

- See EdStem for more info about Peer Evaluation

Last time:

- Application of Hedge to Online Convex Opt.
- Lower bound on regret for any alg. in Online Alloc.
- Start Online Concept Learning
a.k.a. Online Mistake-Bound Learning

Online concept learning (a.k.a. mistake-bound model)

- Concept class $\mathcal{C}$ of $\{0,1\}$ -valued functions over input space $\mathcal{X}$

For $t=1,2,\dots$

Learner gets input $x_t \in \mathcal{X}$

Learner predicts $a_t \in \{0,1\}$

Learner gets true label $y_t \in \{0,1\}$

$\mathcal{C}$ -realizability: $\forall t, \exists c \in \mathcal{C}$ s.t. $c(x_s) = y_s \quad \forall s \leq t$
(there is always a concept $c \in \mathcal{C}$ consistent with all labeled examples (x_s, y_s) seen so far.)

Experts scenario: $\mathcal{C}_{\text{expert}} = \{c_i \mid i \in [N]\}$
 $\mathcal{X} = \{0,1\}^N$
 $c_i(x) = x_i$

Halving w/ $\mathcal{C}$: mistake bound $\log_2 |\mathcal{C}|$
but runtime per round is $\Omega(|\mathcal{C}|)$

Boolean function concepts

$$x_1 = \begin{cases} 1 & \text{if clouds} \\ 0 & \text{if no clouds} \end{cases}$$
$$x_2 = \begin{cases} 1 & \text{if birds} \\ 0 & \text{if no birds} \end{cases}$$

Input space: $X = \{0, 1\}^n$ (Boolean n -cube)

False $\uparrow$ $\uparrow$ True

Monotone conjunctions $MCONJ_n = \{c_S \mid S \subseteq [n]\}$
where $c_S(x) = \bigwedge_{i \in S} x_i$

$c_\emptyset(x) = 1 \quad \forall x$

$$c_{[n]}(x) = \begin{cases} 1 & \text{if } x = (1, 1, \dots, 1) \\ 0 & \text{otherwise} \end{cases}$$

$$|MCONJ_n| = 2^n$$

Halving alg: mistake bound $\log_2 2^n = n$

Attribute Elimination (Valiant, Angluin)

Init: $S_1 = [n]$ $\leftarrow$ most stringent S

In round t :

- Get $x_t = (x_{t,1}, \dots, x_{t,n}) \in \{0, 1\}^n$
- Predict $a_t = c_{S_t}(x_t)$
- Get $y_t \in \{0, 1\}$
- Update

$$S_{t+1} = \begin{cases} S_t & \text{if } a_t = y_t \\ S_t \setminus \{i \in [n] \mid x_{t,i} = 0\} & \text{if } a_t \neq y_t \end{cases}$$

relaxing the current hypothesis

Say c_{S^*} is a monotone conj consistent w/ all examples so far
Only kind of mistake is a "false negative"
(we say 0 but true outcome is 1)

$$x_{t,i} = 0 \Rightarrow i \notin S^*$$

If AE makes a mistake, we remove something from S_t

$M_t = \#$ mistakes after t rounds

Claim: $\forall t$,

$$S^* \subseteq S_t$$

$$M_{t-1} \leq |S_1| - |S_t|$$

Proof:

Say in round t , $a_t \neq y_t$

$$a_t = \bigwedge_{i \in S_t} x_{t,i} = \underbrace{\bigwedge_{i \in S^*} x_{t,i}}_{= y_t} \wedge \underbrace{\bigwedge_{i \in S_t \setminus S^*} x_{t,i}}_{= 0}$$

$$a_t = 0$$

$$S_t \setminus S^* \neq \emptyset$$

Some $i \in S_t \setminus S^*$ has $x_{t,i} = 0$

i.e., Any i removed from S_t is not in S^*

$$\text{So } S^* \subseteq S_{t+1}$$

$$|S_1| - |S_t| \geq M_{t-1}$$

$$|S_t| - |S_{t+1}| \geq \mathbb{1}_{\{\text{mistake in round } t\}}$$

$$\Rightarrow |S_1| - |S_{t+1}| \geq M_t$$

Conjunctions $CONJ_n = \{C_{S,T} \mid S \subseteq [n], T \subseteq [n], S \cap T = \emptyset\}$

$$C_{S,T}(x) = \bigwedge_{i \in S} x_i \wedge \bigwedge_{i \in T} (\neg x_i)$$

$\nwarrow$ NOT x_i

Extend every input $x = (x_1, \dots, x_n) \in \{0,1\}^n$
 to $2n$ Boolean attributes
 $\tilde{x} = (\tilde{x}_1, \dots, \tilde{x}_n, \tilde{x}_{n+1}, \tilde{x}_{n+2}, \dots, \tilde{x}_{2n})$

$$\tilde{x}_i = x_i \quad \tilde{x}_{n+i} = \neg x_i \quad \forall i \in [n]$$

$$\underline{C_{S,T}(x)} = \bigwedge_{i \in S} \tilde{x}_i \wedge \bigwedge_{i \in T} \neg \tilde{x}_i = \bigwedge_{i \in S \cup (n+T)} \tilde{x}_i$$

