HW2. Problem:

$\|x_i\|_\infty \leq X_\infty$

* Thm: If A is OARO-stable w/ rate g, then

$$\mathop{\mathbb{E}}_{S\sim D^m}\left[L_D(A(S)) - L_S(A(S))\right] \leq g(m).$$

* l_2-regularization gives stability:

Assume: ① $\ell(w,z)$ is ρ-Lipschitz wrt w, for any z

② $\ell(w,z)$ is convex wrt w, for any z.

③ $\hat{w} = A(S) = \underset{w\in\mathbb{R}^d}{\operatorname{argmin}}\left(\frac{\lambda}{2}\|w\|_2^2 + L_S(w)\right)$ ($F_S(w)$)

①②③ $\Rightarrow$ A is $g(m) = \frac{2\rho^2}{\lambda m}$-OARO stable.

Need to find λ. such that it minimizes $\in\mathbb{R}^d$

$$\mathbb{E}_{S\sim D^m} L_D(A(S)) = \mathbb{E}_{S\sim D^m} L_S(A(S)) + \mathbb{E}_{S\sim D^m}\left[L_D(A(S)) - L_S(A(S))\right]$$

$\overset{?}{\leq} L_D(w^*) + \frac{\lambda}{2}\|w^*\|^2$

$\mathbb{E}_{S\sim D^m} L_S(A(S))$: can increase when $\lambda\uparrow$.

$\mathbb{E}_{S\sim D^m}\left[L_D(A(S)) - L_S(A(S))\right] \leq \frac{2\rho^2}{\lambda m}\downarrow$ with $\lambda\uparrow$

$\hat{w}$ (output of A).

optimality of [illegible]

$$L_S(A(S)) \le F_S(A(S)) \le F_S(w^*)$$

taking expectation:

$$\mathbb{E}_S L_S(A(S)) \le \mathbb{E}_S[F_S(w^*)]$$

$$= \mathbb{E}_{S\sim D^m}\left[\frac{\lambda}{2}\|w^*\|^2 + L_S(w^*)\right]$$

$L_S(w^*) = \frac{1}{m}\sum_{i=1}^{m} \ell(w^*, z_i)$

$$= \frac{\lambda}{2}\|w^*\|^2 + L_D(w^*)$$

In summary:

$$\forall w^*, \quad \mathbb{E}_{S\sim D^m} L_D(A(S)) \le L_D(w^*) + \frac{\lambda}{2}\|w^*\|^2 + \frac{2\rho^2}{\lambda m}$$

$\lambda m >? 1$

Interpretations:

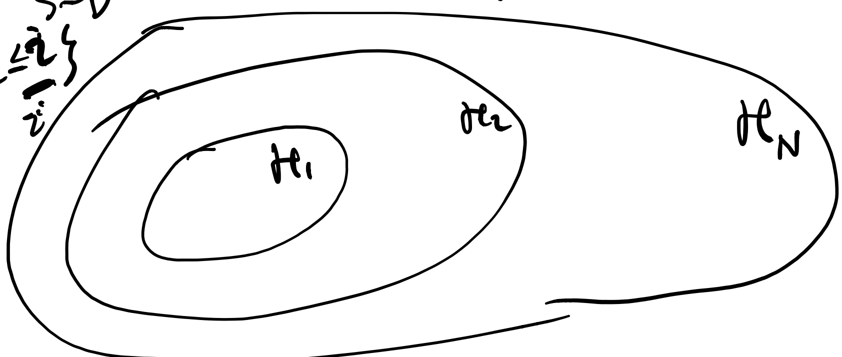
(1) $\lambda = \frac{1}{\sqrt{m}}$. (or any other fn of m $\frac{1}{m} \ll \lambda \ll 1$)

$$\frac{1}{\sqrt{m}}\|w^*\|^2 + \frac{\rho^2}{\sqrt{m}}$$

$$\mathbb{E}_{S\sim D^m} L_D(A(S)) \le \min_{w^*}\left(L_D(w^*) + \frac{\|w^*\|^2 + \rho^2}{\sqrt{m}}\right).$$

$i \in \mathbb{N}$..

$\mathcal{H}_i = \{w : \|w\|_2 \le i\}$



$$\frac{B^2+\rho^2}{\sqrt{m}} \ge \frac{B\cdot\rho}{\sqrt{m}}$$

This is a model selection result.

② Fix hypothesis class

$$\mathcal{H} = \{ w \in \mathbb{R}^d : \|w\|_2 \le B \}.$$

$$\mathbb{E}_S\, L_D(A(S)) \le \min_{w \in \mathcal{H}} L_D(w) + \underbrace{\frac{\lambda}{2} B^2 + \frac{2\rho^2}{\lambda m}}_{2 \cdot \rho \cdot B \cdot \sqrt{\frac{1}{m}}}.$$

$$\lambda^* = 2\frac{\rho}{B}\sqrt{\frac{1}{m}}.$$

This is a PAC-like guarantee, except that the guarantee is in expectation. not in high probability.

$$m \ge \frac{4\rho^2 B^2}{\varepsilon^2} \Rightarrow \mathbb{E}_S\left[L_D(A(S)) \right] - \min_{w \in \mathcal{H}} L_D(w) \le \varepsilon.$$

Online learning:

Ex: 1. spam detection

At each time step $t = 1, 2, \cdots T$

$\Omega = \{ w : \|w\|_2 \le B \}$

— receive an email $x_t \in \mathcal{X}$

— predict $\hat{y}_t \in \{-1, +1\}$ (−1: nonspam, +1: spam)

— see $y_t \in \{-1, +1\}$. $f_t(w) = I(y_t \langle w, x_t \rangle \le 0)$. $f_w(x_t)$

minimize $\sum_{t=1}^{T} I(\hat{y}_t \ne y_t)$ $= \ln\left(1 + e^{-y_t \langle w, x_t \rangle}\right)$

online classification. [regression.

