A DNF has PTF Degree of Between $O(n^{1/3})$ and $O(n^{1/3}\log n)$

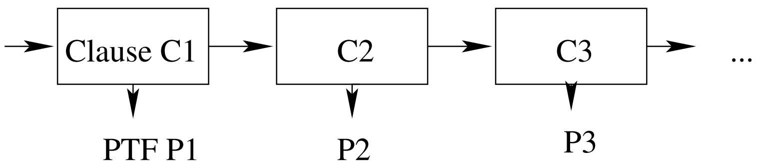
4.1 Upper Bound

4.1.1 Intuition

When you have two completely different methods for doing the same thing, you can often gain by combining them. Consider what we have when trying to learn a Disjunctive Normal Formula (DNF). [Bshouty 96] gave us the idea of converting to an Augmented Decision Tree (ADT), and in the last class, we used the Chebyshev trick of converting to a Polynomial Threshold Function (PTF). So, given an s-term DNF on n variables, let's construct an ADT of rank $\frac{2n}{t}\log s$ with DNFs of term length t at the leaves. We'll then apply the Chebyshev trick to each leaf node DNF. We hope to get a PTF of low rank that we'll be able to learn quickly and with few mistakes in the Mistake Bounded Model. This discussion follows [Klivans, Servedio 01].

4.1.2 Construction

Leaving t as an unfixed parameter at the moment, we will first convert the s-term DNF on n variables to an ADT of rank $\frac{2n}{t}\log s$ with DNFs of term length t at the leaves. Then we'll convert this into a Decision List (DL), analogous to how we converted a normal Decision Tree (DT) into a DL earlier, and represent it as below:



We are trying to convert this entire structure into one single PTF. First note that for each C_i, there exists a polynomial of degree $\frac{2n}{t}\log s$ that equals 1 when C_i is true and 0 otherwise. Call this polynomial $\widetilde{C_i}$. For example, the clause $x_1 \wedge x_3 \wedge x_4 \wedge \overline{x_5} \mapsto x_1x_3x_4(1-x_5)$. Next see that the length of this DL is less than $3^{\frac{2n}{t}\log s}\binom{n}{\frac{2n}{t}\log s} < (3n)^{\frac{2n}{t}\log s}$. (This is an upper bound on the number of different conjunctions with $\leq n$ variables that we've seen earlier in class. The logic behind it is that the number of different conjunctions with exactly n variables is exactly $2^{\frac{2n}{t}\log s}\binom{n}{\frac{2n}{t}\log s}$, the extra power of 2 factor coming because each variable can either be negated or not. To count the conjunctions with $< n$ variables as well, let each variable be negated, not negated, or not even present at all. Then we have the claimed upper bound. This will overestimate the true number, but the loss is not significant for our purposes.) Therefore, a logical choice for PTF of the entire structure is

$$A_1P_1\widetilde{C_1} + A_2P_2\widetilde{C_2} + A_3P_3\widetilde{C_3} + ...$$

Where there are $< (3n)^{\frac{2n}{t}\log s}$ terms in total. We can assume without loss of generality that the P_i are scaled so that $\forall i, \forall x$, $|P_i(x)| \geq 1$. The A_i are constants defined by

$$A_{last-one} = 1$$
$$A_i > \max_{x\in\{0,1\}^n} |A_{i+1}P_{i+1}\widetilde{C_{i+1}} + A_{i+2}P_{i+2}\widetilde{C_{i+2}} + ...|$$

This PTF actually works because for each i, if the DL returned its value at the i'th clause, all the previous clauses are false and this one is true. That is, $C_i = 1$ and $\forall j < i, C_j = 0$. So, partially evaluated, the formula is

$$A_iP_i + A_{i+1}P_{i+1}\widetilde{C_{i+1}} + A_{i+2}P_{i+2}\widetilde{C_{i+2}} + ...$$

We can upper bound the magnitude of the sum of every term but the first one as $< A_i$. Therefore, the sign of P_i will dictate the sign of the overall PTF, as required.

