RL Theory[1]: Meeting 5 (Chapter 2)

Shivam Garg, RLAI@UAlberta

20th October 2020

[1] based on https://rltheorybook.github.io/

Sample Complexity

Q-value iteration

- What is it? Why do we need it?
- Previously, we discussed DP algorithms like value iteration: $\boxed{Q^{(k+1)} = \mathcal{T} Q^{(k)}}$ for $k = 0, 1, 2, \ldots$. Let $\pi^{(k)} = \pi_{Q^{(k)}}$. Then for $k \geq \frac{1}{1-\gamma} \log\left(\frac{2}{\epsilon(1-\gamma)}\right)$,

$$V^{\pi^{(k)}} \geq V^* - \epsilon \mathbb{1}.$$

- This assumes access to the true transition dynamics P, which is not available.
- So we address this question: How do these methods perform when we don't have the true P?

Sample Complexity (contd)

$s_1 a_1$ — s_1 — 10, s_2 — 20, s_3 — 30

- We will begin by assuming a naïve model of the environment $\hat{P}$, defined as

$s' \sim P(\cdot|s,a)$

$$\hat{P}(s'|s,a) = \frac{\text{count}(s', a, \cancel{a}\, s)}{N(s,a)}$$

- Define $\hat{M}$, $\hat{V}^{\pi}$, $\hat{Q}^{\pi}$, $\hat{Q}^{*}$, $\hat{\pi}^{*}$
- Then we will see that given the inaccuracy in this model, how accurate our estimates of, say, $\hat{Q}^{\pi}$ can be?

$\hat{V}^{\pi}$

$\hat{\pi}^{*}$

$= \hat{Q}^{*}$

$\hat{M} = (S, A, R, \hat{P}, \gamma)$

$r(s,a)$

Sample Complexity for a Naïve Model

- There exists a constant c. Let $\epsilon \in \left(0, \frac{1}{1-\gamma}\right)$. If,

$N(s,a) = N$ $\quad$ $N \to \#(s,a)$

$|S||A|N$

$$\#\text{ of samples} \geq \frac{\gamma}{(1-\gamma)^4} \frac{|\mathcal{S}|^2|\mathcal{A}| \log\left(\frac{c|\mathcal{S}||\mathcal{A}|}{\delta}\right)}{\epsilon^2}$$

$s_1 a_1 - 100$

$s_1 a_2 - 10$

$\gamma \longrightarrow \hat{P} \longrightarrow Q$

then following hold with a probability of greater than $1-\delta$:

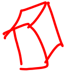
- (Model Accuracy) $\quad \max_{s,a} \|P(\cdot|s,a) - \hat{P}(\cdot|s,a)\|_1 \leq (1-\gamma)^2 \frac{\epsilon}{2}$ $\quad \leq \epsilon'/2$
- (Uniform Value Accuracy) $\quad \|Q^\pi - \hat{Q}^\pi\|_\infty \leq \frac{\epsilon}{2} \quad \forall \pi$
- (Near Optimal Planning) $\quad \|Q^* - \hat{Q}^*\|_\infty \leq \epsilon$

$\epsilon' = (1-\gamma)\epsilon$

Hoeffding's Inequality

- $$\mathbb{P}\left(\left|\mathbb{E}[X] - \frac{\sum_{i=1}^{N} X_i}{N}\right| \leq (b_+ - b_-)\sqrt{\frac{\ln(2/\delta)}{2N}}\right) \geq 1 - \delta.$$

scalar



$X_1, X_2, X_3, \dots \sim P$

$X_i \in [b_-, b_+]$

$b_- \to 1$, $b_+ \to 6$

Hoeffding

Union Bound

$A_1\ A_2\ A_3 \dots \in \mathcal{A}$

(s,a) ▶ $\to P(\bigcup_i A_i) \le \sum P(A_i)$

$$\mathbb{P}\left(|A_i| \le c(\delta)\right) \ge 1-\delta \quad \Rightarrow \quad \mathbb{P}\left(\max_i |A_i| \le c(\delta)\right) \ge 1 - \underbrace{|\mathcal{A}|\delta}_{\delta'}.$$

