CS285 Homework 1

Jeremiah Chen

June 5, 2025

1 Analysis

1.1 Part 1

Using Hint 1, we have:

$$
\begin{aligned}
p_e(t) &= \Pr\left(\bigcup_{t=1}^{T} \pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)\right) \\
&\leq \sum_{t=1}^{T} \Pr\left(\pi_\theta(a_t \neq \pi^*(s_t) \mid s_t)\right) \\
&= \sum_{t=1}^{T} \mathbb{E}_{P_{\pi^*(s_t)}} \pi_\theta(a \neq \pi^*(s_t) \mid s_t) \\
&\leq \varepsilon T
\end{aligned}
$$

Since

$$
\begin{aligned}
p_{\pi_\theta}(s_t) &= (1 - p_e(t)) \cdot p_{\pi^*}(s_t) + p_e(t) \cdot p_{\text{mistake}}(s_t) \\
|p_{\pi_\theta}(s_t) - p_{\pi^*}(s_t)| &\leq p_e(t) \cdot |p_{\pi^*}(s_t) - p_{\text{mistake}}(s_t)|
\end{aligned}
$$

As a result:

$$
\begin{aligned}
\sum_{s_t} |p_{\pi_\theta}(s_t) - p_{\pi^*}(s_t)| &\leq \sum_{s_t} p_e(t) \cdot |p_{\pi^*}(s_t) - p_{\text{mistake}}(s_t)| \\
&\leq \max_t p_e(t) \sum_{s_t} (p_{\pi^*}(s_t) + p_{\text{mistake}}(s_t)) \\
&\leq 2\varepsilon T
\end{aligned}
$$

1.2 Part 2

Notice that

$$
\begin{aligned}
J(\pi^*) - J(\pi_\theta) &= \sum_t \mathbb{E}_{P_{\pi^*(s_t)}} r(s_t) - \sum_{t=1}^{T} \mathbb{E}_{P_{\pi_\theta}(s_t)} r(s_t) \\
&\leq \sum_t \sum_{s_t} |P_{\pi^*}(s_t) - P_{\pi_\theta}(s_t)| \cdot r(s_t) \\
&\leq 2\varepsilon T \sum_t r(s_t) \qquad \cdots
\end{aligned}
$$

Therefore, for the first case $\forall t < T, r(s_t) = 0$

$$
\begin{aligned}
2\varepsilon T \sum_t r(s_t) &\leq 2\varepsilon T \cdot R_{\max} \\
&= \mathcal{O}(T\varepsilon)
\end{aligned}
$$

For an arbitrary case:

$$
\begin{aligned}
2\varepsilon T \sum_t r(s_t) &\leq 2\varepsilon T^2 \cdot R_{\max} \\
&= \mathcal{O}(T^2\varepsilon)
\end{aligned}
$$

2 Editing Code

See the file `hw1.py` for the code implementation.

3 Behavioral Cloning

See the file under textttperformance/BC/ for the results of the behavioral cloning task.

4 DAgger

See the file under textttperformance/DAgger/ for the results of the DAgger task.