Miray Yüce, Mostafa Hussein, Arlette Michelle Perez Espinoza

Exercise 3 solutions

(a), calculate *V*3(*i*) for all states *i* based on the illustrated episodes 1 to 3 (right part of Figure 1).

Using Monte-Carlo policy evaluation

$$V(S_t) \leftarrow V(S_t) + \alpha(G_t - V(S_t))$$

$V_0(i) = 0$ for all i

Episode 1:

$\alpha = \frac{1}{t} = \frac{1}{1} = 1$

$\gamma = 1$

$$V(S_{2,1}) \leftarrow V(S_{2,1}) + \alpha(G_t - V(S_{2,1}))$$
$$G_t = R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{T-t-1} R_T = -5 + 10 = 5$$
$$V(S_{2,1}) \leftarrow 0 + 1 * (5 - 0)$$
$$V(S_{2,1}) \leftarrow 5$$

$$V(S_{2,2}) \leftarrow V(S_{2,2}) + \alpha(G_t - V(S_{2,2}))$$
$$G_t = -4 + 10 = 6$$
$$V(S_{2,2}) \leftarrow 0 + 1 * (6 - 0)$$
$$V(S_{2,2}) \leftarrow 6$$

$$V(S_{2,3}) \leftarrow V(S_{2,3}) + \alpha(G_t - V(S_{2,3}))$$
$$G_t = -3 + 10 = 7$$
$$V(S_{2,3}) \leftarrow 0 + 1 * (7 - 0)$$
$$V(S_{2,3}) \leftarrow 7$$

$$V(S_{1,3}) \leftarrow V(S_{1,3}) + \alpha(G_t - V(S_{1,3}))$$
$$G_t = -2 + 10 = 8$$
$$V(S_{1,3}) \leftarrow 0 + 1 * (8 - 0)$$
$$V(S_{1,3}) \leftarrow 8$$

$$V(S_{1,4}) \leftarrow V(S_{1,4}) + \alpha(G_t - V(S_{1,4}))$$
$$G_t = -1 + 10 = 9$$
$$V(S_{1,4}) \leftarrow 0 + 1 * (9 - 0)$$
$$V(S_{1,4}) \leftarrow 9$$

$$V(S_{1,5}) \leftarrow V(S_{1,5}) + \alpha(G_t - V(S_{1,5}))$$
$$G_t = 10$$
$$V(S_{1,5}) \leftarrow 0 + 1 * (10 - 0)$$
$$V(S_{1,5}) \leftarrow 10$$

$$V(S_{2,5}) \leftarrow V(S_{2,5}) + \alpha(G_t - V(S_{2,5}))$$
$$G_t = 10$$
$$V(S_{2,5}) \leftarrow 0 + 1 * (10 - 0)$$
$$V(S_{2,5}) \leftarrow 10$$

Episode 2:

$\alpha = \frac{1}{t} = \frac{1}{2}$

$\gamma = 1$

$$V(S_{2,1}) \leftarrow V(S_{2,1}) + \alpha(G_t - V(S_{2,1}))$$
$$G_t = R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{T-t-1} R_T = -2 - 100 = -102$$
$$V(S_{2,1}) \leftarrow 5 + \frac{1}{2} * (-102 - 5)$$
$$V(S_{2,1}) \leftarrow \frac{10}{2} - \frac{107}{2} = -\frac{97}{2} = -48.5$$

$$V(S_{2,2}) \leftarrow V(S_{2,2}) + \alpha(G_t - V(S_{2,2}))$$
$$G_t = -1 - 100 = -101$$
$$V(S_{2,2}) \leftarrow 6 + \frac{1}{2} * (-101 - 6)$$
$$V(S_{2,2}) \leftarrow \frac{12}{2} - \frac{107}{2} = -\frac{95}{2} = -47.5$$

$$V(S_{2,3}) \leftarrow V(S_{2,3}) + \alpha(G_t - V(S_{2,3}))$$
$$G_t = -100$$
$$V(S_{2,3}) \leftarrow 7 + \frac{1}{2} * (-100 - 7)$$
$$V(S_{2,3}) \leftarrow \frac{14}{2} - \frac{107}{2} = -\frac{93}{2} = -46.5$$

$$V(S_{3,3}) \leftarrow V(S_{3,3}) + \alpha(G_t - V(S_{3,3}))$$
$$G_t = -100$$
$$V(S_{3,3}) \leftarrow 0 + \frac{1}{2} * (-100 - 0)$$
$$V(S_{3,3}) \leftarrow -50$$

