COMS 6998-2 Advanced Topics in Computational Learning Theory **Spring 2005**

Lecture 5 : February 17, 2005

Lecturer: Rocco A. Servedio *Scribe: Jian Huang, Pinxing Ye*

1 Today's Lecture

In this lecture we cover the following:

- Learning intersections of halfspaces;
- Upper bound for learning intersections of halfspaces via PTFs;
- Lower bound for learning intersections of halfspaces via PTFs.

2 Learning intersections of halfspaces

We are going to look at weight-W halfspaces. Let $a(x_1, \ldots, x_n) = \text{sign}(\mu_1 x_1 + \cdots + \mu_n x_n - \theta)$, and $b(x_1, \ldots, x_n) = \text{sign}(\nu_1 x_1 + \cdots + \nu_n x_n - \tau)$, with $\sum_{i=1}^n |\mu_i| \leq W$, and $\sum_{i=1}^n |\nu_i| \leq W$, where μ_i and ν_i are integers.

Without loss of generality, we can suppose, for all $x \in \{0,1\}^n$, $1 \leq |\bar{\mu}\bar{x} - \theta| \leq W$ and $1 \leq |\bar{\nu}\bar{x} - \tau| \leq W$. For if some x, $\bar{\mu}\bar{x} - \theta = 0$, we can change it into $2(\bar{\mu}\bar{x} - \theta) - 1$, which can not be zero but still has $\text{sign}(\bar{\mu}\bar{x} - \theta) = \text{sign}(2(\bar{\mu}\bar{x} - \theta) - 1)$.

We study the problem of learning $a(x) \wedge b(x)$.

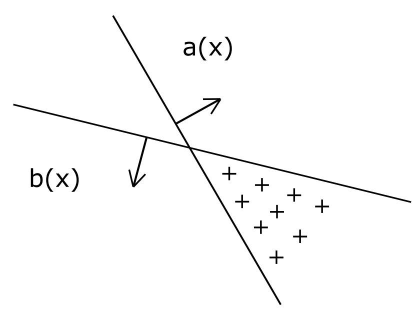

Figure 1: $a(x) \wedge b(x)$

What is $a(x) \oplus b(x)$?

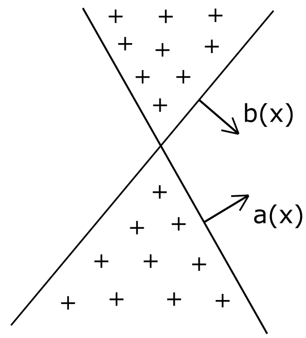

Figure 2: $a(x) \oplus b(x)$

It is easy to learn $a(x) \oplus b(x)$. Notice that the truth table for $a(x) \oplus b(x)$ is,

$(\bar{\mu}\bar{x} - \theta)$	$(\bar{\nu}\bar{x} - \tau)$	$a(x) \oplus b(x)$
$+$	$+$	$+$
$+$	$-$	$-$
$-$	$+$	$-$
$-$	$-$	$+$

So we can learn $a(x) \oplus b(x)$ easily via the degree 2 PTF $\operatorname{sign}((\bar{\mu}\bar{x} - \theta) \cdot (\bar{\nu}\bar{x} - \tau))$.

We will study how to learn learn $a(x) \wedge b(x)$ via PTFs in the following sections.

3 The upper bound for Learning halfspaces by PTF

To learn $a(x) \wedge b(x)$ via PTFs, a natural approach is to build a 2-variable polynomial $p(x, y)$ such that for all integers x, y with $1 \leq |x|,\ |y| \leq W$:

$$\begin{cases} p(x,y) > 0 & \text{if } x \geq 0,\ y \geq 0 \\ p(x,y) \leq 0 & \text{if } x < 0, \text{ or } y \geq 0 \end{cases} \tag{1}$$

Take $p(\bar{\mu}\bar{x} - \theta, \bar{\nu}\bar{x} - \tau)$; this is a PTF for $a(x) \wedge b(x)$. We want $d = \deg(p)$ to be as small as possible. We may achieve this by the rational fuction approximation to the sign function.

Definition 1 *A* Rational Function *of degree d is a quotient of two degree-d polynomials*

$$Q = \frac{r(\cdot)}{s(\cdot)},$$

where $r(\cdot)$ and $s(\cdot)$ are degree-d polynomials.

