lect-8

$V^{\pi}(s) =$

$Q^{\pi}(s, a)$

- Optimal value function —

$$V^{*}(s) = \max_{\pi \in \Pi} V^{\pi}(s)$$

Define $\pi \geq \pi'$ if $\forall s \in S$

$$V^{\pi}(s) \geq V^{\pi'}(s)$$

- Partial ordering of policy

Stuck with π, π'

Intuitively given π & π' we could decode π'' that applies π where $V^{\pi}(s) \geq V^{\pi'}s$ & applies π' where $V^{\pi'}(s) > V^{\pi}(s)$

- Formally, there exist at least one π^*, such that

$$\pi^* \geq \pi \quad \forall \pi \in \Pi$$

(finite MDP, bounded rewards)

- π^* called "optimal policy"
- Stricter definition of π^*
  - $\pi^* \geq \pi$, $\forall \pi$ used in value function based methods
  - $J(\pi)$ used in policy search papers
- Properties
  - $V^* = \cancel{\pi} V^{\pi^*}$
  - Given a finite MDP, π^* is not unique but V^* is unique.

π^* is not guaranteed to be unique
V^* is unique
with bounded reward

$V =$

- Optional q function

$$q^*(s,a) \triangleq \max_{\pi} q^{\pi}(s,a)$$

- if we have, q^*

$$\pi^*(s,a) \neq 0 \text{ only if } a \in \operatorname{argmax}_{\cancel{a}} q^*(s,a)$$

- if we have V^*,

$$\pi^*(s,a) \neq 0 \text{ only if } \arg\max_{a} \sum_{s'} P(s,a,s')\left[R(s,a,s') + \gamma V^*(s')\right]$$

- Bellman optimality eq^n

$$V^{\pi^*}(s) = \sum_a \pi^*(s,a) \sum_{s'} P(s,a,s')\left(R(s,a,s') + \gamma \cdot V^*(s')\right)$$

$$= \max_a \sum_{s'} P(s,a,s')\left[R(s,a,s') + \gamma \cdot V^*(s')\right]$$

- Bellman optimality eq^n for q^*

$$q^*(s,a) = \sum_{s'} P(s,a,s')\left(R(s,a,s') + \gamma . \max_{a} q^*(s,a)\right)$$

Bellman optimality eq^n only hold for v^* & q^*

- Planning algorithm
    - Write out Bellman optimality eq^n for q^*
    - System of $|S| \times |A|$ eq^n with $|S| \times |A|$ unknowns
    - Solve & retrieve π^*