$\downarrow$ $\tilde{x}_i$ $\downarrow$ $\neg \tilde{x}_{n+i}$

$$= C_{S \cup (n+T)}(\tilde{x})$$

$\uparrow$

monotone conj over $2n$ variables

So can use AE for $MCONJ_{2n}$ to learn $CONJ_n$

Mistake bound $\leq 2n$

$\uparrow$
 under $CONJ_n$ -realizability

MDISJ_n monotone disjunctions

$$d_S(x) = \bigvee_{i \in S} x_i$$

De Morgan's law ✓

DISJ_n disjunctions

$$d_{S,T}(x) = \bigvee_{i \in S} x_i \vee \bigvee_{i \in T} \neg x_i$$

$$|\text{CONJ}_n| = 3^n$$

Linear threshold functions (LTFs)

$$\text{LTF}_n = \left\{ f: \{0,1\}^n \rightarrow \{0,1\} \mid \exists w \in \mathbb{R}^n, \theta \in \mathbb{R} \text{ s.t. } f(x) = \begin{cases} 1 & \text{if } \langle w, x \rangle > \theta \\ 0 & \text{o.w.} \end{cases} \right\}$$

$\forall x \in \{0,1\}^n$

w weight vector
 θ threshold.

(w, θ) gives a representation for f iff

$$\text{LTF}_n \subseteq \text{ALL}_n$$

$$|\text{LTF}_n| < +\infty$$

$$|\text{ALL}_n| = 2^{2^n} < +\infty$$

$$\log |\text{LTF}_n| = \Theta(n^2)$$

We'll see the U.B.
using "Sauer's lemma"

$$(1 - o(1)) n^2$$

Also turns out every $(M) \text{CONJ}_n$
 $(M) \text{DISJ}_n$

can be written by a LTF

$$\text{Halving alg: } \log |\text{LTF}_n| = \Theta(n^2)$$

Open: $\exists$ poly time algorithm w/ $MB \leq O(n^2)$?

Best ^{known} poly time alg. has $MB \ O(n^2 \log n)$
— Ellipsoid algorithm

Perceptron

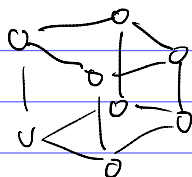
LTF's over $\mathbb{R}^n$

$$\text{LTF}_{\mathbb{R}^n} = \left\{ f: \mathbb{R}^n \rightarrow \{0,1\} \mid \exists (w, \theta) \in \mathbb{R}^n \times \mathbb{R} \text{ s.t. } f(x) = \mathbb{1}\{ \langle w, x \rangle > \theta \} \forall x \right\}$$

Can use algo designed for $\text{LTF}_{\mathbb{R}^n}$

for learning LTF_n

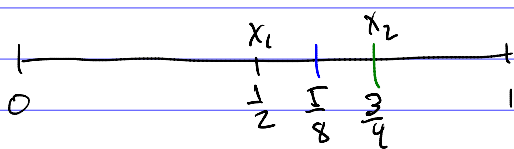
Embed $\{0,1\}^n$ into $\mathbb{R}^n$



Perceptron d/n have a finite m.B.
(even under $\text{LTF}_{\mathbb{R}^n}$ -realizability)

Theorem Even under $\text{LTF}_{\mathbb{R}^n}$ -realizability,
every alg can be forced to make ∞ -many mistakes

PF idea $c^*(x) = \mathbb{1}\{x_1 > \theta_t\}$ $\theta_t \in \mathbb{R}$



$$a_1 = 1 \quad y_1 = 0$$

$$a_2 = 0 \quad y_2 = 1$$

Even though learner can make progress
in the "length" of the "unknown region"

adversary can still force a mistake

γ -Margin assumption

$$\bullet \|x_t\| \leq 1 \quad \forall t$$

$$\bullet \exists (w^*, \theta^*) \text{ w/ } \|w^*\| \leq 1, |\theta^*| \leq 1$$

$$y_t = 1 \quad \langle w^*, x_t \rangle \geq \theta^* + \gamma$$

$$y_t = 0 \quad \langle w^*, x_t \rangle \leq \theta^* - \gamma$$

} $\forall t$

This assumption rules out the adversary from