Monday, October 19, 2020

Lecture 17

CS 59000_RL

MDP

- Optimality
- Value iteration
- Policy iteration

Bellman optimality operator is a contraction.

In particular:

For $U, V \in R^{|X|}$

$$\|TU - TV\|_\infty \leq \lambda \|U - V\|_\infty$$

Theorem (Banach Fixed-Point) Suppose Ω is a Banach space and $\mathcal{L} : \Omega \to \Omega$ is a contraction map under a norm $\|\cdot\|$, and parameter λ. Then

i) There exist a unique V^+ in Ω such that $\mathcal{L}V^+ = V^+$

ii) For an arbitrary $V_0 \in \Omega$, the sequence $\{V_n\}$ defined as $V_{n+1} = \mathcal{L}V_n = \mathcal{L}^{n+1} V_0$ converges to V^+

Monday, October 19, 2020

Proof:

$$\underbrace{\|V_{n+m} - V_n}_{V_{n+k}-V_{n+k}}\| \leqslant \sum_{k=0}^{m-1} \|V_{n+k+1} - V_{n+k}\|$$

$$= \sum_{k=0}^{m-1} \|\underbrace{\mathcal{L}^{n+k} V_1 - \mathcal{L}^{n+k} V_0}\|$$

$$\leqslant \sum_{k=0}^{m-1} \lambda^{n+k} \|V_1 - V_0\| = \underbrace{\frac{\lambda^n (1-\lambda^m)}{1-\lambda} \|V_1 - V_0\|}_{\leqslant \varepsilon}$$

Therefore, $\{V_n\}$ is a cauch sequence, i.e. for any ε, there exist a sufficiently large n such that $\|V_{n+m} - V_n\| \leqslant \varepsilon$

Since Ω is Banach, then the cauchy sequence $\{V_n\}$ converges. Let V^+ denote the limit.

Now we show that $\mathcal{L} V^+ = V^+$

Monday, October 19, 2020

For $n \geq 1$ $\quad \in \pm V_n$

$$0 \leq \| \mathcal{L} V^+ - V^+ \|$$

$$\leq \| \mathcal{L} V^+ - V_n \| + \| V_n - V^+ \|$$

$$\leq \lambda \| V^+ - V_{n-1} \| + \| V_n - V^+ \|$$

when taking the limit $n \to \infty$; $\lim_{n \to \infty} \| V_n - V^+ \| = 0$

Ergo $\| \mathcal{L} V^+ - V^+ \| = 0 \Rightarrow \underline{\mathcal{L} V^+ = V^+}$

So, there exist a solution

Uniqness: let's imagine there exist another solution U^+ such that

(i) $U^+ \neq V^+$
(ii) and $\mathcal{L} U^+ = U^+$

$$V^+ - U^+ = \mathcal{L} V^+ - \mathcal{L} U^+$$

$$\hookrightarrow \| V^+ - U^+ \| = \| \mathcal{L} V^+ - \mathcal{L} U^+ \| \leq \lambda \| V^+ - U^+ \|$$

$$\Rightarrow \| V^+ - U^+ \| = 0 \Rightarrow V^+ = U^+$$

Monday, October 19, 2020

How we use it for MDPs?

Using the fact that $R^{|\mathcal{X}|}$ is a Banach space we have have the Bellman optimality equation has a unique solution. $\underline{V^+ = T V^+}$

we need to show V^* is V^+.

$$V^+(x) = \max_a \bar{r}(x,a) + \lambda \sum_{x'} P(x'|x,a) V^+(x')$$
$$\geq \bar{r}(x,a) + \lambda \sum_{x'} P(x'|x,a) V^+(x')$$

Therefore:

$$V^+ \geq \bar{r}_{\pi_1} + \lambda P_{\pi_1} V^+$$
$$\geq \bar{r}_{\pi_1} + \lambda P_{\pi_1} (\bar{r}_{\pi_2} + \lambda P_{\pi_2} V^+)$$
$$= \bar{r}_{\pi_1} + \lambda P_{\pi_1} \bar{r}_{\pi_2} + \lambda^2 P_{\pi_1} P_2 V^+$$
$$\geq \bar{r}_{\pi_1} + \lambda P_{\pi_1} \bar{r}_{\pi_2} + \dots \dots \lambda^n \overbrace{P_{\pi_1} \dots P_{\pi_n}}^{\bar{P}_{\pi_n}} V^+$$

$$V^* = V^{\pi^*} = \sum_{t=1}^{\infty} \lambda^{t-1} \bar{P}^{t-1}_{\pi^*_t} \bar{r}_{\pi}$$

Monday, October 19, 2020

This inequality hold for any sequence of policies $\bar{\pi}$

$$\Rightarrow V^{+} \geq V^{\bar{\pi}} + \lambda^{n} \bar{P}^{n}_{\bar{\pi}_n} V^{+} - \sum_{k=n}^{\infty} \lambda^{k} \bar{P}^{k}_{\bar{\pi}_k} r_{\bar{\pi}_{k+1}}$$

since $V^{\bar{\pi}} = \sum \lambda^{t-1} \bar{P}^{t-1}_{\bar{\pi}_t} r_{\bar{\pi}_t}$