Episode 3:

$\alpha = \frac{1}{t} = \frac{1}{3}$

$\gamma = 1$

$$V(S_{2,1}) \leftarrow V(S_{2,1}) + \alpha(G_t - V(S_{2,1}))$$
$$G_t = R_{t+1} + \gamma R_{t+2} + \cdots + \gamma^{T-t-1} R_T = -5 + 10 = 5$$
$$V(S_{2,1}) \leftarrow -\frac{97}{2} + \frac{1}{3} * (5 + \frac{97}{2})$$
$$V(S_{2,1}) \leftarrow -\frac{582}{6} + \frac{107}{6} = -\frac{475}{6} = -79.16$$

$$V(S_{2,2}) \leftarrow V(S_{2,2}) + \alpha(G_t - V(S_{2,2}))$$
$$G_t = -4 + 10 = 6$$
$$V(S_{2,2}) \leftarrow -\frac{95}{2} + \frac{1}{3} * (6 + \frac{95}{2})$$
$$V(S_{2,2}) \leftarrow -\frac{570}{6} + \frac{107}{6} = -\frac{463}{6} = -77.16$$

$$V(S_{2,3}) \leftarrow V(S_{2,3}) + \alpha(G_t - V(S_{2,3}))$$
$$G_t = -3 + 10 = 7$$
$$V(S_{2,3}) \leftarrow -\frac{93}{2} + \frac{1}{3} * (7 + \frac{93}{2})$$
$$V(S_{2,3}) \leftarrow -\frac{558}{6} + \frac{107}{6} = -\frac{451}{6} = -75.16$$

$$V(S_{2,4}) \leftarrow V(S_{2,4}) + \alpha(G_t - V(S_{2,4}))$$

$$G_t = -2 + 10 = 8$$
$$V(S_{2,4}) \leftarrow 0 + \frac{1}{3} * (8 - 0)$$
$$V(S_{2,4}) \leftarrow \frac{8}{3} = 2.66$$

$$V(S_{1,4}) \leftarrow V(S_{1,4}) + \alpha(G_t - V(S_{1,4}))$$
$$G_t = -1 + 10 = 9$$
$$V(S_{1,4}) \leftarrow 9 + \frac{1}{3} * (9 - 9)$$
$$V(S_{1,4}) \leftarrow 9$$

$$V(S_{2,4}) \leftarrow V(S_{2,4}) + \alpha(G_t - V(S_{2,4}))$$
$$G_t = 10$$
$$V(S_{2,4}) \leftarrow \frac{8}{3} + \frac{1}{3} * (10 - \frac{8}{3})$$
$$V(S_{2,4}) \leftarrow \frac{8}{3} + \frac{22}{3} = 10$$

$$V(S_{2,5}) \leftarrow V(S_{2,5}) + \alpha(G_t - V(S_{2,5}))$$
$$G_t = 10$$
$$V(S_{2,5}) \leftarrow 10 + \frac{1}{3} * (10 - 10)$$
$$V(S_{2,5}) \leftarrow 10$$

