CSE 150. Assignment 8 *Summer 2016*

Out: *Thu Jul 21*
Due: *Tue Jul 26*
Reading: *Sutton & Barto*, Chapters 1-4.

8.1 Two-state MDP

Consider the Markov decision process (MDP) with two states $s \in \{0, 1\}$, two actions $a \in \{\texttt{up}, \texttt{down}\}$, discount factor $\gamma = \frac{1}{2}$, and rewards and transition matrices as shown below:

s	$R(s)$
0	-1
1	2

s	s'	$P(s'\|s, a=\texttt{up})$
0	0	$\frac{3}{4}$
0	1	$\frac{1}{4}$
1	0	$\frac{1}{4}$
1	1	$\frac{3}{4}$

s	s'	$P(s'\|s, a=\texttt{down})$
0	0	$\frac{1}{2}$
0	1	$\frac{1}{2}$
1	0	$\frac{1}{2}$
1	1	$\frac{1}{2}$

(a) **Policy evaluation**

Consider the policy π that chooses the action $a = \texttt{up}$ in each state. For this policy, solve the linear system of Bellman equations (by hand) to compute the state-value function $V^\pi(s)$ for $s \in \{0, 1\}$. Your answers should complete the following table. (*Hint:* the missing entries are whole numbers.) **Show your work for full credit.**

s	$\pi(s)$	$V^\pi(s)$
0	up	
1	up	

(b) **Policy improvement**

Compute the greedy policy $\pi'(s)$ with respect to the state-value function $V^\pi(s)$ from part (a). Your answers should complete the following table. **Show your work for full credit.**

s	$\pi(s)$	$\pi'(s)$
0	up	
1	up	

8.2 Three-state MDP

Consider the Markov decision process (MDP) with three states $s \in \{1, 2, 3\}$, two actions $a \in \{\texttt{up}, \texttt{down}\}$, discount factor $\gamma = \frac{2}{3}$, and rewards and transition matrices as shown below:

s	$R(s)$
1	-15
2	30
3	-25

s	s'	$P(s'\|s, a=\texttt{up})$
1	1	$\frac{3}{4}$
1	2	$\frac{1}{4}$
1	3	0
2	1	$\frac{1}{2}$
2	2	$\frac{1}{2}$
2	3	0
3	1	0
3	2	$\frac{3}{4}$
3	3	$\frac{1}{4}$

s	s'	$P(s'\|s, a=\texttt{down})$
1	1	$\frac{1}{4}$
1	2	$\frac{3}{4}$
1	3	0
2	1	0
2	2	$\frac{1}{2}$
2	3	$\frac{1}{2}$
3	1	0
3	2	$\frac{1}{4}$
3	3	$\frac{3}{4}$

(a) **Policy evaluation**

Consider the policy π that chooses the action shown in each state. For this policy, solve the linear system of Bellman equations (by hand) to compute the state-value function $V^\pi(s)$ for $s \in \{1, 2, 3\}$. Your answers should complete the following table. (*Hint:* the missing entries are whole numbers.) **Show your work for full credit.**

s	$\pi(s)$	$V^\pi(s)$
1	up	
2	up	
3	down	

(b) **Policy improvement**

Compute the greedy policy $\pi'(s)$ with respect to the state-value function $V^\pi(s)$ from part (a). Your answers should complete the following table. **Show your work for full credit.**

s	$\pi(s)$	$\pi'(s)$
1	up	
2	up	
3	down	