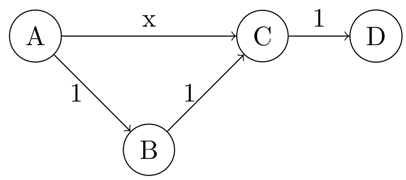
CS 188 Spring 2024 | Introduction to Artificial Intelligence | HW 7B Sol.

Solutions for HW 7B

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, $k =$ 2 and 3

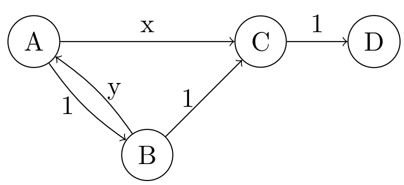
(b) State B, $k =$ 2

(c) State C, $k =$ 1

(d) State D, $k =$ 0

C will find its optimal value in one iteration. B will find its optimal value one iteration after that (2 iterations total). A will find its optimal value one iteration after C (2 iterations total) or one iteration after B (3 iterations total), depending on the value of x.

(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, $k =$ ∞

(b) State B, $k =$ ∞

(c) State C, $k =$ 1

(d) State D, $k =$ 0

A and B will never find their optimal value because they can get infinite value. C and D are the same as above.

(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, $k =$ 2

(c) State C, $k =$ 0

(d) State D, $k =$ 0

First, evaluate π_0 :

$$\begin{aligned}
&V^{\pi_0}(D) = 0 \\
&\Rightarrow V^{\pi_0}(C) = 1 + V^{\pi_0}(D) = 1 \\
&\Rightarrow V^{\pi_0}(B) = 1 + V^{\pi_0}(C) = 2 \\
&\Rightarrow V^{\pi_0}(A) = x + V^{\pi_0}(C) = x + 1
\end{aligned}$$

Now do policy improvement to obtain π_1:

$$\begin{aligned}
&\pi_1(D) = D \\
&\pi_1(C) = D \\
&\pi_1(B) = \text{argmax}_{A,C}\{A : x + y + 1, C : 2\} = C \\
&\pi_1(A) = \text{argmax}_{B,C}\{B : 3, C : x + 1\} = B
\end{aligned}$$

Now, evaluate π_1 :

$$\begin{aligned}
&V^{\pi_1}(D) = 0 \\
&\Rightarrow V^{\pi_1}(C) = 1 + V^{\pi_1}(D) = 1 \\
&\Rightarrow V^{\pi_1}(B) = 1 + V^{\pi_1}(C) = 2 \\
&\Rightarrow V^{\pi_1}(A) = 1 + V^{\pi_1}(B) = 3
\end{aligned}$$

Now run policy improvement to obtain π_2 :

$$\begin{aligned}
&\pi_2(D) = D \\
&\pi_2(C) = D \\
&\pi_2(B) = \text{argmax}_{A,C}\{A : y + 3, C : 2\} = A \\
&\pi_2(A) = \text{argmax}_{B,C}\{B : 3, C : x + 1\} = B
\end{aligned}$$

Observe that this policy is optimal, because the value $V^{\pi_2}(A) = V^{\pi_2}(B) = \infty$. The other values are trivially optimal because the agent has only one choice of action.

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.

- ■ 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, $V^{\pi}(s) \leq V^{*}(s)$.
- ■ For all states s and for all policies π, $V^{\pi}(s) \leq V^{*}(s)$.
- ○ None of the above.

By definition of the optimal policy and value function, $V^{\pi}(s) \leq V^{*}(s)$ for all states s and all policies π. All the other answers can be derived from this fact.

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

- □ 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.
- ■ Once the value function is optimal at all states, value iteration will not change any value at any state.
- ○ None of the above.

Before convergence, the values can fluctuate. Once the value function is optimal, it has converged.