COMS 6998-2 Advanced Topics in Computational Learning Theory **Spring 2007**

Lecture 5 : February 15, 2007

Lecturer: Rocco Servedio *Scribe: Phong Pham*

In the last lecture we gave upper and lower bounds on the PTF degree of depth-2 circuits (DNF formulas). There are many open questions about upper and lower bounds on the PTF degree of more complicated circuits. In particular, very little is known about the PTF degree that is required even for poly(n)-size, depth-3 circuits. The best lower bound that is known is $\Omega(n^{\frac{1}{3}} \log^{\frac{2}{3}} n)$ (note that this is just a little bit stronger than the $\Omega(n^{1/3})$ lower bound we saw for depth-2 circuits), and the best upper bound known is a trivial upper bound of $O(n)$.

This lecture:

1. We will discuss a new class of functions: intersections of halfspaces

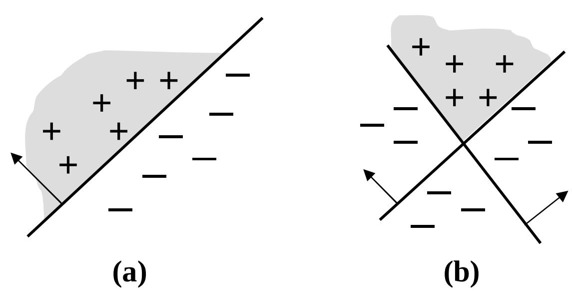

Figure 1: (a) One halfspace (b) An intersection of 2 halfspaces

2. We will derive a PTF-degree upper bound, and will give some idea about PTF-degree lower bounds for intersections of halfspaces.

3. There is no known efficient algorithm for learning intersections of arbitrary halfspaces. Hence, to prove any positive results, we will need to restrict ourselves to considering only weight W halfspaces.

Readings: [Beigel et al. 1991, section 3], [Klivans et al. 2002, section 4], [O'Donnell and Servedio 2003, section 5]

1 Upper bound on PTF degree of intersections of 2 halfspaces

Consider 2 halfspaces

$$\begin{aligned} a(x_1, x_2, \dots, x_n) &= \operatorname{sign}(u_1 x_1 + u_2 x_2 + \cdots + u_n x_n - \theta) \quad (\theta, \tau, u_i, v_i \in \mathbb{Z}\ \forall i = 1, \dots, n) \\ b(x_1, x_2, \dots, x_n) &= \operatorname{sign}(v_1 x_1 + v_2 x_2 + \cdots + v_n x_n - \tau) \end{aligned}$$

We will assume that weights of $a(x)$ and $b(x)$ are bounded by W, i.e.

$$W \geq \max\left\{\sum_{i=1}^{n} |u_i|, \sum_{i=1}^{n} |v_i|\right\}$$

As a warmup to thinking about the PTF degree of an intersection of halfspaces, let's consider the PTF-degree of $a(x) \oplus b(x)$. Notice that the sign of $a(x) \oplus b(x)$ is positive if and only if $a(x)$ and $b(x)$ are of opposite signs. Thus we can simply take $-a(x)b(x)$, a degree-2 polynomial, as a PTF for $a(x) \oplus b(x)$, and therefore the XOR of any two halfspaces has a PTF of degree 2, regardless of the weight W.

1.1 Intersection of 2 halfspaces

We now derive a PTF degree upper bound for $a(x) \wedge b(x)$. Without loss of generality[1], assume that for all $x \in \{0,1\}^n$

$$\begin{aligned} u_1 x_1 + u_2 x_2 + \cdots + u_n x_n - \theta &\neq 0 \\ v_1 x_1 + v_2 x_2 + \cdots + v_n x_n - \tau &\neq 0 \end{aligned}$$

Since $\theta, \tau, u_i, v_i \in \mathbb{Z}$ for all $i = 1 \dots n$, we must have that $(\overline{ux} - \theta)$ and $(\overline{vx} - \tau)$ are integers in the range $[-W, -1] \cup [1, W]$ for all $x \in \{0,1\}^n$.

