5.4 Follow the leader

[Algorithm 2] follow the leader (FTL)

Let $f_1, \cdots, f_T$ be the sequence of loss functions played by nature.

The learner chooses $w_t \in S$ that min. $\sum_{i=1}^{t-1} f_i(w)$, i.e.

$$w_t \in \underset{w \in S}{\operatorname{argmin}} \sum_{i=1}^{t-1} f_i(w) \quad \cdots \quad (476)$$

Note: We can think FTL as an empirical risk minimizer where the training set is the first t-1 samples.

Lemma 7 (compare FTL with one-step lookahead cheater)

Let $f_1, \cdots, f_T$ be any sequence of loss function.

Let $w_1, \cdots, w_T$ be produced by FTL according to (476).

For any $u \in S$:

$$\text{Regret}(u) := \sum_{t=1}^{T} f_t(w_t) - f_t(u) \leq \sum_{t=1}^{T} [f_t(w_t) - f_t(w_{t+1})]$$

Pf: It suffices to show $\sum_{t=1}^{T} f_t(w_{t+1}) \leq \sum_{t=1}^{T} f_t(u) \quad \forall u \in S$.

By induction:

① $T=1$, $f_1(w_2) = \min_{u \in S} f_1(u) \leq f_1(u)$

② Assume the inductive hypothesis on $T-1$:

$$\sum_{t=1}^{T-1} f_t(w_{t+1}) \leq \sum_{t=1}^{T-1} f_t(u) \quad \text{for all } u \in S$$

Add $f_T(w_{T+1})$ to both sides:

$$\sum_{t=1}^{T} f_t(w_{t+1}) \leq \sum_{t=1}^{T-1} f_t(u) + f_T(w_{T+1}) \quad \text{for all } u \in S$$

In particular, set $u = w_{T+1}$ and we have

$$\sum_{t=1}^{T} f_t(w_{t+1}) \leq \sum_{t=1}^{T} f_t(w_{T+1}) = \min_{u \in S} \sum_{t=1}^{T} f_t(u)$$

which finishes the proof. □

Note: (i) RHS $f_t(w_t) - f_t(w_{t+1})$ measure the stability of Algorithm 2.

[Example 28] quadratic optimization: FTL works.

Assume nature always chooses quadratic functions:

$$f_t(w) = \frac{1}{2}\|w - z_t\|_2^2$$

where $\|z_t\|_2 \leq L$ for all $t = 1, \cdots, T$.

FTL (minimizing over $S = \mathbb{R}^d$) has a closed form solution, which is $w_t = \frac{1}{t-1}\sum_{i=1}^{t-1} z_i$. (by differential we have this)

Bounded one term of $f_t(w_t) - f_t(w_{t+1})$:

$$\begin{aligned} f_t(w_t) - f_t(w_{t+1}) &= \frac{1}{2}\|w_t - z_t\|_2^2 - \frac{1}{2}\|(1-\tfrac{1}{t})w_t + \tfrac{1}{t}z_t - z_t\|_2^2 \\ &= \frac{1}{2}[1-(1-\tfrac{1}{t})^2]\|w_t - z_t\|_2^2 \\ &\leq \frac{1}{t}\|w_t - z_t\|_2^2 \\ &\leq \frac{4L^2}{t} \end{aligned}$$

$4L^2$ = the square of diam($\|z\|_2$)

Summing: $\sum_{t=1}^{T}[f_t(w_t) - f_t(w_{t+1})] \leq 4L^2 \sum_{t=1}^{T}\frac{1}{t} \leq 4L^2(\log(T)+1)$

$$\begin{aligned} \sum_{t=1}^{T}\frac{1}{t} &= 1 + \frac{1}{2} + \cdots + \frac{1}{T} = 1 + \int_1^2 \frac{1}{2}dt + \cdots + \int_{T-1}^{T}\frac{1}{T}dt \\ &\leq 1 + \int_1^2 \frac{1}{t}dt + \cdots + \int_{T-1}^{T}\frac{1}{t}dt \\ &= 1 + \log(T) \end{aligned}$$

Note: $f_t(w_t) - f_t(w_{t+1}) \sim O(\frac{1}{t})$, which implies FTL for $\|\cdot\|_2^2$ is stable.

[Example 29] linear optimization: FTL fails.

Let $S = [-1, 1]$ be FTL's possible predictions (convex set)

Consider linear functions $f_t(w_t) = wz_t$ in $d=1$ dimention:

$$(z_1, z_2, \cdots) = (-0.5, 1, -1, 1, -1, 1, -1, \cdots)$$

The FTL give w_t like:

$$(w_1, w_2, \cdots) = (0, 1, -1, 1, -1, 1, -1, \cdots)$$

$$\Rightarrow \quad \sum_{t=1}^{T-1} f_t(w_t) = T-1$$

Expert $u=0$ obtains 0 cumulative loss.

$\Rightarrow$ Regret $\geq T-1$

Note:

(i) For $\|\cdot\|_2^2$, w_t and w_{t+1} must get closer

(ii) For linear functions, w_t and w_{t+1} vary largely.

(iii) It seems FTL only works when f_t offers 'stability'.