CS 395T Computational Learning Theory

Lecture 5: September 20, 2006 — Lecturer: Adam Klivans

Scribe: Jeremy Stober

5.1 Degree Bounds for Polynomial Threshold Functions

Recall the observation that any boolean function has a polynomial threshold function of degree less than or equal to the number of variables n. For example, let

$$b(\bar{x}) = x_1 \wedge x_2 \wedge \neg x_3 \vee \cdots .$$

We may construct a polynomial threshold function of degree $\leq n$ for b as follows

$$p(\bar{x}) = x_1 x_2 (1 - x_3) + \cdots .$$

Note that since $\bar{x} \in \{0,1\}^n$ we have that $x_i^k = x_i$ for $k \neq 0$. Hence we will only consider multilinear polynomial threshold functions below.

The natural question arises: Is there a boolean function for which any polynomial threshold function will have degree at least n? The answer is yes.

5.1.1 PARITY

Theorem 1 *If p is a polynomial threshold function of degree d for computing PARITY, then $d \geq n$.*

To demonstrate this, we begin with the following definition.

Definition 1 *A polynomial p in n variables is symmetric if p it can be written as*

$$\sum_{0 \leq \alpha \leq n} c_\alpha \binom{x_1 + \cdots + x_n}{\alpha}$$

where

$$\binom{x_1 + \cdots + x_n}{\alpha}$$

denotes the set of monomials of degree α.

In otherwords, f is symmetric if all monomials of the same degree share the same coefficient. For example, consider $f(x_1, x_2, x_3) = 3x_1 + 3x_2 + 3x_3 + 4x_1x_2 + 4x_2x_3 + 4x_1x_3$. Then f is symmetric since monomial terms of degree 1,2 and 3 have coefficients of 3,4, and 0 respectively.

We also require the following helpful lemmas in the proof below.

Lemma 1 *If p is a symmetric polynomial of degree d then there exists a polynomial q of degree d such that*

$$p(x_1, \dots, x_n) = q(x_1 + \dots + x_n).$$

Lemma 2 *If p is a polynomial of degree d then the polynomial*

$$p'(x_1, \dots, x_n) = \sum_{\pi} p(\pi(x_1, \dots, x_n))$$

is a symmetric polynomial of degree d.

We proceed with the proof the the theorem.

Proof:

Let p be a polynomial threshold function of degree d that computes $PARITY(x_1, \dots, x_n)$. We construct a new polynomial

$$p'(x_1, \dots, x_n) = \sum_{\pi} p(\pi(x_1, \dots, x_n))$$

where π varies over all permutations of the input variables $x_1, \dots, x_n$.

PARITY is invariant under permutation of the input bits, so p is also invariant under permutation of input bits. Since p' is the sum of p under every permutation, p' is a polynomial threshold function for PARITY.

p' is symmetric of degree d by lemma 2 above. By lemma 1, there exists a polynomial q of degree d such that

$$p'(x_1, \dots, x_n) = q(x_1 + \dots + x_n)$$

so q is also a polynomial threshold function for $PARITY(x_1, \dots, x_n)$ of degree d.

Consider the values of q on the integers 0 to n. We have

$$\begin{aligned} q(0) &< 0 \\ q(1) &\geq 0 \\ q(2) &< 0 \\ &\vdots \end{aligned}$$

since q computes PARITY. q has at least n roots implying that $d \geq n$ as desired.

■

5.1.2 Disjunctive Normal Forms

Recall that we have already seen a construction for polynomial threshold functions of degree $\tilde{O}(n^{\frac{1}{3}})$ for DNF formulae. We continue with a theorem from [1].

Theorem 2 (Minsky and Papert '68) *There exists a polynomial size DNF formula that requires PTF degree* $\Omega(n^{\frac{1}{3}})$*.*

Proof: We construct the following DNF

$$f = T_1 \vee \cdots \vee T_m$$

where each T_i term is the conjunction of $4m^2$ distinct variables. Let p be a polynomial threshold function of degree d that computes f. Let X_i denote the set of variables contained in term T_i. We consider the following polynomial

$$p' = \sum_{\pi_1,\ldots,\pi_m} p(\pi_1(X_1),\ldots,\pi_m(X_m))$$

and note that p' has degree d and is symmetric within each block of variables X_i.

Furthermore, p' is a polynomial threshold function of f since the value of p is invariant under the permutation of variables within a single set, e.g. $p(X_1,\ldots,X_m) = p(\pi_1(X_1),\ldots,\pi_m(X_m))$ for any set of permutations $\{\pi_i\}$ since f is similarly invariant.

