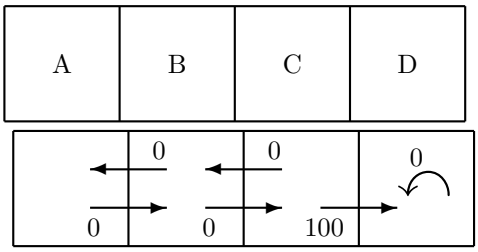
Reinforcement Learning Examples

Example 1

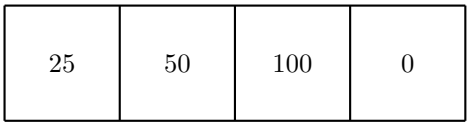
A simple board game:

$r(s,a) =$

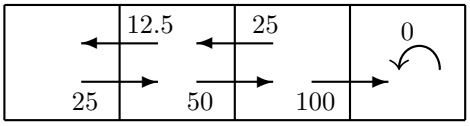


A discount factor: $\gamma = 1/2$.
If we can somehow figure out $V^*(s)$, it is easy to compute Q from V^*, π^* from Q, V^* from π^*.

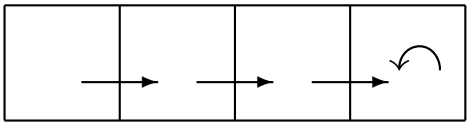
$V^*(s) =$



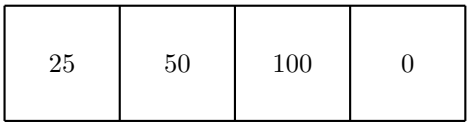
$Q(s,a) =$



$\pi^*(s) =$

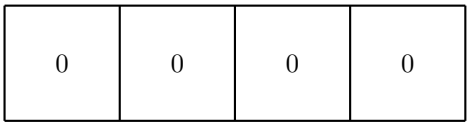


$V^*(s) =$

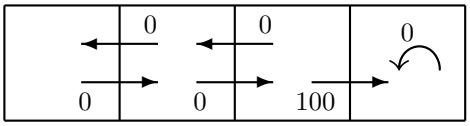


The offline algorithm: start by guessing $\hat{V}$.

$\hat{V} =$

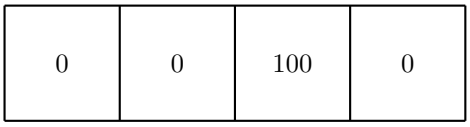


$\hat{Q}(s,a) =$

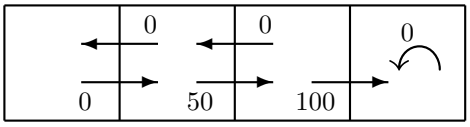


$\hat{\pi} =$

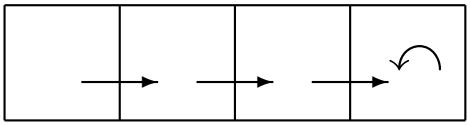
$\hat{V} =$



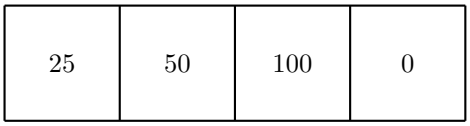
$\hat{Q}(s,a) =$



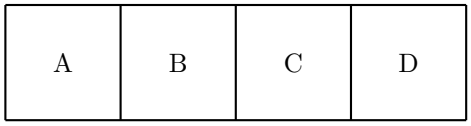
$\hat{\pi} =$



$\hat{V} =$

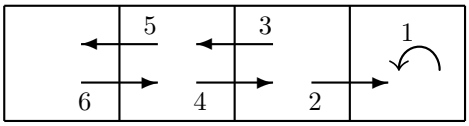


The online algorithm.



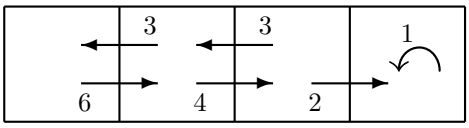
start with arbitrary Q

$\hat{Q}(s,a) =$



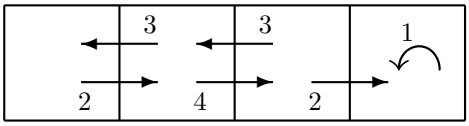
$Q(B, \text{left}) = 0 + 0.5 \cdot 6 = 3$

