

PS1 due tonight

Peer evaluation distributed Thursday (or come to my office after that)

PS2 out on Thursday

Last time: Online concept learning w/ realizability

Concept classes over $\{0,1\}^n$

- Monotone Conjunctions

- Halving: Mistake bound n , but run time $\Omega(2^n)$

- Attribute Elimination

- Mistake bound: n

- Run time per round: $O(n)$

- Conjunctions

- Monotone disjunctions

- Disjunctions

- Linear Threshold Functions

- Halving: Mistake bound n^2 , but run time $2^{\Omega(n^2)}$

- Ellipsoid-type alg: Mistake bound $O(n^2 \log(n))$
poly(n) run time per round

Concept classes over $\mathbb{R}^n$

- $\text{LTF}_{\mathbb{R}^n} = \{f: \mathbb{R}^n \rightarrow \{0,1\} \mid \exists w \in \mathbb{R}^n, \theta \in \mathbb{R} \text{ s.t.}$

$$f(x) = \mathbb{1}\{\langle w, x \rangle > \theta\} \quad \forall x \in \mathbb{R}^n\}$$

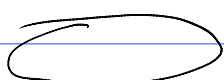
- No alg has finite mistake bound under $\text{LTF}_{\mathbb{R}^n}$ -realizability

Two simplifications:

$$\tilde{x} = (x, -1)$$

$$\tilde{w} = (w_1, \dots, w_n, w_{n+1}) \in \mathbb{R}^{n+1}$$

$$\begin{aligned} \langle \tilde{w}, \tilde{x} \rangle &= \sum_{i=1}^{n+1} \tilde{w}_i \tilde{x}_i \\ &= \sum_{i=1}^n \tilde{w}_i x_i - \tilde{w}_{n+1} \end{aligned}$$



- Lift every LTF in $\mathbb{R}^n$ to a **homogeneous LTF** (i.e. $b=0$) in $\mathbb{R}^{n+1}$
 - by mapping $x \in \mathbb{R}^n$ to $(x, -1) \in \mathbb{R}^{n+1}$
- Map $y \in \{0, 1\}$ to $2y - 1 \in \{-1, 1\}$

$$\begin{aligned} 0 &\rightarrow -1 \\ 1 &\rightarrow 1 \end{aligned}$$

Homogeneous γ -margin assumption in $\mathbb{R}^n$ ($\gamma > 0$)

- $\forall t, \|x_t\| \leq 1$
- $\exists w^* \in \mathbb{R}^n$ s.t. $\|w^*\| \leq 1$ and

$$\forall t, \begin{cases} y_t = 1 \Rightarrow \langle w^*, x_t \rangle \geq \gamma \\ y_t = -1 \Rightarrow \langle w^*, x_t \rangle \leq -\gamma \end{cases}$$

$$y_t \langle w^*, x_t \rangle \geq \gamma$$

Perceptron alg. for HLTF $_{\mathbb{R}^n}$

Init: $w_1 = 0 \in \mathbb{R}^n$

In round t :

- Get $x_t \in \mathbb{R}^n$
- Predict: $a_t = \begin{cases} 1 & \text{if } \langle w_t, x_t \rangle > 0 \\ -1 & \text{ow.} \end{cases}$

- Get $y_t \in \{-1, 1\}$

- Update: if $a_t \neq y_t$

$$w_{t+1} = w_t + y_t x_t$$

$$\text{if } a_t = y_t$$

$$w_{t+1} = w_t$$

$$y_t = 1$$

$$\begin{aligned} \langle w_{t+1}, x_t \rangle &= \langle w_t + x_t, x_t \rangle \\ &= \underbrace{\langle w_t, x_t \rangle}_{\geq 0} + \|x_t\|^2 \end{aligned}$$

Theorem In online concept learning w/ homogeneous γ -margin assumption, Perceptron makes at most $\frac{1}{\gamma^2}$ mistakes

Pf: If there's a mistake in round t ,

$$y_t \langle w_t, x_t \rangle \leq 0$$

Want to track $\langle w^*, w_t \rangle$ & $\|w_t\|$

Initially, $w_1 = 0$, so $\langle w^*, w_1 \rangle = 0$
 $\|w_1\| = 0$

Suppose mistake in round t .

$$w_{t+1} = w_t + y_t x_t$$

$$\begin{aligned} \langle w^*, w_{t+1} \rangle &= \langle w^*, w_t + y_t x_t \rangle \\ &= \langle w^*, w_t \rangle + y_t \underbrace{\langle w^*, x_t \rangle}_{\geq \gamma} \end{aligned}$$

$$\geq \langle w^*, w_t \rangle + \gamma$$

So if M_t mistakes after t rounds,

$$\langle w^*, w_{t+1} \rangle \geq M_t \cdot \gamma$$

$$\begin{aligned} \|w_{t+1}\|^2 &= \|w_t + y_t x_t\|^2 \\ &= \|w_t\|^2 + 2 \underbrace{\langle w_t, y_t x_t \rangle}_{\leq 0} + \underbrace{\|y_t x_t\|^2}_{\leq 1} \\ &\leq \|w_t\|^2 + 1 \end{aligned}$$

