CSCI4230 Computational Learning Theory — Spring 2023

Lecturer: Siu On Chan — *Based on Rocco Servedio's notes*

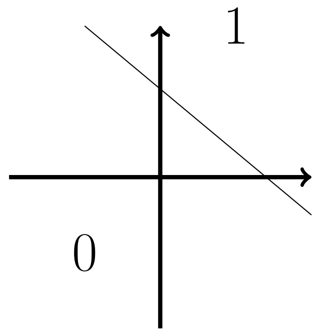
Notes 3: Winnow algorithms

1. Linear threshold functions (LTF)

Let $w \cdot x \overset{\text{def}}{=} \sum_{1 \leqslant i \leqslant n} w_i x_i$ (inner product between $w \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$)
An **LTF** $f : \mathbb{R}^n \to \{0, 1\}$ has the form

$$f(x) = \begin{cases} 1 & \text{if} \quad w \cdot x \geqslant \theta \\ 0 & \text{otherwise} \end{cases}$$

for some weight vector $w \in \mathbb{R}^n$ and threshold $\theta \in \mathbb{R}$



Every disjunction is LTF, e.g. for $x \in \{0, 1\}^n$

$$x_1 \vee x_2 \vee \overline{x}_3 \text{ true} \iff x_1 + x_2 + (1 - x_3) \geqslant 1 \iff x_1 + x_2 - x_3 \geqslant 0$$

Every 1-DL is LTF (why?)

---

2. Winnow1

Update weights **multiplicatively**
Learn **k-sparse** (i.e. involves k literals) monotone disjunctions using LTF hypothesis
$O(k \log n)$ mistakes
When k really small (e.g. 5) and n really big, $O(k \log n)$ is better than n (in Elimination Algorithm)

Winnow1

Initialize: $w_1 = \cdots = w_n = 1$, θ fixed to be n
On input x, output hypothesis $h(x) = \mathbb{1}(w \cdot x \geqslant \theta)$ and get $c(x)$
False positive ($h(x) = 1, c(x) = 0$): For every i s.t. $x_i = 1$
    Set $w_i = 0$ (demotion, in fact elimination)
False negative ($h(x) = 0, c(x) = 1$): For every i s.t. $x_i = 1$
    Double w_i (promotion)

In fact non-zero w_i is always $1, 2, 4, 8, \ldots$ (power of 2)
Observation: no w_i is ever negative
Observation: in every promotion step, some x_i in c has its w_i doubled

---

Claim: Each w_i always $< 2n$
Reason: When w_i is doubled, x_i must be 1 and $w \cdot x < n$
Claim: #promotion steps $\leqslant k \log(2n)$
Reason: No x_i in c is ever eliminated, and is promoted $\leqslant \log(2n)$ times (k many such x_i)

Lemma 1. *#elimination steps $\leqslant$ #promotion steps $+ 1$*

Proof. Let W = total weight $= \sum_{1 \leqslant i \leqslant n} w_i$ (initially n)
In each elimination step W decreases by $w \cdot x \geqslant n$ (w_i becomes 0 iff $x_i = 1$)
In each promotion step W increases by $w \cdot x < n$ (w_i doubled iff $x_i = 1$)
After e elimination steps and p promotion steps, $0 \leqslant W \leqslant n - en + pn$, so $e \leqslant p + 1$. □

Winnow1 makes $\leqslant 2k\log(2n)+1 = O(k\log n)$ mistakes on k-sparse monotone disjunction
Variation: During promotion, instead of doubling w_i, can multiply w_i with $\alpha > 1$; Threshold θ need not be n; See Littlestone if interested

---

Can Winnow1 learn non-monotone disjunction? (False positive kills it e.g. $c(x) = \overline{x}_1, x^1 = 11$)
Or LTF with nonnegative weights? (Not without new ideas such as Winnow2)

---

3. Winnow2

Can assume threshold $\theta = 1$ (by rescaling w)
An LTF $x \in \{0,1\}^n \mapsto \mathbb{1}(w\cdot x \geqslant 1)$ is **δ-separated** if

$$\forall x \in \{0,1\}^n, \quad \text{either } w\cdot x \geqslant 1 \text{ or } w\cdot x \leqslant 1-\delta$$

e.g. r-out-of-k threshold function

$$c(x) = \mathbb{1}(x_{i_1} + \cdots + x_{i_k} \geqslant r) = \mathbb{1}\left(\frac{1}{r}x_{i_1} + \cdots + \frac{1}{r}x_{i_k} \geqslant 1\right)$$

is $1/r$-separated

Winnow2

Initialize: $w_1 = \cdots = w_n = 1$, θ fixed to be n, α fixed to be $1+\delta/2$
On input x, output hypothesis $h(x) = \mathbb{1}(w\cdot x \geqslant \theta)$ and get $c(x)$
False positive $(h(x)=1, c(x)=0)$: For every i s.t. $x_i = 1$
  Divide w_i by α (demotion)
False negative $(h(x)=0, c(x)=1)$: For every i s.t. $x_i = 1$
  Multiply w_i by α (promotion)

Claim 2. *Winnow2 can learn δ-separated LTF with nonnegative weights $w \in \mathbb{R}^n$ with $O((\log n)\delta^{-2}\sum_{1\leqslant i\leqslant n} w_i)$ mistakes*

Proof in Littlestone §5

k-sparse monotone disjunctions are 1-out-of-k threshold functions
Winnow2 learns k-sparse monotone disjunctions with $O(k\log n)$ mistakes (direct proof in Blum §3.2)