We will try to construct a 2-variable polynomial $p(x, y)$ such that

$$\forall x, y \in \mathbb{Z}, 1 \leq |x|, |y| \leq W \qquad \begin{cases} p(x,y) > 0 & \text{if } x > 0 \wedge y > 0 \\ p(x,y) < 0 & \text{if } x < 0 \vee y < 0 \end{cases}$$

The behavior of $p(x, y)$ is described graphically in Figure 2. It is easy to see that if we can construct a low degree (says degree d) polynomial p, then $p(\overline{ux} - \theta, \overline{vx} - \tau)$ is our desired degree-d PTF.

[1] If, for example, $u_1 x_1 + u_2 x_2 + \cdots + u_n x_n - \theta = 0$ for some $x \in \{0,1\}^n$, we can do a simple fix by doubling all weights u_i and replace θ with $2\theta + 1$. The integrality of all weights are still preserved, while our non-zero assumption is now satisfied. While doing this will double the weight bound W, it does not affect our analysis later on.

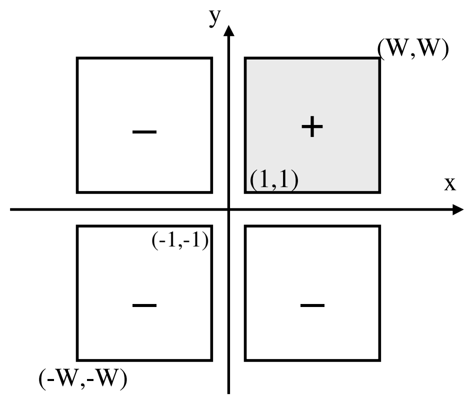

Figure 2: Desired behavior for polynomial $p(x, y)$

Definition 1 *A rational function of degree d is a quotient $Q = \frac{r}{s}$ in which r and s are polynomials of degree at most d.*

Observation 1 *It would be sufficient to construct a degree-d rational function $Q(x, y) = \frac{r(x,y)}{s(x,y)}$ which behaves like $p(x, y)$.*

Consider $p(x, y) = r(x, y)s(x, y) = \frac{r(x,y)}{s(x,y)}s(x, y)^2 = Q(x, y)s(x, y)^2$. Since $Q(x, y)$ is defined on all our points of interest (integer pairs in $([-W, -1] \cup [1, W]) \times ([-W, -1] \cup [1, W])$), we must have $s(x, y) \neq 0$, which means $s(x, y)^2 > 0$ and therefore $p(x, y)$ has the same sign as $Q(x, y)$. If $Q(x, y)$ has the desired behavior, so will $p(x, y)$.

Observation 2 *Suppose we have a low degree rational function $s(x)$ such that $s(x) = \operatorname{sign}(x)$, $\forall x \in \mathbb{Z}, 1 \leq |x| \leq W$, then we can simply construct*

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

Observation 3 *We don't need $s(x)$ to be exactly -1 and 1 on these points, it's sufficient to be close.*

Suppose $s(x)$ can approximate $\operatorname{sign}(x)$ to accuracy ϵ, i.e. $s(x) \in [1 - \epsilon, 1 + \epsilon]$ $\forall x \in \mathbb{Z}, 1 \leq x \leq W$, and $s(x) \in [-1 - \epsilon, -1 + \epsilon]$ $\forall x \in \mathbb{Z}, -W \leq x \leq -1$. Then we have:

$$\begin{aligned} s(x) + s(y) &\geq 1 - \epsilon + 1 - \epsilon = 2 - 2\epsilon && \text{if } x, y \geq 0 \\ s(x) + s(y) &\leq (-1 + \epsilon) + (1 + \epsilon) = 2\epsilon && \text{if either } x \text{ or } y \text{ is negative} \end{aligned}$$

For the positive and negative parts to be separable, we must have $2-2\epsilon > 2\epsilon \Rightarrow \epsilon < \frac{1}{2}$. With this condition on ϵ, it is easy to see that $Q(x,y) = s(x)+s(y)-1$ is the rational function we want. This means it's sufficient for $s(x)$ to approximate $\text{sign}(x)$ up to accuracy $\epsilon < \frac{1}{2}$.