So

$$\|w_{t+1}\| \leq \sqrt{M_t}$$

$$1 \geq \cos(\text{angle bet. } w^*, w_{t+1}) = \frac{\langle w^*, w_{t+1} \rangle}{\|w^*\| \|w_{t+1}\|} \geq \frac{M_t \cdot \gamma}{\|w^*\| \sqrt{M_t}} = \frac{\sqrt{M_t} \cdot \gamma}{\|w^*\|} \geq \sqrt{M_t} \cdot \gamma$$

$$1 \geq \sqrt{M_t} \cdot \gamma$$

$$M_t \leq \frac{1}{\gamma^2} \quad \square$$

$$\|w_{t+1} - w^*\|^2 = \|w_t - w^*\|^2$$

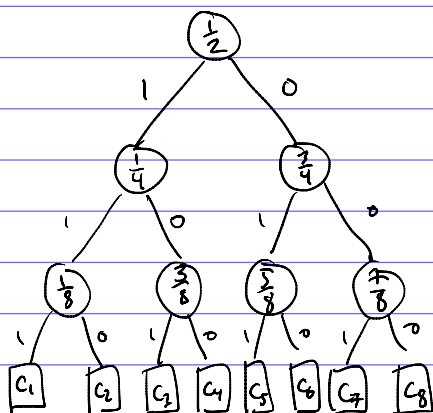
Winnow - multiplicative update alg. for LTF $_{\mathbb{R}^n}$

OPTIMAL MISTAKE BOUNDS

(Littlesstone 88)

(Complete) mistake tree of depth d for concept class $\mathcal{C}$

- Rooted complete binary tree of depth d
 - 2^d leaves
- Annotate every non-leaf node w/ an input $x \in \mathcal{X}$
- Every edge is either 0-edge or 1-edge
- Annotate every leaf w/ a concept $c \in \mathcal{C}$



$$c_i(x) = \mathbb{1}\{x > \frac{i-1}{8}\}$$

$$\mathcal{X} = \mathbb{R}$$

Thm If $\mathcal{C}$ has a mistake tree of depth d ,
 then in online concept learning w/ $\mathcal{C}$ -realizability,
 every learner can be forced to make d mistakes.

$Ldim(\mathcal{C}) =$ largest int. d s.t. $\exists$ mistake tree of depth d for $\mathcal{C}$
 (or ∞ if arbitrarily deep mistake trees exist for $\mathcal{C}$)

$$Ldim(\emptyset) = -1$$

- If $\mathcal{C}' \subseteq \mathcal{C}$, $Ldim(\mathcal{C}') \leq Ldim(\mathcal{C})$
- $Ldim(\mathcal{C}) = 0$ iff $|\mathcal{C}| = 1$
- If $\exists$ alg w/ MB M for $\mathcal{C}$, then $Ldim(\mathcal{C}) \leq M$.

$$Ldim(MCONJ_n) \leq n$$

$$Ldim(MDISJ_n) \leq n$$

$$Ldim(CONJ_n) \leq n+1$$

$$Ldim(\mathcal{C}) \leq \lfloor \log_2 |\mathcal{C}| \rfloor$$

$$Ldim(LTF_n) \leq n^2$$

$$Ldim(LTF_{\mathbb{R}^n}) = \infty$$

Standard Optimal Algorithm (SOA) for $\mathcal{C}$

• Init: $V_1 = \mathcal{C}$

• On round t :

- Get x_t

- $V_t^y = \{c \in V_t \mid c(x_t) = y\} \quad \forall y \in \{0, 1\}$

- If $\text{Ldim}(V_t^1) > \text{Ldim}(V_t^0)$, Halving $\log |V_t^1| > \log |V_t^0|$

$a_t = 1$

else

$a_t = 0$

- Get y_t

- Update $V_{t+1} = V_t^{y_t}$

Thm SOA makes $\leq \text{Ldim}(\mathcal{C})$ mistakes

Pf V_t is never empty (by realizability)

$\text{Ldim}(V_t) \geq 0$

If $\text{Ldim}(V_t) = 0$, then no more mistakes.

Claim: mistake $\rightarrow$ Ldim reduces by ≥ 1

Suppose mistake in round t ($a_t \neq x_t$)

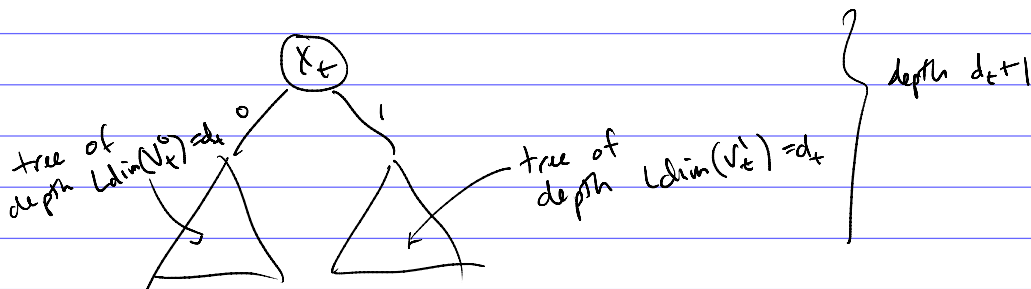
WTS: $\text{Ldim}(V_{t+1}) \leq \text{Ldim}(V_t) - 1$

Suppose for contradiction,

$$\text{Ldim}(V_{t+1}) = \text{Ldim}(V_t) = d_t$$

$$\text{Ldim}(V_{t+1}) = \text{Ldim}(V_t^{y_t}) \leq \text{Ldim}(V_t^{a_t}) \leq \text{Ldim}(V_t) = \text{Ldim}(V_{t+1})$$

$$\Rightarrow \text{Ldim}(V_t^0) = \text{Ldim}(V_t^1)$$



this is a mistake tree for V_t
of depth $d_t + 1$
 $\rightarrow$ contradiction.