$\hat{Q}(s,a) =$



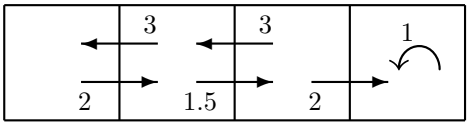
$Q(A, \text{right}) = 0 + 0.5 \cdot 4 = 2$

$\hat{Q}(s,a) =$



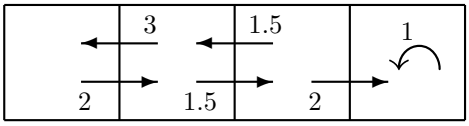
$Q(B, \text{right}) = 1.5$

$\hat{Q}(s,a) =$



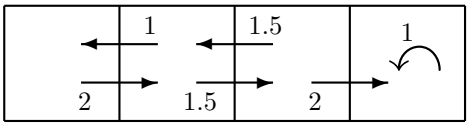
$Q(C, \text{left}) = 1.5$

$\hat{Q}(s,a) =$



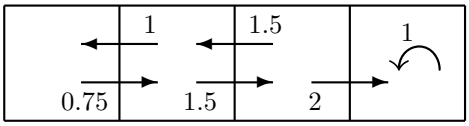
$Q(B, \text{left}) = 1$

$\hat{Q}(s,a) =$



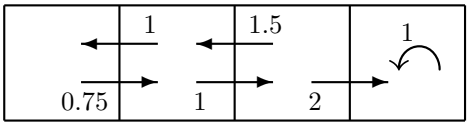
$Q(A, \text{right}) = 0.75$

$\hat{Q}(s,a) =$



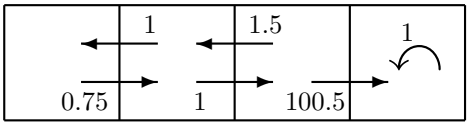
$Q(B, \text{right}) = 1$

$\hat{Q}(s,a) =$



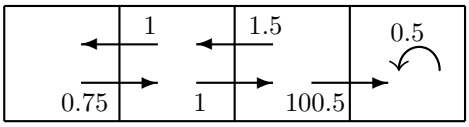
$Q(C, \text{right}) = 100.5$

$\hat{Q}(s,a) =$



$Q(D, \text{itself}) = 0.5$

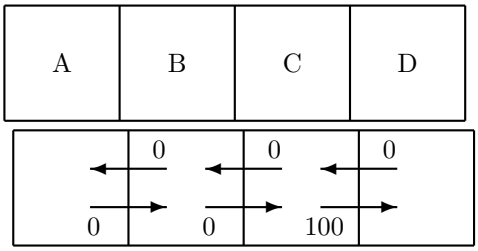
$\hat{Q}(s,a) =$



additional iterations are needed

Example 2

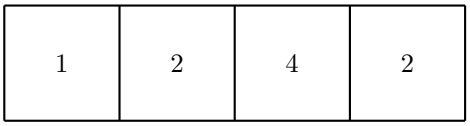
$r(s,a) =$



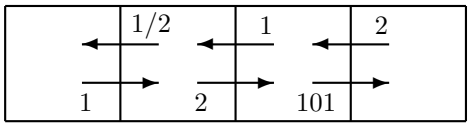
A discount factor: $\gamma = 1/2$.
Use intuition to compute V^*. If we are at D we can go to C, and then back to D and get rewarded.

$$\begin{aligned}
V^*(D) &= 0 + \frac{1}{2}\cdot 100 + 0 + \frac{1}{2^3}\cdot 100 + \ldots \\
&= 100 \cdot \frac{1}{2} \cdot (1 + (\frac{1}{4}) + (\frac{1}{4})^2 + \ldots) \\
&= 100 \cdot \frac{1}{2} \cdot \frac{1}{1 - 1/4} = 100 \cdot \frac{2}{3} \\
V^*(C) &= 100 + \frac{1}{2}V^*(D) = 100 \cdot \frac{4}{3} \\
V^*(B) &= 0 + \frac{1}{2}V^*(C) = 100 \cdot \frac{2}{3} \\
V^*(A) &= 0 + \frac{1}{2}V^*(B) = 100 \cdot \frac{1}{3}
\end{aligned}$$

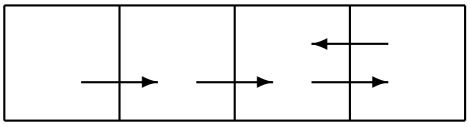
$V^*(s) = \frac{100}{3}$



$Q(s,a) = \frac{100}{3}$



$\pi^*(s) =$



$V^*(s) = \frac{100}{3}$