Policy Iteration

Init π_0. for $k = 0, 1, 2, \ldots, K$.

- Policy Eval: Compute Q^{π_k}
- Policy Improvement: $\pi_{k+1} \leftarrow \pi_{Q^{\pi_k}}$

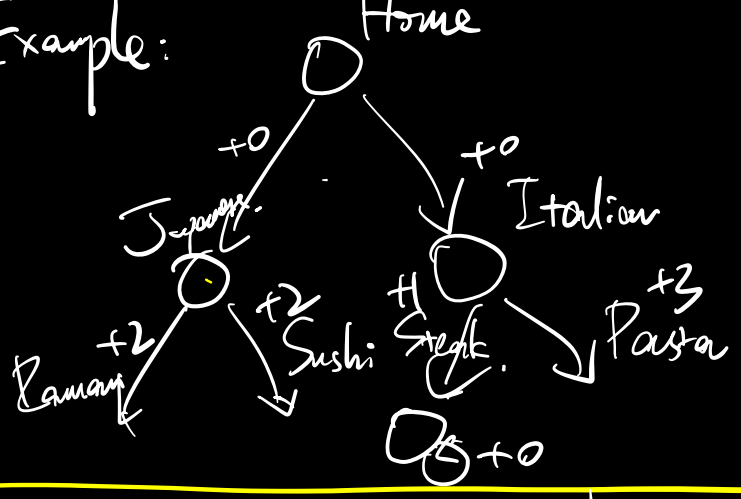
$\pi_k = \pi^*. \Rightarrow Q^{\pi_k} = Q^{\pi^*} = Q^*$

$\Rightarrow \pi_{k+1} = \pi_{Q^{\pi_k}} = \pi_{Q^*} = \pi^*$ ✓

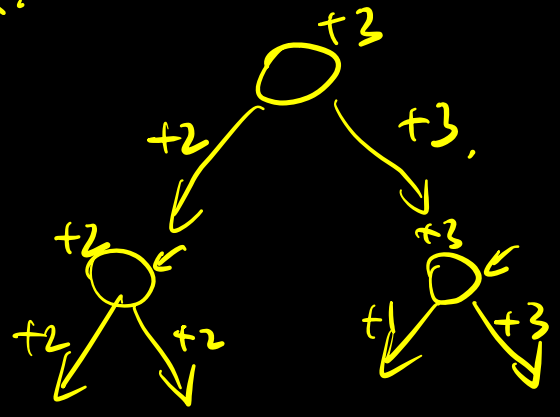
$\pi_{f(s)} = \arg\max_a f(s,a)$

Δ.

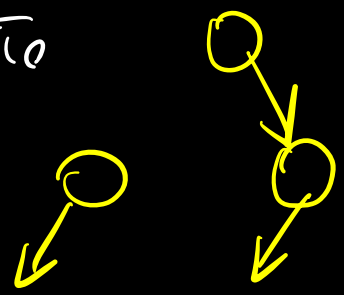
Example:



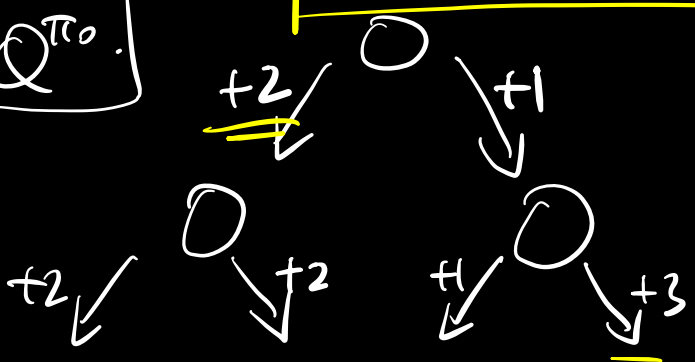
VI:



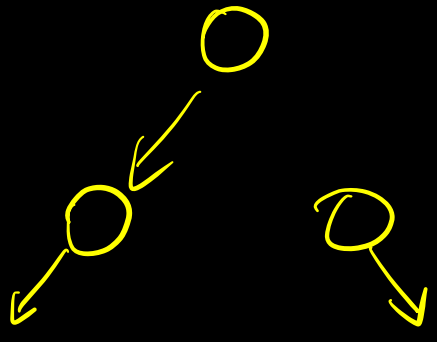
π_0



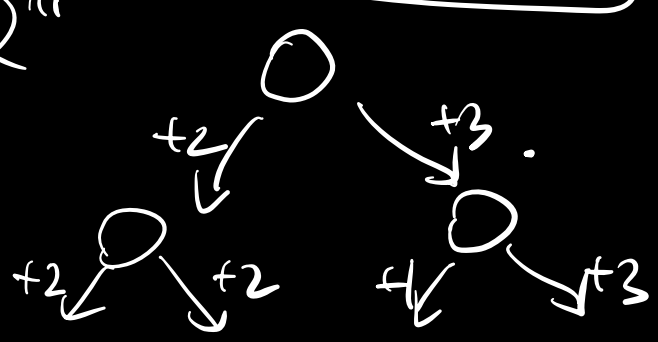
Q^{π_0}



π_1

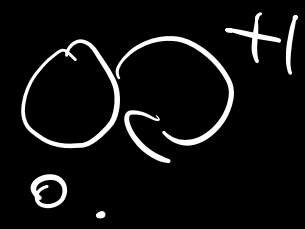


Q^{π_1}



Convergence:

PI converges in finite #iter



(Tf) $\quad P(s,a) + \gamma \mathbb{E}_{\cdot\cdot}[\max f]$

Monotone Policy Improvement.

$\forall k, \quad Q^{\pi_{k+1}} \geq Q^{\pi_k}.$

if $\pi_k \neq \pi^*$, then "$\geq$" is strict in at least one (s,a).

Corollary: if $\pi_k \neq \pi^*$. then $\pi_0, \pi_1, \pi_2, \cdots, \pi_k$ are all distinct

VI: $f_0, f_1, f_2, \cdots, f_k.$

PI: $Q^{\pi_0}, Q^{\pi_1}, Q^{\pi_2}, \cdots.$

$\|f_k - Q^*\|_\infty$ $\quad \forall f \in \mathbb{R}^{S \times A}.$

$2 \cdot \|f - Q^*\|_\infty$

$$\|V^* - V^{\pi_f}\|_\infty \leq \frac{\qquad\qquad}{1-\delta}.$$

$$\|f\|_\infty = \max_{s,a} |f(s,a)|$$