By repeated application of lemma 1 above, there exists a polynomial q such that

$$q(\sum_{x_i \in X_1} x_i, \ldots, \sum_{x_i \in X_m} x_i) = p'.$$

q is a degree d PTF for f and in particular, $q \geq 0$ if and only if

$$\sum_{x_i \in X_j} x_i = 4m^2$$

for at least one set of variables X_j.

We now construct the following univariate polynomial

$$q(a_1(t),\ldots,a_m(t))$$

where

$$a_i(t) = 4m^2 - (t - (2i-1))^2.$$

We now examine the value of q for values of $t = 1,2,3,\ldots,m$. We have

$$\begin{aligned} q(1) &\geq 0 \\ q(2) &< 0 \\ q(3) &\geq 0 \\ &\vdots \end{aligned}$$

q has at least m roots and since q has the same degree as p, p must have degree at least m. Since we have m terms of $4m^2$ distinct variables each, we have at least $4m^3$ variables. This gives the desired lower bound of $\Omega(n^{\frac{1}{3}})$ on the degree of p. ■

5.1.3 Intersections of Halfspaces

Definition 2 *A halfspace is a function $f(\bar{x}) = SIGN(\sum_i \alpha_i x_i - \theta)$ where each α_i and θ are integers and $\bar{x} \in \{0,1\}^n$.*

Definition 3 *The weight of a halfspace h is given by*

$$weight(h) = \sum_i |\alpha_i| + |\theta|.$$

For convenience we let $H_i(\bar{x}) = \sum_i \alpha_i x_i - \theta$ where the corresponding halfspace is then denoted $h_i(\bar{x}) = SIGN(H_i(\bar{x}))$.

We consider the problem of learning intersections of halfspaces where each halfspace has the same weight. We also require that $H_i(\bar{x}) \neq 0$. We have the following theorem regarding the degree of PTFs for calculating the intersection of weight W halfspaces.

Theorem 3 *The intersection of k weight W halfspaces*

$$\bigwedge_{i=1}^{k} h_i$$

can be computed by a polynomial threshold function of degree $O(k \log k \log W)$

The proof of this theorem relies on the existence of a sequence of rational functions that approximate the SIGN function. We omit the construction of these functions but summarize the important properties of these functions in the following lemma.

Lemma 3 *For a given l and t there exists a rational function f^l_t of degree $O(t \log l)$ such that*

$$1 \leq f^l_t(x) \leq 1 + \frac{1}{l} \qquad \forall x \in [1, \ldots, 2^t]$$

$$-1 - \frac{1}{l} \leq f^l_t(x) \leq -1 \qquad \forall x \in [-2^t, \ldots, -1]$$

We can know proceed with the construction of a polynomial threshold function of the desired degree for computing halfspace intersection.

Proof:

Choosing t to be $\log W$ and l to be $2k$ in lemma 3, we construct the following threshold function

$$F(\bar{x}) = f^{2k}_{\log W}(H_1(\bar{x})) + \cdots + f^{2k}_{\log W}(H_k(\bar{x})) - k.$$

Since each of the halfspaces h_i is of weight W we have that $|H_i(\bar{x})| \leq W$. Since we choose t to be $\log W$ and l to be $2k$, we have that for $H_i(\bar{x}) \in [1, \ldots, W]$ each term will contribute between 1 and

$1 + \frac{1}{2k}$ to the sum. Similarly, for $H_i(\bar{x}) \in [-W, \ldots, -1]$ each term will contribute between -1 and $-1 - \frac{1}{2k}$ by lemma 3.

When all halfspaces are satisfied, $H_i(\bar{x}) \geq 1$ for all i and each corresponding term will contribute between 1 and $1 + \frac{1}{2k}$ resulting in $F(\bar{x}) \geq 0$. If one or more halfspaces are not satisfied, the corresponding terms will contribute a negative value in the range -1 and $-1 - \frac{1}{2k}$ resulting in $F(\bar{x}) < 0$.

The degree of F is $O(\log k \log W)$. To convert F to a polynomial threshold function, we must multiply through by the square of the product of the denominators, giving a PTF of order $O(k \log k \log W)$.

■

We end with a lower bound result for the intersection of two halfspaces. Let $h_1 = MAJORITY(x_1, \ldots, x_n)$ and let $h_2 = MAJORITY(y_1, \ldots, y_n)$. h_1 and h_2 are halfspaces since majority is equivalent to $SIGN(\sum_i x_i - \frac{n}{2})$.

Theorem 4 *Any polynomial threshold function computing $h_1 \wedge h_2$ has degree $\Omega(\frac{\log n}{\log \log n})$.*

References

[1] Marvin Minsky and Seymour A. Papert. *Perceptrons.* MIT Press, 1968.