یادگیری تقویتی دانیال پرنیان

Subject :

Year . Month . Date . ( ) تمرین ۱۱ ۴۰۱۱۱۰۳۰۷

(۱) ابتدا توجه کنید که می دانیم: $P_\pi(S_{t+1}) = \sum_{S_t, a_t} P_\pi(S_t) \times \pi(a_t|S_t) \times P(S_{t+1}|S_t, a_t)$

بنابراین می توان $\sum_{S_t} |P_{\pi_\theta}(S_t) - P_{\pi^*}(S_t)|$ را به صورت بازگشتی برحسب توزیع زمان $t-1$ و πها و transition Prob. نوشت:

$$\sum_{S_t} |P_{\pi_\theta}(S_t) - P_{\pi^*}(S_t)| = \sum_{S_t} \left| \sum_{S_{t-1}, a} \left[P_{\pi_\theta}(S_{t-1}) \times \pi_\theta(a|S_{t-1}) - P_{\pi^*}(S_{t-1}) \pi^*(a|S_{t-1}) \right] \times P(S_t|S_{t-1}, a) \right|$$

حالا یک $P_{\pi^*}(S_{t-1}) \times \pi_\theta(a|S_{t-1})$ اضافه و کم می کنیم و نامساوی مثلث می زنیم:

$$\sum_{S_t} |P_{\pi_\theta}(S_t) - P_{\pi^*}(S_t)| \leq \sum_{S_t} \left| \sum_{S_{t-1}, a} (P_{\pi_\theta}(S_{t-1}) - P_{\pi^*}(S_{t-1})) \pi_\theta(a|S_{t-1}) P(S_t|S_{t-1}, a) \right|$$
$$+ \sum_{S_t} \left| \sum_{S_{t-1}, a} P_{\pi^*}(S_{t-1}) (\pi_\theta(a|S_{t-1}) - \pi^*(a|S_{t-1})) P(S_t|S_{t-1}, a) \right|$$

عبارت اول را A و عبارت دوم را B در نظر بگیرید. نامساوی مثلث می زنیم:

$$A \leq \sum_{S_{t-1}} |P_{\pi_\theta}(S_{t-1}) - P_{\pi^*}(S_{t-1})| \underbrace{\sum_a \pi_\theta(a|S_{t-1})}_{=1} \underbrace{\sum_{S_t} P(S_t|S_{t-1}, a)}_{=1}$$
$$= \sum_{S_{t-1}} |P_{\pi_\theta}(S_{t-1}) - P_{\pi^*}(S_{t-1})|$$

$$B \leq \sum_{S_{t-1}, a} \left| P_{\pi^*}(S_{t-1}) (\pi_\theta(a|S_{t-1}) - \pi^*(a|S_{t-1})) \right| \underbrace{\sum_{S_t} P(S_t|S_{t-1}, a)}_{=1}$$
$$= \sum_{S_{t-1}} P_{\pi^*}(S_{t-1}) \sum_a |\pi_\theta(a|S_{t-1}) - \pi^*(a|S_{t-1})|$$

حالا توجه کنید $\pi^*(a|S_{t-1})$ به ازای تمام a ها ۰ و یکی از آن ۱ است.

بنابراین:

$$\sum_a |\pi_\theta(a|S_{t-1}) - \pi^*(a|S_{t-1})| = \left[\sum_{a \neq \pi^*(S_{t-1})} \pi_\theta(a|S_{t-1}) \right] + 1 - \pi(a = \pi^*(S_{t-1})|S_{t-1})$$
$$= 2\pi_\theta(a \neq \pi^*(s)|s) \rightsquigarrow B \leq \sum_{S_{t-1}} P_{\pi^*}(S_{t-1}) \times 2\pi_\theta(a \neq \pi^*(S_{t-1})|S_{t-1})$$
$$= 2 E_{P_{\pi^*}(S_{t-1})} [\pi_\theta(a \neq \pi^*(S_{t-1}))$$

PAPCO

Subject :
Year . Month . Date . ( )

الکنون به صورت بازگشتی می‌نویسیم :

$$\sum_{s_t} |P_\pi(s_t) - P_{\pi^*}(s_t)| \leq \sum_{s_{t-1}} |P_{\pi_\theta}(s_{t-1}) - P_{\pi^*}(s_{t-1})| + 2E_{P_{\pi^*}(s_{t-1})}[\pi_\theta(a \neq \pi^*_t(s$$

$$\leq \dots \leq 2\sum_{i=1}^{t-1} E_{P_\pi(s_i)}[\pi_\theta(a \neq \pi^*(s_i) | s_i)] \leq 2T\varepsilon$$

نامساوی پایانی هم از بخش ۱.۱ داریم (و توجه به اینکه $t-1 < T$)

پس حکم ثابت شد . ■

$$J(\pi) = \sum_{t=1}^{T} E_{P_\pi(s_t)}[r(s_t)] \xRightarrow{r(s_{t<T})=0} J(\pi) = E_{P_\pi(s_T)}[r(s_T)]$$ ۲

$$\Rightarrow J(\pi^*) - J(\pi_\theta) = \sum_{s_T} r(s_T) \times P_{\pi^*}(s_T) - \sum_{s_T} r(s_T) \times P_{\pi_\theta}(s_T)$$

$$\leq \sum_{s_t} r(s_T) \times |P_{\pi_\theta}(s_t) - P_{\pi^*}(s_t)| \leq R_{Max} \sum_{s_t} |P_{\pi_\theta}(s_t) - P_{\pi^*}(s_t)|$$

بخش قبل

$$\leq R_{Max} \times T\varepsilon \Rightarrow J(\pi^*) - J(\pi_\theta) = O(T\varepsilon)$$ ■

$$J(\pi^*) - J(\pi_\theta) = \sum_{t=1}^{T} E_{P_{\pi^*}}(r(s_t)) - \sum_{t=1}^{T} E_{P_{\pi_\theta}}(r(s_t))$$ ۳

از بخش قبل (با همان اثبات) می‌دانیم $R_{Max} T\varepsilon \geq E_{P_{\pi^*}}(r(s_t)) - E_{P_{\pi_\theta}}(r(s_t))$

$$\Rightarrow J(\pi^*) - J(\pi_\theta) \leq \sum_{t=1}^{T} R_{Max} T\varepsilon = T^2\varepsilon \times R_{Max}$$

$$\Rightarrow J(\pi^*) - J(\pi_\theta) = O(T^2\varepsilon)$$ ■