Observation 1 *It suffices to construct a degree-d rational function $Q(x, y)$ satisfying Equation 1 because $Q(x, y) \cdot s(x, y)^2$, a polynomial of degree-$2d$, also satisfies Equation 1.*

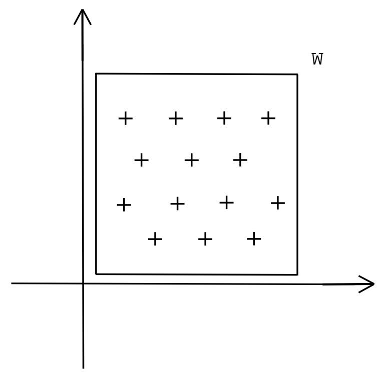

Figure 3: $p(\bar{\mu}\bar{x} - \theta, \bar{\nu}\bar{x} - \tau)$

How can we obtain $Q(x, y)$?

Observation 2 *If we had a rational function $s(x)$ computing sign$(\cdot)$, i.e. $\forall$ integers x w/ $1 \leq |x| \leq W$*

$$s(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}$$

then we can construct $Q(x, y)$ as

$$Q(x, y) = s(x) + s(y) - \frac{3}{2}$$

Observation 3 *It is ok to find $s(x) \approx$ sign(x) but that is not exactly equal to sign(x).*

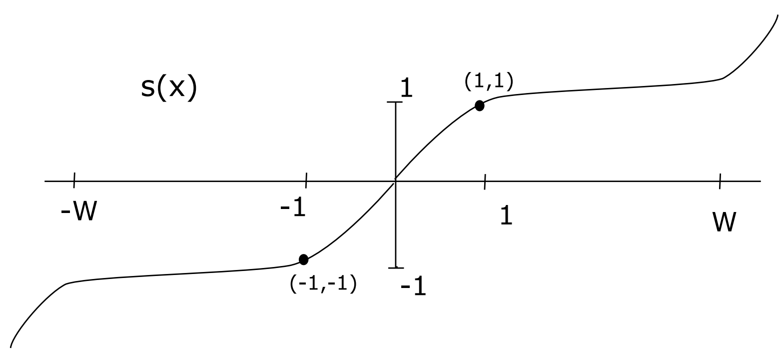

Figure 4: $s(x)$ approximates sign(x)

We'll shoot for degree-$\log W$, say $t = \log W$. Intuitively, let $s(x) = \frac{\alpha(x)}{\beta(x)}$. If we can "control" $\alpha(x)$, $\beta(x)$ at $x = 1, 2, 4, 8, \ldots, 2^t$, then we can control $s(x)$ at these points.

Define

$$P_t(x) = (x-1)(x-2)^2(x-4)^2 \ldots (x-2^t)^2 = (x-1)\prod_{i=1}^{t}(x-2^i)^2$$

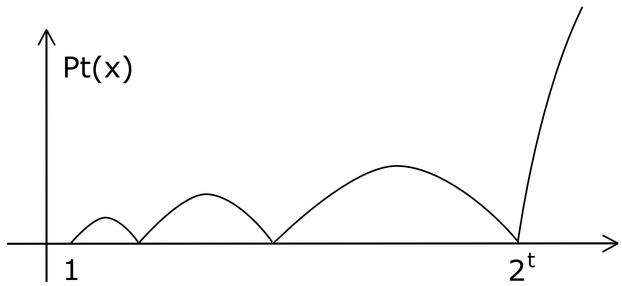

Figure 5: $Pt(x)$

Obviously, $deg(P_t) = 2t+1$

Lemma 1 *For $1 \le x \le 2^t$, we have that $0 \le 4P_t(x) \le -P_t(-x)$.*

Proof: Fix $x \in [1, 2^t]$, it is easy to see that $0 \le P_t(x)$. For any $i = 1, 2, \ldots, t$, we have

$$0 \le (x-2^i)^2 \le (-x-2^i)^2.$$

If k is s.t. $2^{k-1} \le x \le 2^k$, have

$$4(x-2^k)^2 \le 4 \cdot 2^{2k-2} = 2^{2k} < (-x-2^k)^2.$$

Clearly,

$$0 \le x-1 \le x+1.$$

Multiplying the $k-1$ inequalities for $i \ne k$ with the inequality for k, we get that

$$0 < 4(x-1)(x-2)^2(x-4)^2 \cdots (x-2^t)^2 < (x+1)(-x-2)^2(-x-4)^2 \cdots (-x-2^t)^2$$

i.e. $0 \le 4P_t(x) \le -P_t(-x)$.
■

Define

$$S_t(x) = \frac{P_t(-x) - P_t(x)}{P_t(-x) + P_t(x)}$$

The lemma below shows that $S_t(x)$ is a good approximation to sign(x) for appropriate values of x.

