1) Imitation Learning can be viewed as interactive online learning.

\[ \text{Goal: Minimize Average Regret} \]

\[ \text{Avg Regret} (\pi_1 : \pi_N) = \frac{1}{N} \left( \sum_{i=1}^{N} l_i(\pi) - \min_{\pi \in \Pi} \sum_{i=1}^{N} l_i(\pi) \right) \]

No Regret Algorithm := \( \lim_{N \to 00} \text{Avg Regret}(\pi_1 : \pi_N) \to 0 \)

2) Follow the Leader

\[ \pi^i = \arg\min_{\pi \in \Pi - \pi^{i-1}} \sum_{i=0}^{\pi} l_i(\pi) \]
If $l_i(\pi)$ is strongly convex

then $\text{FTL}$ is no-regret

3. **DAGGER** is $\text{FTL}$.

$$\pi_{t+1} = \text{Train}(D)$$

$$= \arg\min_{\pi \in \Pi} \sum_{i=0}^{t} l_i(\pi)$$

4. **DAGGER** returns at least one policy $\pi_i$ that does well on its own induced distribution

$$\pi_i \text{ s.t. } \sum_{t=0}^{t_i} \mathbb{E} l(s_t, \pi) \leq O(T)$$

Assumption: [Rich Policy Class]

For any loss function $l(\pi)$, $\exists \pi \in \Pi$ that is good

$$\min_{\pi \in \Pi} l(\pi) = O(T H)$$

Proof: Look at the best policy that DAGGER returns.

$$\min_{i=1, \ldots, N} l_i(\pi_i)$$

$$\leq \frac{1}{N} \sum_{i=1}^{N} l_i(\pi_i)$$

$$\leq \frac{1}{N} \left( \sum_{i=1}^{N} l_i(\pi) - \min_{\pi \in \Pi} \sum_{i=1}^{N} l_i(\pi) \right) + \min_{\pi \in \Pi} \frac{1}{N} \sum_{i=1}^{N} l_i(\pi)$$
As $N \to \infty$, $\text{Avg}\,\text{Reg} \to 0$

$$\log N \quad \frac{1}{N}$$