$\boxed{\hat{P}(s'|s,a) \le C}$

$1 - P(|A_i| \le c) \le \delta$

$c\left(\frac{\delta'}{|\mathcal{A}|}\right)$

$P(|A_i| \ge c) \le \delta$

$1 - P\left(\bigcup(|A_i| \ge c)\right) \ge 1 - |\mathcal{A}|\delta$

$P\left(\overline{\bigcup_i |A_i| \ge c}\right) \ge 1 - |\mathcal{A}|\delta$

$\bigcap |A_i| \le c \Rightarrow \max_i |A_i| \le c$

Proof of Claim 1: (Step 1)

$\delta \ll 1$

$$P\left(\|P(\cdot|s,a) - \hat{P}(\cdot|s,a)\|_1 \le c\sqrt{\frac{|S|\log(1/\delta)}{m}}. \right) \ge 1-\delta$$

(s,a)

$q = [\ \ ]_{d\times 1}$

$$P\left(\|\hat{q} - \bar{q}\|_1 \le \sqrt{|S|}\left(\frac{1}{\sqrt{m}} + \epsilon\right)\right) \ge \underbrace{1 - e^{-N\epsilon^2}}_{\ge 1-\delta}$$

$\frac{\sum q_i}{N}$ $\quad E[q]$

$\epsilon \le \sqrt{\frac{1}{m}\log(1/\delta)}$

$\epsilon\left(0, \frac{1}{1-\gamma}\right)$

$\delta \simeq 1$

$$\left(\frac{\sqrt{|S|}}{\sqrt{m}} + \sqrt{\frac{|S|}{m}\log(1/\delta)}\right) \le c\sqrt{\overbrace{\frac{|S|}{m}\log\left(\frac{1}{\delta}\right)}^{\simeq 0}}$$

Proof of Claim 1: (Step 2)

- $$\#\text{ samples} \geq \frac{\gamma}{(1-\gamma)^4}\frac{|\mathcal{S}|^2|\mathcal{A}|\log\left(\frac{c|\mathcal{S}||\mathcal{A}|}{\delta}\right)}{\epsilon^2} \Rightarrow \max_{s,a}\|P(\cdot|s,a)-\hat{P}(\cdot|s,a)\|_1 \leq (1-\gamma)^2\frac{\epsilon}{2}$$

$|S||A|m$

$$P\left(\max_{s,a}\|p(\cdot|s,a)-\hat{p}(\cdot|s,a)\| \leq c\sqrt{\frac{|S|\log\left(\frac{|S||A|}{\delta}\right)}{m}}\right) \geq 1-\delta$$

$(1-\gamma)^2\,\epsilon/2$

$$\frac{1}{m} = \frac{(1-\gamma)^4\epsilon^2}{4|S|\log(|S||A|/\delta)}$$

$$|S||A|m \geq \frac{4|A||S|^2\log(|S||A|/\delta)}{(1-\gamma)^4\epsilon^2 c^2}$$

Proof of Claim 1: (Step 3)

- $$\# \text{ samples} \geq \frac{\gamma}{(1-\gamma)^4} \frac{|\mathcal{S}|^2 |\mathcal{A}| \log\left(\frac{c|\mathcal{S}||\mathcal{A}|}{\delta}\right)}{\epsilon^2} \Rightarrow \max_{s,a} \|P(\cdot|s,a) - \hat{P}(\cdot|s,a)\|_1 \leq (1-\gamma)^2 \frac{\epsilon}{2}$$