(b) Consider now Episode 4 (magenta).
Temporal Difference error

$$V(S_t) \leftarrow V(S_t) + \alpha(R_{t+1} + \gamma V(S_{t+1}) - V(S_t))$$

TD error:

$$\delta_t = R_{t+1} + \gamma V(S_{t+1}) - V(S_t)$$

$$V(S_{2,1}) \leftarrow V(S_{2,1}) + \frac{1}{4}(R_{t+1} + \gamma V(S_{1,1}) - V(S_{2,1}))$$
$$V(S_{2,1}) \leftarrow -\frac{475}{6} + \frac{1}{4} * \left(-1 + 1 * 0 + \frac{475}{6}\right) = -59.625$$
$$\delta_t = -1 + 1 * 0 + \frac{475}{6} = \frac{469}{6} = 78.16$$

$$V(S_{1,1}) \leftarrow V(S_{1,1}) + \frac{1}{4}(R_{t+1} + \gamma V(S_{1,2}) - V(S_{1,1}))$$
$$V(S_{1,1}) \leftarrow 0 + \frac{1}{4} * (-1 + 1 * 0 - 0) = -0.25$$
$$\delta_t = -1 + 1 * 0 - 0 = -1$$

$$V(S_{1,2}) \leftarrow V(S_{1,2}) + \frac{1}{4}(R_{t+1} + \gamma V(S_{1,3}) - V(S_{1,2}))$$

$$V(S_{1,2}) \leftarrow 0 + \frac{1}{4} * (-1 + 1 * 8 - 0) = 1.75$$
$$\delta_t = -1 + 1 * 8 - 0 = 7$$

$$V(S_{1,3}) \leftarrow V(S_{1,3}) + \frac{1}{4}(R_{t+1} + \gamma V(S_{1,4}) - V(S_{1,3}))$$
$$V(S_{1,3}) \leftarrow 8 + \frac{1}{4} * (-1 + 1 * 9 - 8) = 8$$

$$\delta_t = -1 + 1 * 9 - 8 = 0$$

$$V(S_{1,4}) \leftarrow V(S_{1,4}) + \frac{1}{4}(R_{t+1} + \gamma V(S_{1,5}) - V(S_{1,4}))$$
$$V(S_{1,4}) \leftarrow 9 + \frac{1}{4} * (-1 + 10 - 9) = 9$$
$$\delta_t = -1 + 10 - 9 = 0$$

$$V(S_{1,5}) \leftarrow V(S_{1,5}) + \frac{1}{4}(R_{t+1} + \gamma V(S_{2,5}) - V(S_{1,5}))$$
$$V(S_{1,5}) \leftarrow 10 + \frac{1}{4} * (-1 + 10 - 10) = 9.75$$
$$\delta_t = -1 + 10 - 10 = -1$$

$$V(S_{2,5}) \leftarrow V(S_{2,5}) + \frac{1}{4}(R_{t+1} + \gamma V(S_{2,5}) - V(S_{2,5}))$$
$$V(S_{2,5}) \leftarrow 10 + \frac{1}{4} * (-1 + 10 - 10) = 9.75$$
$$\delta_t = -1 + 10 - 10 = -1$$

(c) Using the TD(λ)-algorithm,

$$V(S_t) \leftarrow V(S_t) + \alpha\left(\sum_{i=0}^{\infty} \lambda^i \delta_{t+i}\right)$$
$$\delta_t = \{78.16, -1, 7, 0, 0, -1, -1\}$$

$$\lambda = 0$$
$$V(S_{2,1}) \leftarrow -\frac{475}{6} + \frac{1}{4}(0 * 78.16 + 0 * -1 + 0 * 7 + 0 * 0 + 0 * 0 + 0 * -1 + 0 * -1) = -79.16$$

$$\lambda = 0.5$$
$$V(S_{2,1}) \leftarrow -\frac{475}{6} + \frac{1}{4}(0.5^0 * 78.16 + 0.5^1 * -1 + 0.5^2 * 7 + 0.5^3 * 0 + 0.5^4 * 0 + 0.5^5 * -1 + 0.5^6 * -1) = -59.32$$

$$\lambda = 1$$
$$V(S_{2,1}) \leftarrow -\frac{475}{6} + \frac{1}{4}(1 * 78.16 + 1 * -1 + 1 * 7 + 1 * 0 + 1 * 0 + 1 * -1 + 1 * -1) = -58.62$$