UC Berkeley CS285
homework 1: Imitation Learning

dhia naouali

November 7, 2025

1 Analysis

1. given

$$\frac{1}{T}\sum_{t=1}^{\mathrm{T}} E_{p_{\pi^*}(s_t)}\pi_\theta(a_t\pi^*(s_t)|s_t) \leq \varepsilon$$

$$\sum_{t=1}^{\mathrm{T}} E_{p_{\pi^*}(s_t)}\pi_\theta(a_t\pi^*(s_t)|s_t) \leq \varepsilon T$$

we introduce the event $\mathrm{M}_t : \{\pi_\theta(a_t \neq \pi^*(s_t)\}$
and by the union bound

$$\Pr\left(\bigcup_{k=1}^{t} \mathrm{M}_k\right) \leq \sum_{k=1}^{t} \mathrm{M}_k \leq \sum_{k=1}^{T} \mathrm{M}_k \leq \varepsilon T$$

$$|p_{\pi\theta}(s_t) - p_{\theta^*}(s_t)| = \Pr\left(\bigcup_{t} M_t\right)\ |p_{mistake}(s_t) - p_{\pi^*}(s_t)|$$

$$\sum_{s_t} |p_{\pi\theta}(s_t) - p_{\theta^*}(s_t)| = \Pr\left(\bigcup_{t} M_t\right)\ \sum_{s_t} |p_{\pi mistake}(s_t) - p_{\theta^*}(s_t)|$$

therefore

$$\sum_{s_t} |p_{\pi\theta}(s_t) - p_{\theta^*}(s_t)| \leq 2\varepsilon T$$

2. (a) Given a state-dependent reward $\mathrm{r}(s_t)$, where $|\mathrm{r}(s_t)| \leq \mathcal{R}_{max}$, we consider the total expected reward:

$$\mathcal{J}(\pi) = \sum_{t=1}^{T} E_{p_{\pi^*}(s_t)}\mathrm{r}(s_t)$$

$$\mathcal{J}(\pi^*)-\mathcal{J}(\pi_\theta) = \sum_{t=1}^{T} \Big(E_{p_{\pi^*}(s_t)}\mathrm{r}(s_t)-E_{p_{\pi_\theta}(s_t)}\mathrm{r}(s_t)\Big) = \sum_{t=1}^{T} E_{\,s_t}\big(p_{\pi^*}(s_t)-p_{\pi_\theta}(s_t)\big)\mathrm{r}(s_t)$$

since $\mathrm{r}(s_t) = 0$ for all $t < T$, we have:

$$\mathcal{J}(\pi^*) - \mathcal{J}(\pi_\theta) = \sum_{s_T} \big(p_{\pi^*}(s_T) - p_{\pi_\theta}(s_T)\big)\nabla(s_T).$$

thus,

$$\sum_{s_t} \big|p_{\pi^*}(s_t) - p_{\pi_\theta}(s_t)\big|\, \mathcal{R}_{\max} = \mathcal{O}(2\varepsilon T\mathcal{R}_{\max}) = \mathcal{O}(\varepsilon T).$$

(b) Given that

$$\mathcal{J}(\pi^*) - \mathcal{J}(\pi_\theta) = \sum_{t=1}^{T} \Big(E_{p_{\pi^*}(s_t)}\mathrm{r}(s_t) - E_{p_{\pi_\theta}(s_t)}\mathrm{r}(s_t)\Big),$$

where each term is $\mathcal{O}(2\varepsilon T\mathcal{R}_{\max})$, we get:

$$\mathcal{J}(\pi^*) - \mathcal{J}(\pi_\theta) = \mathcal{O}(2\varepsilon T^2\mathcal{R}_{\max}) = \mathcal{O}(\varepsilon T^2).$$