The 3 observations above lead us to the problem of approximating the sign function on integer points in range $[-W,1] \cup [1,W]$ using a low degree rational function. This problem has been studied since the 19^{th} century. Some results are available in [Newman 1964], Beigel et al. [1991].

The problem restated: we want to find a rational function $s(x)$ such that:

$$\begin{aligned} s(x) &\approx 1 \quad \forall x \in [1,W] \cap \mathbb{Z} \\ s(x) &\approx -1 \quad \forall x \in [-W,-1] \cap \mathbb{Z} \end{aligned}$$

For the rest of this section, we will show that we can find such a rational function $s(x)$ with degree roughly $t = \log W$.

Consider

$$P_t(x) = (x-1)(x-2)^2(x-4)^2(x-8)^2 \cdots (x-2^t)^2.$$

Notice that $\deg(P_t) = 2t+1$. We have the following lemma:

Lemma 1

$$0 \le 4P_t(x) \le -P_t(-x) \qquad \forall x, 1 \le x \le 2^t$$

Proof: The first inequality is obvious since $(x-1) \ge 0$ for all $x \ge 1$, and the rest of the terms in P_t are non-negative.

For the second inequality, we have

$$\begin{aligned} -P_t(-x) &= (x+1)(x+2)^2(x+4)^2(x+8)^2 \cdots (x+2^t)^2 \\ P_t(x) &= (x-1)(x-2)^2(x-4)^2(x-8)^2 \cdots (x-2^t)^2 \end{aligned}$$

It's easy to see that $(x+1) \ge (x-1)$ and $(x+2^i)^2 \ge (x-2^i)^2$ for all $x \ge 1$. Furthermore, since $1 \le x \le 2^t$, there exists k, $1 \le k \le t$ such that $2^{k-1} \le x < 2^k$. We have

$$\begin{array}{rrcccc} & 2^{k-1} & \le & x & \le & 2^k \\ \Leftrightarrow & -2^{k-1} = 2^{k-1} - 2^k & \le & x - 2^k & < & 0 \\ \Rightarrow & & & (x-2^k)^2 & \le & \left(2^{k-1}\right)^2 \end{array}$$

and therefore $4(x-2^k)^2 \leq 4\left(2^{k-1}\right)^2 = (2^k)^2 \leq (x+2^k)^2$. We have what we need to prove. ■

Lemma 2 *Let*

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

then

$$\begin{aligned} 1 \leq s_t(x) \leq \frac{5}{3} & \qquad \forall x, 1 \leq x \leq 2^t = W \\ -\frac{5}{3} \leq s_t(x) \leq -1 & \qquad \forall x, -W = -2^t \leq x \leq -1 \end{aligned}$$

Proof: Consider $1 \leq x \leq 2^t$.

If $P_t(x) = 0$, it's easy to see that $P_t(-x) \neq 0$, therefore $s_t(x) = 1$.

If $P_t(x) \neq 0$, we rewrite $s_t(x)$ as follows

$$\begin{aligned} s_t(x) &= \frac{\frac{P_t(-x)}{-P_t(x)} + 1}{\frac{P_t(-x)}{-P_t(x)} - 1} \\ &= 1 + \frac{2}{\frac{-P_t(-x)}{P_t(x)} - 1} \qquad (1) \\ &\leq 1 + \frac{2}{4-1} \qquad \text{(from Lemma 1)} \\ &= \frac{5}{3}. \end{aligned}$$

And since $-P_t(-x) \geq P_t(x)$, (1) also implies $s_t(x) \geq 1$.

