$a=-۲۰$, $b=+۲$, $c=-۱$, $d=+۴$, $e=-۳۰$, $f=+۷$ (۲

$g=+۲$, $h=+۱$, $i=+۱۰$, $j=-۴۰$, $\alpha=۰/۹$, $\gamma=۰/۸$

5

Q-learning :

$Q(S_t, A_t) =$

10 $Q(S_t, A_t) + \alpha\left(R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)\right)$

$Q(S_0, up) =$

15 $\underbrace{Q(S_0, up)}_{b} + ۰/۹\left(R_1 + ۰/۸ \overbrace{Q(S_1, right)}^{d} - Q(S_0, up)\right)$

$= ۲ + ۰/۹\left(-۱ + ۰/۸ \times ۴ - ۲\right) = ۲/۱۸$

20

$Q(S_1, right) = ۴ + ۰/۹\left(-۱ + ۰/۸ \times \overbrace{Q(S_۲, right)}^{f} - ۴\right)$

$= ۳/۸۲$

25 $Q(S_۲, right) = ۷ + ۰/۹\left(-۱ + ۰/۸ \times \overbrace{Q(S_۳, right)}^{i} - ۷\right)$

Soroush $= ۷/۹$

$$Q(S_3, down) = -40 + 0.9\left(-100 + 0.8 \times \overbrace{Q(terminal\_state_5, .)}^{0} - (-40)\right)$$

$$= -94$$

به طور خلاصه : $Q(S_0, up) = 4.18$

$Q(S_1, right) = 13.82$

$Q(S_2, right) = 41.9$

$Q(S_3, down) = -94$

Sarsa :

$Q(S_t, A_t) =$

$$Q(S_t, A_t) + \alpha(R_{t+1} + \gamma Q(S_{t+1}, A_{t+1}) - Q(S_t, A_t))$$

$Q(S_0, up) =$

$$\underbrace{Q(S_0, up)}_{b} + 0.9\left(R_1 + 0.8 \overbrace{Q(S_1, right)}^{d} - Q(S_0, up)\right)$$

Soroush

$$Q(s_0, up) = 2 + 0.9(-1 + 0.8 \times 4 - 2) = 2.18$$

$$Q(s_1, right) = 4 + 0.9(-1 + 0.8 \times \underbrace{Q(s_2, right)}_{f} - 4)$$

$$= 3.82$$

$$Q(s_2, right) = 4 + 0.9(-1 + 0.8 \times \underbrace{Q(s_2, down)}_{j} - 4)$$

$$= -29.1$$

$$Q(s_3, down) = -40 + 0.9(-100 + 0.8 \times \underbrace{Q(terminal\_state)}_{0}) - (-40)$$

$$= -94$$

بطور خلاصه: $Q(s_0, up) = 2.18$

$$Q(s_1, right) = 3.82$$

$$Q(s_2, right) = -29.1$$

$$Q(s_3, down) = -94$$