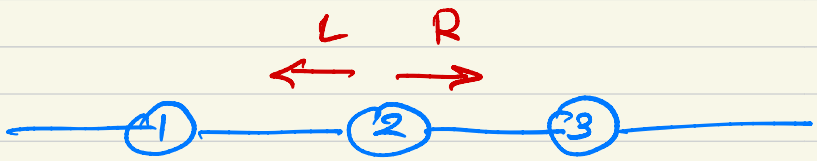
HW5 Solutions Q4

Multi-agent systems (Vrije Universiteit Amsterdam)

HW 5: Q5: Q-learning & SARSA



$\alpha = 0.9$, $\gamma = 2/3$: $2 \xrightarrow{R} 3$ (deterministic transitions)

Q-learning:

$$q(2,R) \leftarrow q(2,R) + \alpha\left[r(2,R) + \gamma \max_{a'} q(3,a') - q(2,R)\right]$$

$q(2,R) = 5$, $\alpha = 0.9$, $r(2,R) = -1$, $\gamma = 2/3$, $\max_{a'} q(3,a') = \max\{6,3\} = 6$, $q(2,R) = 5$

$$= 5 + 0.9\left[-1 + \frac{2}{3}\cdot 6 - 5\right] = 5 + 0.9[-2]$$

$$= 3.2$$

Expected SARSA

$$\sum_a \pi(a|s_{t+1})\, q_\pi(s_{t+1},a) = \sum_a \pi(a|3)\, q_\pi(3,a)$$

$$= \pi(L|3)\, q_\pi(3,L) + \pi(R|3)\, q_\pi(3,R)$$

$\pi(L|3) = 1/2$, $q_\pi(3,L) = 3$, $\pi(R|3) = 1/2$, $q_\pi(3,R) = 6$

$$= 9/2$$

Hence: $q_\pi(2,R) \leftarrow q_\pi(2,R) + 0.9\left[-1 + \gamma \cdot \frac{9}{2} - 5\right]$

$q_\pi(2,R) = 5$, $\gamma = 2/3$

$$= 5 + 0.9[-3] = 2.3$$