It's easy to see that $s_t(-x) = -s_t(x)$, therefore the result for the case $-2^t \leq x \leq -1$ follows immediately. ■

We can then construct $Q(x,y) = s_t(x) + s_t(y) - \frac{3}{2}$. Degree of $Q(x,y)$ is at most 3 times the degree of $P_t(x)$, and hence is at most $6t + 3 = O(\log W)$.

So any intersection of 2 weight-W halfspaces has an $O(\log W)$-degree PTF, and therefore can be learned in $n^{O(\log W)}$ time.

1.2 Extended results

What about a PTF-degree bound for an AND of k halfspaces, $h_1(x) \wedge h_2(x) \wedge \ldots \wedge h_k(x)$? In fact, we can show that there exists a variant $S_t^{(k)}(x)$ with degree $O(t \log k)$ and

$$\begin{aligned} 1 \leq S_t^{(k)}(x) \leq 1 + \frac{1}{k} \qquad & \forall x \in [1, W] \\ -1 - \frac{1}{k} \leq S_t^{(k)}(x) \leq -1 \qquad & \forall x \in [-W, -1] \end{aligned}$$

and from that construct

$$Q(x, y) \quad = \quad S_t^{(k)}(x_1) + S_t^{(k)}(x_2) + \cdots + S_t^{(k)}(x_n) - \left(k - \frac{1}{2}\right)$$

which is a rational function of degree $O(tk \log k)$. Therefore any AND of k weight-W halfspaces has a PTF of degree $O(k \log k \log W)$.

What about more general functions of halfspaces? Let g be an arbitrary boolean function on k input variables, $g : \{-1, 1\}^k \rightarrow \{-1, 1\}$, and let $h_1, h_2, \ldots, h_k$ be halfspaces with weight W. We have the following theorem:

Theorem 1 *The function $g(h_1(x), h_2(x), \ldots, h_k(x))$ has a PTF of degree $O(k^2 \log W)$.*

[Klivans et al. 2004] provides a full proof for this theorem. The idea of the proof is as follows: there is a polynomial p of degree k that exactly computes $g : \{-1, 1\}^k \rightarrow \{-1, 1\}$ on all inputs in $\{-1, 1\}^k$. For each $i = 1, \ldots, k$ we use a rational function that is $\pm\frac{1}{2^{3k}}$-close to $h_i(x)$ for all $x \in \{-1, 1\}^n$ (this rational function is created using the $S_t^{(2^{3k})}$ function mentioned above; it has degree $O(tk)$) and plug this rational function in for the i-th input variable to g. It can be proved that the resulting rational function will take the same sign as $g(h_1, \ldots, h_k)$ on every input $x \in \{-1, 1\}^n$; converting the rational function to a polynomial as described earlier, we get the desired result.

2 Lower bound on PTF degree for intersections of 2 halfspaces

Consider the majority function $\mathrm{MAJ}(x_1, \ldots, x_n) = \mathrm{sign}(x_1 + \cdots + x_n)$, $x \in \{-1, 1\}^n$. We define a $2n$-variable boolean function g as follows

$$g : \{-1, 1\}^{2n} \to \{-1, 1\}$$
$$g(x_1, \ldots, x_n, x_{n+1}, \ldots, x_{2n}) = \mathrm{MAJ}(x_1, \ldots, x_n) \wedge \mathrm{MAJ}(x_{n+1}, \ldots, x_{2n})$$

Minsky and Papert [1969] showed that any PTF for g must have degree $\omega(1)$, i.e. the degree must grow with n. O'Donnell and Servedio [2003] later showed a stronger lower bound as stated in the following theorem:

Theorem 2 *Any PTF for g must have degree $\Omega\left(\frac{\log n}{\log \log n}\right)$.*

Note that this is not far from the best possible, since the result from earlier in this lecture shows that g has a PTF of degree $O(\log n)$.

We will not show the full proof for this theorem, but we now present an important ingredient used in this proof and explain the basic role this ingredient plays in the proof. This ingredient, the "discriminator lemma," will be useful for learning algorithms later in the course.