Simulation Lemma

$$Q^\pi = r + \gamma P^\pi Q^\pi \Rightarrow Q^\pi = (I - \gamma P^\pi)^{-1} r$$

- $$Q^\pi - \hat{Q}^\pi = \gamma\big(I - \gamma \hat{P}^\pi\big)^{-1}\big(P - \hat{P}\big)V^\pi.$$

$$(I - \gamma P^\pi)^{-1} r - (I - \gamma \hat{P}^\pi)^{-1} r$$

$$\gamma (I - \gamma \hat{P}^\pi)^{-1} \left[P^\pi - \hat{P}^\pi \right] \underbrace{(I - \gamma P^\pi)^{-1} r}_{Q^\pi}$$

$$P^\pi(s' a' | s, a)$$

$$= \sum_{a'} \pi(a'|s') \sum_{s'} p(s'|s,a) Q(s',a')$$

$$[P^\pi - \hat{P}^\pi] Q = (P - \hat{P}) V$$

$$\sum_{s'} p(s'|s,a) \underbrace{\sum_{a'} \pi(a'|s') Q^\pi(s',a')}_{V^\pi(s')}$$

Another Useful Result

- For $x \in |\mathcal{S} \times \mathcal{A}|$

$$\frac{1}{1-x} = 1 + x + x^2$$

$$\|\underbrace{(I - \gamma \hat{P}^{\pi})^{-1}}\overset{r}{x}\|_\infty \leq \frac{\|\overset{r}{x}\|_\infty}{1-\gamma}$$

$$\left\| \left(I + \gamma \hat{P}^{\pi} + \gamma^2 \hat{P}^{\pi} + \cdots \right) x \right\|_\infty$$

$$\leq \|x\|_\infty + \gamma \|\hat{P}^{\pi} x\|_\infty + \cdots$$

$$\leq \left(1 + \gamma + \gamma^2 + \cdots\right) \|x\|_\infty$$

Proof of Claim 2

Hölder's

$r \in [0,1]$

- $\sum_i |p_i||q_i| \le \|p\|_1 \|q\|_\infty$

$\|\hat{Q}^\pi - Q^\pi\|_\infty \le \epsilon/2.$

$p(s_1|s,a_1) \quad p(s_1|s,a)$

$|S||A| \times S$

$v(s_1)$

$v(s_2)$

$$\|\hat{Q}^\pi - Q^\pi\|_\infty = \gamma \|\underbrace{(I - \gamma \hat{P}^\pi)^{-1}} \underbrace{(P - \hat{P}) V^\pi}_{x}\|_\infty$$

$$\le \frac{\gamma}{1-\gamma} \underbrace{\|(P - \hat{P}) V^\pi\|_\infty}$$

$p(\cdot|s,a) \quad V^\pi$

$$\max_{s,a} \left| \sum_{s'} \left(p(s'|s,a) - \hat{p}(s'|s,a)\right) v^\pi(s') \right|$$

$$\le \max_{s,a} \sum_{s'} |p(s'|s,a) - \hat{p}(s'|s,a)| |v^\pi(s')|$$

$$\max_{s,a} \|p(\cdot|s,a) - \hat{p}(\cdot|s,a)\|_1 \|V^\pi\|_\infty \quad \frac{1}{1-\gamma}$$

$$\cancel{\frac{\gamma}{(1-\gamma)^2}} \cancel{(1-\gamma)^2} \epsilon^2/2$$

$$\boxed{\gamma(\epsilon/2)} \le \frac{\gamma}{1-\gamma}$$

$$\underbrace{\hat{Q}^{\pi} \qquad Q^{\pi}}$$

$$\hat{Q}^{*} - Q^{*}$$

$$\hat{P}(s'|s,a) \qquad X_i = \mathbb{I}(s_i = s')$$

$$\frac{\sum X_i}{n} \qquad \mathbb{E}[X_i]$$

$$\|(P - \hat{P})V^*\|_\infty = \max_{s,a} \left| \sum_{s'} P(s'|s,a)V^*(s') - \sum_{s'} \hat{P}(s'|s,a)V^*(s') \right|$$

$$\leq \frac{1}{1-\gamma}\sqrt{\frac{2}{n}\log\left(\frac{2|S||A|}{\delta}\right)}$$