

Q learning



learning rate $\alpha \in [0:1]$
discount factor $\lambda \in [0:1]$

$$Q'(s, a) = (1 - \alpha)Q(s, a) + \alpha \text{ Improved Estimate}$$

$$= (1 - \alpha)Q(s, a) +$$

$$\alpha (r + \lambda \cdot \underset{a}{\operatorname{argmax}}_a Q[s', \operatorname{argmax}_a Q[s', a']])$$

S. $\begin{matrix} & \begin{matrix} 1 & 2 & 3 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \end{matrix} & \begin{bmatrix} 4 & 8 & 2 \\ 1 & 0 & 6 \\ 3 & 1 & 2 \end{bmatrix} \end{matrix}$ $r = 4$
 $2 \rightarrow 1$
 $s = 4$
 $\alpha = 0.1$ $\lambda = 0.1$

$$Q'[2, 3] = (1 - 0.1) \cdot 6 +$$

$$0.1 \times [4 + 0.1 \times 8]$$

$$= 5.4$$

