Policy evaluation

Convergence and Spectral Radius

Ahmed Touati

Reinforcement Learning class

Policy evaluation Algorithm

Input π, the policy to be evaluated
Initialize an array $V(s) = 0$, for all $s \in \mathcal{S}^+$
Repeat
  $\Delta \leftarrow 0$
  For each $s \in \mathcal{S}$:
    $v \leftarrow V(s)$
    $V(s) \leftarrow \sum_a \pi(a|s) \sum_{s',r} p(s', r|s, a)\big[r + \gamma V(s')\big]$
    $\Delta \leftarrow \max(\Delta, |v - V(s)|)$
until $\Delta < \theta$ (a small positive number)
Output $V \approx v_\pi$

Convergence and contraction mappings

Banach Fixed-Point Theorem

Suppose U is a Banach Space and $T : U \to U$ is a contraction mapping. Then:

- there exists a unique v^* in U such that $Tv^* = v^*$; and
- for arbitrary v^0 in U. The sequence $\{v^n\}$ defined by

$$v^{n+1} = Tv^n = T^{n+1}v^0$$

  converges to v^*.

Spectral Radius

Definition

Let $A \in \mathbb{R}^{dxd}$ a matrix and (λ_i) are his eigenvalues, we define the **Spectral Radius** of A denoted $\rho(A)$ as $\rho(A) = \max_i\{\lambda_i\}$

Gelfand's Formula

$$\rho(A) = \lim_{n\to\infty} ||A^n||^{\frac{1}{n}}$$

Neuman expansion of inverses

If $\rho(A) < 1$ than $(I - A)^{-1}$ exists and satisfies:

$$(I - A)^{-1} = \lim_{N\to\infty} \sum_{n=0}^{N} A^n$$

proof

On blackboard

Policy iteration revisited

- Vectorized form of Bellman equation: if d is the number of state, we have $V_\pi \in \mathbb{R}^d$
$$V_\pi = R_\pi + \gamma P_\pi V_\pi$$
where
  - $R_\pi \in \mathbb{R}^d, R_\pi(s) = \mathbb{E}[R_t | S_t = s, A_{t:\infty} \sim \pi]$.
  - $P_\pi \in \mathbb{R}^{dxd}$ transition matrix: $(P_\pi)_{i,j} = \mathbb{P}[s_j | s_i]$

Policy iteration revisited

- Vectorized form of Bellman equation: if d is the number of state, we have $V_\pi \in \mathbb{R}^d$

$$V_\pi = R_\pi + \gamma P_\pi V_\pi$$

  where

  - $R_\pi \in \mathbb{R}^d, R_\pi(s) = \mathbb{E}[R_t | S_t = s, A_{t:\infty} \sim \pi]$.
  - $P_\pi \in \mathbb{R}^{dxd}$ transition matrix: $(P_\pi)_{i,j} = \mathbb{P}[s_j | s_i]$

- The solution of Bellman equation is:

$$V_\pi = (I - \gamma P_\pi)^{-1} R_\pi$$

Policy iteration revisited

- Vectorized form of Bellman equation: if d is the number of state, we have $V_\pi \in \mathbb{R}^d$

$$V_\pi = R_\pi + \gamma P_\pi V_\pi$$

where

  - $R_\pi \in \mathbb{R}^d, R_\pi(s) = \mathbb{E}[R_t | S_t = s, A_{t:\infty} \sim \pi]$.
  - $P_\pi \in \mathbb{R}^{dxd}$ transition matrix: $(P_\pi)_{i,j} = \mathbb{P}[s_j | s_i]$

- The solution of Bellman equation is:

$$V_\pi = (I - \gamma P_\pi)^{-1} R_\pi$$

- Policy iteration rewritten: For each k:

$$V_\pi^{k+1} = R_\pi + \gamma P_\pi V_\pi^k$$

Policy iteration revisited

- Vectorized form of Bellman equation: if d is the number of state, we have $V_\pi \in \mathbb{R}^d$
$$V_\pi = R_\pi + \gamma P_\pi V_\pi$$
where
  - $R_\pi \in \mathbb{R}^d, R_\pi(s) = \mathbb{E}[R_t | S_t = s, A_{t:\infty} \sim \pi]$.
  - $P_\pi \in \mathbb{R}^{dxd}$ transition matrix: $(P_\pi)_{i,j} = \mathbb{P}[s_j | s_i]$
- The solution of Bellman equation is:
$$V_\pi = (I - \gamma P_\pi)^{-1} R_\pi$$
- Policy iteration rewritten: For each k:
$$V_\pi^{k+1} = R_\pi + \gamma P_\pi V_\pi^k$$
- By recursion:
$$V^K = (\gamma P)^K V^0 + \sum_{k=0}^{K-1} (\gamma P)^k R$$

Convergence proof: final step

- P is a right stochastic matrix then $\rho(P) = 1$. So, $\rho(\gamma P) = \gamma < 1$.

proof

On blackboard

Convergence proof: final step

- P is a right stochastic matrix then $\rho(P) = 1$. So, $\rho(\gamma P) = \gamma < 1$.
- We can apply the Neuman expansion theorem:

$$\lim_{K \to \infty} \sum_{k=0}^{K} (\gamma P)^k = (I - \gamma P)^{-1}$$

.

- We have also that the term $(\gamma P)^K V^0$ vanishes to zero.
- As result, we show that

$$\lim_{K \to \infty} V^K = (I - \gamma P)^{-1} R$$

.

Questions?