First: $\|\lambda^{n} \bar{P}^{n}_{\bar{\pi}_n} V^{+}\|_{\infty} \leq \lambda^{n} \|V^{+}\|_{\infty}$

second: since $\|r(x,a)\| \leq M < \infty$, we have:

vector of all ones.

$$\frac{-\lambda^{n} M e}{1-\lambda} \leq \sum_{k=n}^{\infty} \lambda^{k} \bar{P}^{k}_{\bar{\pi}_k} r_{\bar{\pi}_{k+1}}$$

Therefore, for any $\varepsilon > 0$, there exists an n such that

$$V^{+} - V^{\bar{\pi}} \geq -\varepsilon \rightarrow V^{+} \geq V^{\bar{\pi}}$$

$$\Rightarrow V^{+} \geq V^{\pi^*} = V^{*}$$

---

we are left with showing $V^{+} \leq V^{*}$

Monday, October 19, 2020

Lemma: If $V \geq 0$, then $(I - \lambda P_\pi)^{-1} V \geq V$

proof: Using the fact that, $\| I - \lambda P_\pi \| < 1$,

and Neumann series of invertable operators:

$$(I - \lambda P_\pi)^{-1} V = V + \underbrace{\lambda P_\pi V + \lambda^2 P_\pi^2 V + \cdots}_{\text{non negative}} \geq V \geq 0$$

For Bellman optimality equation, we have $TV^+ = V^+$

Therefore, for any $\varepsilon > 0$, there exist a π, such that

$$V^+ \leq r_\pi + \lambda P_\pi V^+ + \varepsilon e$$

$$\Rightarrow r_\pi + \varepsilon e - (I - \lambda P_\pi) V^+ \geq 0$$

This is positive: Using the lemma we just proved we have

$$V^+ \leq (I - \lambda P_\pi)^{-1} (r_\pi + \varepsilon e) = V^\pi + (1 - \lambda P_\pi)^{-1} \varepsilon e$$

$$V^+ \leq V^\pi + (1 - \lambda)^{-1} \varepsilon e \quad \text{for any } \varepsilon > 0$$

Monday, October 19, 2020

$$\Rightarrow V^{+} \leq V^{\pi^*}$$

$$\Rightarrow V^{+} = V^{*}$$

Value iteration:

- Initialize with V_0
- Repeatedlly apply $V_{n+1} = T V_n$

---

we had in the analysis:

$$\underbrace{\|V_{n+1} - V^*\|_\infty} = \|T V_n - T V^*\|_\infty \leq \lambda \underbrace{\|V_n - V^*\|_\infty}$$

$$\leq \underline{\lambda}^{n+1} \|V_0 - V^*\|_\infty$$

we showed value iteration converges exponentially fast.

Policy iteration:

- start with an arbitrary stationary and memory-less policy π_0, then solve for V

Policy evalution step $\longrightarrow (I - \lambda P_{\pi_n}) V = r_{\pi_n}$

Monday, October 19, 2020

call the solution V_n (value of Π_n)

- Improve policy as $\Pi_{n+1} \in \arg\max (r_\Pi + \lambda P_\Pi V_n)$

Policy improvement step

Is Π_{n+1} better than Π_n?

we know: $r_{\Pi_{n+1}} + \underbrace{\lambda P_{\Pi_{n+1}}}\, V_n > r_{\Pi_n} + \lambda P_{\Pi_n} V_n = V_n$

$$\Rightarrow \underbrace{r_{\Pi_{n+1}}} \geq \underbrace{\left(I - \lambda P_{\Pi_{n+1}}\right)} V_n$$

Applying the lemma we previously used
(for $v \geq 0 \to (I - \lambda P_\Pi)^{-1} v \geq v$)

we have: applying $\left(I - \lambda P_{\Pi_{n+1}}\right)^{-1}$

$$V_{n+1} = \left(I - \lambda P_{\Pi_{n+1}}\right)^{-1} r_{\Pi_{n+1}} \geq V_n$$

Policy iteration algorithm converges to an optimal policy and optimal value.

refer to 6.5 MDP Book by Martin. Puterman.

Monday, October 19, 2020

Example:

For $\lambda' < \lambda$

$$\downarrow \qquad \downarrow$$

$$V^*_{\lambda'} \qquad V^*_{\lambda}$$

after n step

Iteration with $\lambda \rightarrow \lambda^n$

$$\|V^*_\lambda - V^n_\lambda\|_\infty < \lambda^n$$

$$\Downarrow$$

$$\|V^*_{\lambda'} - V^n_{\lambda'}\|_\infty < \lambda'^n$$

If n is small and $\|V^*_{\lambda'} - V^*_\lambda\| \leqslant \delta$

$$\lambda^n > \underbrace{\delta} + \underbrace{\lambda'^n}$$