2. sequential investment (portfolio selection).

W_1 = initial capital.

For $t = 1, 2 \cdots T$

- decide $p_t \in \Delta^{N-1} = \{ p \in \mathbb{R}_+^N : \sum_i p(i) = 1 \}$ $\subseteq \Omega$

  for asset $i \in \{1 \cdots N\}$ investing $W_t \cdot p_t(i)$

- observe relative prices $r_t \in \mathbb{R}_+^N$.

$r_t(i) > 1$ ↑

$= 1$ –

< 1 ↓

maximizing $\prod_{t=1}^{T} \langle p_t, r_t \rangle$

$f_t(p) = -\ln(\langle p, r_t \rangle)$.

asset i: $W_t \cdot p_t(i) \cdot r_t(i)$

$$W_{t+1} = \sum_i W_t \, p_t(i) \, r_t(i) = W_t \cdot \langle p_t, r_t \rangle.$$

Goal: maximize W_{T+1}.

3. aggregating weather prediction:

For each day $t = 1, 2 \cdots T$

- obtain weather temperature predictions from N experts / model.
- make final prediction by following an expert drawn from $p_t \in \Delta^{N-1}$ $\subseteq \Omega$
- observe the losses of each model

$$l_t \in [0,1]^N.$$

$f_t(p) = \langle p, l_t \rangle$.

Goal: minimize $\sum_t \langle p_t, \ell_t \rangle$

4. production recommendation (multi-armed bandits)

For $t = 1, 2, \cdots T$

— recommend $a_t \in \{1 \cdots K\}$ (product) to a new customer

— observe loss of a_t: $\ell_t(a_t)$ (e.g. clicked $\Rightarrow 0$, not clicked $\Rightarrow 1$)

$\ell_t \in [0,1]^K$

$\ell_t(i)$

bandit feedback

full-info feedback.

② $p_t \in \Delta^{K-1}$

① $= \Omega$

MAB.

$f_t(a) = \ell_t(a)$

$f_t(p) = \langle p, \ell_t \rangle$

$(a_t, \ell_t(a_t)) \rightarrow \langle p_t, \ell_t \rangle$

Goal: minimize $\sum_t \ell_t(a_t)$

5. personalized product recommendation:

given N policies: $\pi^1 \cdots \pi^N$ each $\pi^i : \mathcal{X} \to \{1 \cdots K\}$.

(take $N = K$. $\pi^i(x) \equiv i$)

$\Rightarrow$ MAB.

For $t = 1, 2, \cdots T$

— observe contextual info of customer x_t.

— random select on of N policies $\sim w_t \in \Delta^{N-1}$.

selected policy π_t.

recommend $\pi_t(x_t)$

— observe loss $\ell_t(\pi_t(x_t))$ (we don't observe other entries of $\ell_t \in [0,1]^K$)

$\Omega = \Delta^{N-1}$

$f_t(w) = \mathbb{E}_{i \sim w} \ell_t(\pi^i(x_t)) = \sum_{i=1}^N w_i \ell_t(\pi^i(x_t))$

contextual bandits.

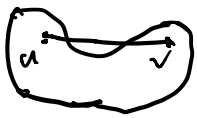
Online (convex) optimization:

- decision set Ω (action space) often cvx

(Ω is a convex set : $\forall u, v \in \Omega \;.\; \forall \alpha \in (0,1).$

$\alpha u + (1-\alpha) v \in \Omega$.

(cvx).



(non cvx)

For $t = 1, 2 \cdots T$.

- learner picks $w_t \in \Omega$
- environment picks loss fn $f_t : \Omega \to \mathbb{R}$.
- learner suffers loss $f_t(w_t)$,

  & observes information on f_t.

performance measure : $\sum_{t=1}^{T} f_t(w_t)$.

setting of OL:

- stochastic : $\{f_t\}_{t=1}^{T} \overset{iid}{\sim} D$.
- oblivious adversary : $\{f_t\}_{t=1}^{T}$ chosen ahead of time.
- adaptive adversary : $\forall t.$ f_t can depend on $\{w_s\}_{s=1}^{t}$.

categorized by feedback model:

- full info : see f_t (or $\nabla f_t(w_t)$).

- bandit : $f_t(w_t) \in \mathbb{R}$
- other feedback settings.

key performance measure :

regret

$$R_T = \sum_{t=1}^{T} f_t(w_t) - \min_{w \in \mathcal{W}} \sum_{t=1}^{T} f_t(w) .$$

(want $= o(T)$)

Next class :

Online to batch conversion.

small R_T $\rightarrow$ get stat. learning alg. w/ small excess loss.

Next Thursday (Mar. 25) : pre-recorded lecture.