Lemma 2 a. *For $1 \le x \le 2^t$, have*

$$1 \le S_t(x) \le \frac{5}{3}$$

b. For $-2^t \le x \le -1$*, have*

$$-\frac{5}{3} \le S_t(x) \le -1$$

Proof:

a. If $1 \le x \le 2^t$: If $P_t(x) = 0$, then $S_t(x) = 1$. If $P_t(x) \ne 0$, then we can write

$$\begin{aligned} S_t(x) &= \frac{\frac{P_t(-x)}{-P_t(x)} + 1}{\frac{P_t(-x)}{-P_t(x)} - 1} \\ &= 1 + \frac{2}{\frac{P_t(-x)}{-P_t(x)} - 1} \\ &\le 1 + \frac{2}{3} \\ &= \frac{5}{3} \end{aligned}$$

b. This follows from above and the fact that $S_t(x)$ is an odd funtion, i.e., $S_t(-x) = -S_t(x)$.

■

Define

$$Q_t(x, y) = S_t(x) + S_t(y) - \frac{3}{2}$$

Lemma 3 *a. Have* $Q_t(x, y)$ *is a rational function of degree* $6t + 3$*,*

b. If $1 \le |x|, |y| \le W$*,*

$$Q_t(x, y) \text{ is } \begin{cases} > 0, & \text{if } x > 0 \text{ and } y > 0 \\ < 0, & \text{if } x < 0 \text{ or } y < 0 \end{cases}$$

Proof:

a. Let $S_t(x) = \frac{a(x)}{b(x)}$, then clear the denominator.

b. If $x > 0$ and $y > 0$,

$$Q_t(x, y) \ge 1 + 1 - \frac{3}{2} > 0$$

If $x < 0$ or $y < 0$,

$$Q_t(x, y) \le \frac{5}{3} - 1 - \frac{3}{2} < 0$$

■

Theorem 1 *Any intersectin of two weight-W halfspaces over $\{0,1\}^n$ has a PTF of degreee $O(\log W)$.*

Thus we can learn such halfspaces in $n^{O(\log W)}$ time. Similarly, we can extend this result to the cases that $K > 2$.

Interserction of K $(K > 2)$ weight-W halfspaces need "sharper" $S_t(x)$. Let

$$S_t^{(k)}(x) = \frac{P_t(-x)^{h(k)} - P_t(x)^{h(k)}}{P_t(-x)^{h(k)} + P_t(x)^{h(k)}}$$

$$h(k) = \text{ the smallest odd integer} \geq \log_4(2k+1)$$

Lemma 4 *a. degree of $S_t^{(k)}(x)$ is $O(t \log k)$*

b. For $1 \leq x \leq 2^t$, have

$$1 \leq S_t^{(k)}(x) \leq 1 + \frac{1}{k}$$

c. For $-2^t \leq x \leq -1$, have

$$-1 - \frac{1}{k} \leq S_t^{(k)}(x) \leq -1$$

Proof:

a. Trivial.

b. Have $4 < \frac{-P_t(-x)}{P_t(x)}$, for $1 \leq x \leq 2^t$,

$$4^{\log(2k+1)} = 2^{k+1} < \frac{P_t(-x)^{h(k)}}{(-P_t(x))^{h(k)}}$$

$$S_t^{(k)}(x) = 1 + \frac{2}{\frac{-P_t(-x)^{h(k)}}{P_t(x)^{h(k)}} - 1}$$

$$1 \leq S_t^{(k)}(x) \leq 1 + \frac{2}{2k} = 1 + \frac{1}{k}$$

c. As $S_t^{(k)}(x)$ is odd, for $-2^t \leq x \leq -1$, we get

$$-1 - \frac{1}{k} \leq S_t^{(k)}(x) \leq -1$$

■

Let our rational function be,

$$Q_t^{(k)}(x_1, \ldots, x_k) = S_t^{(k)}(x_1) + S_t^{(k)}(x_2) + \cdots + S_t^{(k)}(x_k) - (k-1)$$

Lemma 5 *a.* $Q_t^{(k)}(x_1,\ldots,x_k)$ *is a rational function of degree* $O(tk\log k)$*.*

b. For integers $x_1,\ldots,x_k$ *with* $1 \le |x_1|,\ldots,|x_k| \le W = 2^t$*.*

- *If each* $x_i > 0$*, then* $Q_t^{(k)}(x_1,\ldots,x_k) > 0$*.*
- *If any* $x_i < 0$*, then* $Q_t^{(k)}(x_1,\ldots,x_k) < 0$*.*

Proof:

a. Trivial. Just clear the denominators.

b. If each $x_i > 0$, then $Q_t^{(k)}(x_1,\ldots,x_k) \ge k-(k-1)$.
If any $x_i < 0$, then

$$\begin{aligned} Q_t^{(k)}(x_1,\ldots,x_k) &\le (k-1)(1+\frac{1}{k})-1-(k-1) \\ &= (k-1)\frac{k+1}{k}-k \\ &= -\frac{1}{k} \end{aligned}$$

■

Theorem 2 *Any intersection of* k *weight* W *halfspaces has a PTF of degreee* $O(k\log k\log W)$*.*

Such intersections can be learned in $n^{O(k\log k\log W)}$ time. Can we learn arbitary boolean functions of k halfspaces?

