Q learning

learning rate
$\alpha \in [0:1]$
$\lambda \in (0:1)$ → discount rate

$Q'[S,a] = (1-\alpha) Q[S,a] + \alpha$: Improved Estimate

$= (1-\alpha) Q[S,a] +$

$\alpha \left(r + \lambda \cdot Q[S', \underset{a'}{\text{argmax}} [Q(S,a')] \right)$

S	1	2	3
1	4	8	2
2	1	0	6
3	3	1	2

$r = 4$

$2 \rightarrow 1$

5.4

$\alpha = 0.1 \quad \lambda = 0.1$

$Q'[2,3] = (1-0.1) \cdot 6 + 0.1 \times [4 + 0.1 \times 8]$

$= 5.9$