Lemma 3 ("Discriminator Lemma") *Let $f(x_1, \ldots, x_n) : \{-1, 1\}^n \to \{-1, 1\}$ be $f(x) = \mathrm{sign}(w_1 x_1 + \cdots w_n x_n)$, where*

1. *$w_i \in \mathbb{Z}$ for all $i = 1 \ldots n$.*
2. *$\overline{wx} \neq 0$ for all $x \in \{-1, 1\}^n$.*

and let $W = \sum_{i=1}^n |w_i|$, then for any distribution $\mathcal{D}$ over $\{-1, 1\}^n$, there exists some x_i, $1 \leq i \leq n$, such that

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

Proof: For all $x \in \{-1,1\}^n$, we have $1 \leq |\overline{wx}| \leq W$. Therefore

$$\begin{aligned} 1 &\leq E_{x\in\mathcal{D}}\left[|\overline{wx}|\right] \\ &= E_{x\in\mathcal{D}}\left[f(x)\overline{wx}\right] \\ &= E_{x\in\mathcal{D}}\left[f(x)w_1x_1 + f(x)w_2x_2 + \cdots + f(x)w_nx_n\right] \\ &= w_1E_{x\in\mathcal{D}}\left[f(x)x_1\right] + \cdots + w_nE_{x\in\mathcal{D}}\left[f(x)x_n\right] \\ &\leq W\max_{i=1}^{n}\left|E_{x\in\mathcal{D}}\left[f(x)x_i\right]\right| \end{aligned}$$

Let $j = \arg\max_{i=1}^{n}\left|E_{x\in\mathcal{D}}\left[f(x)x_i\right]\right|$, then we will have

$$\left|E_{x\in\mathcal{D}}\left[f(x)x_j\right]\right| \geq \frac{1}{W}$$

■

From this lemma, if f is any LTF, then for any distribution $\mathcal{D}$ on the input space, there exists some x_i that has non-zero correlation with f.

Rephrasing this in a different way, if we have a function f, and we can prove that there exists some distribution $\mathcal{D}$ on the input space such that $E\left[f(x)x_i\right] = 0$ for all i, then f is not an LTF.

We can similarly extend this statement to PTFs: if we have a function f, and we can prove that there exists some distribution $\mathcal{D}$ on input space such that $E[f(x)M(x)] = 0$ for all polynomials $M(x)$ of degree at most d, then f is not a degree-d PTF.

The proof of the $\Omega(\frac{\log n}{\log\log n})$ lower bound uses this basic approach. It constructs a particular distribution $\mathcal{D}$ over $\{0,1\}^n$, and shows that for this distribution, every monomial $M(x)$ of degree $O(\frac{\log n}{\log\log n})$ has $E_{\mathcal{D}}[f(x)M(x)] = 0$.

References

R. Beigel, N. Reingold, and D. Spielman. PP is closed under intersection. In *STOC '91: Proceedings of the twenty-third annual ACM symposium on Theory of computing*, pages 1–9, New York, NY, USA, 1991. ACM Press. ISBN 0-89791-397-3.

A. Klivans, R. O'Donnell, and R. Servedio. Learning intersections and thresholds of halfspaces. In *Proceedings of 43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 177–186, 2002.

A. R. Klivans, R. O'Donnell, and R. A. Servedio. Learning intersections and thresholds of halfspaces. *Journal of Computer & System Sciences*, 68(4):808–840, 2004. ISSN 0022-0000.

M. Minsky and S. Papert. *Perceptrons: An Introduction to Computational Geometry*. MIT Press, 1969.

D. Newman. Rational approximation to $|x|$. *Michigan Mathematical Journal*, 11: 11–14, 1964.

R. O'Donnell and R. A. Servedio. New degree bounds for polynomial threshold functions. In *STOC '03: Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 325–334, New York, NY, USA, 2003. ACM Press. ISBN 1-58113-674-9.