- Any boolean function $g(x_1,\ldots,x_k)$;
- $g(h_1(x),\ldots,h_k(x))$;

Theorem 3 *Let* $g : \{-1,1\}^k \to -1,1$ *be any boolean function over* k *variables. Let* $h_1(x_1,\ldots,x_n),\ldots,h_k(x_1,\ldots,x_n)$ *all be weight* W *halfspaces.*
Then $g(h_1(x),\ldots,h_k(x))$ *has a PTF of degree* $O(k^2\log W)$*.*

Sketch of Proof Take $h(k) = 3k$, and get that $S_t(x)$ always lies in $[1, 1+1/2^{3k}]$, or $[-1-1/2^{3k}, -1]$.
$g(Y_1,\ldots,Y_k)$ is computed by a degree k polynomial. i.e. $P(Y_1,\ldots,Y_k) = g(Y_1,\ldots,Y_k)$ on all $\bar{y} \in \{-1,1\}^k$.
Agree that $P(S_t(Z_1),\ldots,S_t(Z_k))$ is $\approx \pm 1$, for all $Z \in [1,2^t]$or$[-2^t,-1]$.
Degree k polynomial in k rational functions, each of degree $O(k\cdot t)$.
Multiply, clear denominators $\Rightarrow$ get $O(k^2\log W)$.

■

These are our upper bounds for learning halfspaces by PTFs.

4 Lower Bounds

Let $MAJ(x_1, \ldots, x_n) = \text{sign}(x_1 + \cdots + x_n)$, where $\bar{x} \in \{-1, 1\}^n$. Weight of MAJ is n.
Let $g(x, y) = MAJ(\bar{x}) \wedge MAJ(\bar{y})$. i.e. $g : \{-1, 1\}^{2n} \to \{-1, 1\}$.

Theorem 4 *The PTF degree of $g(x, y)$ is $\Omega(\log n / \log \log n)$.*

The proof uses the "discriminator lemma."

Lemma 6 (Discriminator Lemma) *Let $f(x_1, \ldots x_n) : \{-1, 1\}^n \to \{-1, 1\}$ be $f(\bar{x}) = sign(w_1 x_1 + \cdots + w_n x_n)$, with*

a. each $x_i \in \{-1, 1\}$

b. each w_i is an integer and $W = \sum |w_i|$.

c. For all $Z \in \{-1, 1\}^n$, have $w_1 x_1 + \cdots + w_n x_n \neq 0$

Then for any probability distribution $\mathcal{D}$ over $\{-1, 1\}^n$, there must be some x_i such that,

$$\Big| E_{x \in \mathcal{D}}[f(x) \cdot x_i] \Big| \geq \frac{1}{W}$$

For example, if $f(x) = x_1$, then for any distribution $\mathcal{D}$ we have that $E_{\mathcal{D}}[f(x) \cdot x_1] = 1$. If $f(x) = x_2$, then $E_{\mathcal{D}}[f(x) \cdot x_i]$ can be 1 (if the distribution $\mathcal{D}$ puts all its weight on inputs for which $x_1 = x_2$), or can be 0 (if for instance the distributino $\mathcal{D}$ is uniform over $\{-1, 1\}^n$.)
NOTE:Intuitively, the lemma says that "A LTF is always correlated with some variable."

Proof: Have $1 \leq |\bar{w}\bar{x}| \leq W$, for all $x \in \{-1, 1\}^n$. So

$$\begin{aligned}
1 &\leq E_{x \in \mathcal{D}}\big[|\bar{w} \cdot \bar{x}|\big] \\
&= E_{x \in \mathcal{D}}[f(x)(\bar{w} \cdot \bar{x})] \\
&= E[f(x)(w_1 x_1 + \cdots + w_n x_n)] \\
&= E[f(x) w_1 x_1] + \cdots + E[f(x) w_n x_n] \\
&= w_1 E[f(x) x_1] + \cdots + w_n E[f(x) x_n] \\
&\leq |w_1| E[f(x) x_1] + \cdots + |w_n| E[f(x) x_n] \\
&\leq W \cdot max \Big| E[f(x) x_i] \Big|
\end{aligned}$$

■

We'll sketch the proof of Theorem 3 in the next lecture.