2.4 Summary

In this chapter, we learned about the sample complexity based on a generative model.

First we studied a naive approach. We estimate the the number of samples to control the accuracy of the approximate model with any policy, which leads to Prop 2.1.

[Prob. 2.1] Suppose $\varepsilon \le \frac{1}{1-\gamma}$, when #samples $= |S||A|N$ and

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

For all policies π,

$$\|Q^\pi - \hat{Q}^\pi\|_\infty \le \varepsilon. \quad \text{with prob.} \geq 1-\delta$$

Then we tried to understand how many samples do we need to estimate $\hat{Q}^*$ with policy $\hat{\pi}^*$, which leads to Prop. 2.4.

[Proposition 2.4] Let $\delta \geq 0$. with prob. $\geq 1-\delta$

$$\left\{ \begin{array}{l} \|Q^* - \hat{Q}^*\|_\infty, \\ \|Q^* - \hat{Q}^{\pi^*}\|_\infty \end{array} \right\} \le \Delta_{\delta,N}$$

where $\Delta_{\delta,N} = \frac{\gamma}{(1-\gamma)^2}\sqrt{\frac{2\log(2|S||A|/\delta)}{N}}$

Finally, we proved that model based algorithm is minmax optimal, which means that less error means larger samples, which is shown in Theorem. 2.8:

[Theorem 2.8.] for $\varepsilon \in (0, \varepsilon_0)$, $\delta \in (0, \delta_0)$, if algorithm A is (ε, δ)-good, then A must use a number of samples which is lower bounded as:

$$\#\text{ samples} \geq \frac{c}{(1-\gamma)^3} \frac{|S||A| \log(c|S||A|/\delta)}{\varepsilon^2}$$

note: (ε, δ)-good : $\|Q^* - \hat{Q}^*\|_\infty \leq \varepsilon$ prob. $\geq 1-\delta$.

Besides, with lemma 1.11, we have the amplification on $Q^{\hat{\pi}^*}$

[lemma 1.11] For any vector $Q \in \mathbb{R}^{|S||A|}$

$$V^{\pi_Q} \geq V^* - \frac{2\|Q - Q^*\|}{1-\gamma} \mathbb{1}$$

By regarding $\hat{Q}^*$ as Q, $\hat{\pi}^*$ as π^Q, we can use Lemma 1.11 to quantify $\|Q^* - Q^{\hat{\pi}^*}\|_\infty$.