improvement is alphabetical. The initial policy is given in the table below.

State s	Policy $\pi_0(s)$
A	C
B	C
C	D
D	D

For each state A, B, C and D, find k; if the policy never converges set $k = \infty$ for that state.

(a) State A, $k =$ 1

(b) State B, $k =$ 1

(c) State C, $k =$ 0

(d) State D, $k =$ 0

Th following two questions are conceptual.

(c) [2 pts] Which of the following statements are guaranteed to be correct for any MDP? Select all that apply.

- [x] There exists a state s and some policy π such that $V^\pi(s) \leq V^*(s)$.
- [ ] There does not exist a state s such that for all policies π, $V^\pi(s) \leq V^*(s)$.
- [x] For all states s and for all policies π, $V^\pi(s) \leq V^*(s)$.
- ○ None of the above.

(d) [2 pts] Which of the following statements are guaranteed to be correct for Value Iteration? Select all that apply.

- [x] At each iteration, and for all states, the value at the next iteration is $\geq$ the value at the current iteration.
- [ ] At each iteration, and for all states, the value at the next iteration is $>$ the value at the current iteration.
- [ ] At each iteration, the value function can be lower than the earlier values for some state.
- [x] Once the value function is optimal at all states, value iteration will not change any value at any state.
- ○ None of the above.