CS221 Fall 2018 Homework [blackjack]

SUNet ID: prabhjot
Name: Prabhjot Singh Rai

By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

Problem 1

(a) The equation V_{opt} is given by:

$$V_{opt} = \begin{cases} 0 & \text{if } endState \\ max_{\text{a } \epsilon \text{ actions}} Q_{opt}(s,a) & \text{otherwise} \end{cases}$$

where,

$$Q_{opt}(s,a) = \sum_{s'} T(s,a,s')[\text{Reward}(s,a,s') + \gamma V_{opt}(s')]$$

Iteration 1
As per the question, V is assigned values of 0. Therefore,

$$V_{opt} = \{-2:0, -1:0, 0:0, 1:0, 2:0\}$$

Iteration 2

(a) State 0

i. Action +1

$$\begin{aligned} Q_{opt} &= 0.3 * [-5+0] + 0.7 * [-5+0] \\ &= -5 \end{aligned}$$

ii. Action −1

$$\begin{aligned} Q_{opt} &= 0.8[-5+0] + 0.2 * [-5+0] \\ &= -5 \end{aligned}$$

Therefore, $V^1_{opt}(0) = -5$

(b) State 1

i. Action +1

$$\begin{aligned} Q_{opt} &= 0.3 * [100 + 0] + 0.7 * [-5 + 0] \\ &= 26.5 \end{aligned}$$

ii. Action −1

$$\begin{aligned} Q_{opt} &= 0.8 * [-5 + 0] + 0.2 * [100 + 0] \\ &= 16 \end{aligned}$$

Therefore, $V^1_{opt}(1) = 26.5$

(c) State −1

i. Action +1

$$\begin{aligned} Q_{opt} &= 0.3 * [-5 + 0] + 0.7 * [20 + 0] \\ &= 12.5 \end{aligned}$$

ii. Action −1

$$\begin{aligned} Q_{opt} &= 0.8 * [20] + 0.2 * [-5] \\ &= 15 \end{aligned}$$

Therefore, $V^1_{opt}(-1) = 15$

(d) State 2, since it's an end state $V^1_{opt}(2) = 0$

(e) State −2, since it's an end state $V^1_{opt}(-2) = 0$

Therefore,

$$V^1_{opt} = \{-2 : 0, -1 : 15, 0 : -5, 1 : 26.5, 2 : 0\}$$

Iteration 3

(a) State 0

i. Action +1

$$\begin{aligned} Q_{opt} &= 0.3 * [-5 + 26.5] + 0.7 * [-5 + 15] \\ &= 6.45 + 7 \\ &= 13.45 \end{aligned}$$

ii. Action −1

$$\begin{aligned} Q_{opt} &= 0.8 * [-5 + 15] + 0.2 * [-5 + 26.5] \\ &= 8 + 4.3 \\ &= 12.3 \end{aligned}$$

Therefore, $V^2_{opt}(0) = 13.45$

(b) State 1

i. Action +1

$$\begin{aligned} Q_{opt} &= 0.3 * [100 + 0] + 0.7 * [-5 + -5] \\ &= 30 - 7 \\ &= 23 \end{aligned}$$

ii. Action −1

$$\begin{aligned} Q_{opt} &= 0.8 * [-5 + -5] + 0.2 * [100 + 0] \\ &= -8 + 20 \\ &= 12 \end{aligned}$$

Therefore, $V^2_{opt}(1) = 23$

(c) State −1

i. Action +1

$$\begin{aligned} Q_{opt} &= 0.3 * [-5 + -5] + 0.7 * [20] \\ &= -3 + 14 \\ &= 9 \end{aligned}$$

ii. Action −1

$$\begin{aligned} Q_{opt} &= 0.8 * [20] + 0.2 * [-5 + -5] \\ &= 16 - 2 \\ &= 14 \end{aligned}$$

Therefore, $V^2_{opt}(-1) = 14$

Therefore,

$$V^2_{opt} = \{-2 : 0, -1 : 14, 0 : 13.45, 1 : 23, 2 : 0\}$$

(b) From the solution in $1a$, the different [Action, $Q_{opt}(s, a)$] after iteration 2 for

(a) State 0 = $[+1, 13.45][-1, 12.3]$, therefore, $\pi_{opt}(0) = +1$

(b) State 1 = $[+1, 23][-1, 12]$, therefore, $\pi_{opt}(1) = +1$

(c) State −1 = $[+1, 9][-1, 14]$, therefore, $\pi_{opt}(-1) = -1$

Therefore,

$$\pi_{opt}(s) = \{-1 : -1, 0 : +1, 1 : +1\}$$

Problem 2

(a) (your solution)

(b) (your solution)