The degree of our PTF is $\max_i(deg(P_i) + deg(\widetilde{C_i})) \leq \frac{2n}{t}\log s + t^{\frac{1}{2}}\log s$. We will now choose $t = O(n^{\frac{2}{3}})$ to optimize this expression, and (allowing s to be poly-logarithmic in n) we get a degree of $O(n^{\frac{1}{3}}\log n)$.

4.2 Lower Bound

We will show that there is a DNF that has PTFs of degree $\Omega(n^{\frac{1}{3}})$, obtaining tightness to within a $\log n$ factor.

4.2.1 The Parity Function

First, let's give some ground knowledge. Every boolean function can be written as a DNF $f = \sum_{i=1}^{2^n} T_i$, where each T_i has degree $\leq n$.

Lemma: $PARITY$ on n variables attains degree n. For fun, we'll prove this in two different ways:

Proof: [Minsky, Papert 68] For any symmetric polynomial of degree d, $p(x_1,...x_n) = \sum_{k=0}^{d} c_k \binom{x_1+...+x_n}{k}$, where the c_k are constants and the notation $\binom{x_1+...+x_n}{k}$ is meant as the sum of all monomials of k variables. That is, a symmetric function depends only on the sum of the input bits.

Now look at $PARITY$ – assume it can be computed by a polynomial of degree d which we'll call P. Consider permutations π of $\{1...n\}$. Clearly for any π, $PARITY \circ \pi(x_1,...,x_n) = PARITY(x_1,...,x_n)$. Now consider $Q = \sum_\pi P \circ \pi(x_1,...,x_n)$. Q is a PTF of parity, and it is clearly symmetric. By the previous paragraph, there exists some Q' such that $Q(x_1,...,x_n) = Q'(x_1 + ... + x_n)$. The key point is that $Q'(z) \leq 0$ when z is even but $Q'(z) > 0$ when z is odd, since Q is a PTF. The input ranges through from 0 to n, and therefore Q' has at least n zeros, and its degree is at least n. But P was chosen arbitrarily, and given P, the degrees of Q and Q are no bigger than the degree of P. Therefore, for any PTF representing $PARITY$, its degree must be at least n. ■

Proof: Consider the vector space over $\mathbb{R}$ of functions mapping $\{-1,1\}^n \mapsto \{-1,1\}$. You can translate a formula from the $\{0,1\}^n \mapsto \{0,1\}$ representation to this one by making the replacement $x_i \mapsto \frac{1+x_i}{2}$ for each i. The set of all monomials form an orthonormal basis for this new vector space with inner product $< f,g >= \frac{1}{2^n}\sum_{x\in\{-1,1\}^n} f(x)g(x)$. This is easy to verify – when $f(x) = g(x)$, $f(x)g(x) = 1$ and the expression is 1. When $f(x) \neq g(x)$, $f(x)g(x)$ is a monomial but not 1. For some variable x_i in the monomial, we can re-express the inner product as $\frac{1}{2^n}\sum_{x_1=0}^{1}\cdots\sum_{x_n=0}^{1} = \frac{1}{2^n}\sum_{x_1=0}^{1}\cdots\sum_{x_{i-1}=0}^{1}\sum_{x_{i+1}=0}^{1}\cdots\sum_{x_n=0}^{1}(f(x)g(x)|_{x_i=-1} + f(x)g(x)|_{x_i=1}) = 0$.

We know $PARITY$ can be computed by the monomial $x_1...x_n$. This is the only monomial of degree $\geq n$ of course. So if you had some function q that was the sum of $< n$ monomials then $0 < \frac{1}{2^n}\sum_{x\in\{-1,1\}^n} PARITY * q =< PARITY, q >= 0$. The last equality follows because $PARITY$ and q have no common monomials and so the previous paragraph shows that the inner product is zero. Thus, a contradiction is obtained, and there can be no such q that computes $PARITY$ which doesn't have the n term monomial. ■

4.2.2 One in a Box Function

[Minsky, Papert 68] describes a DNF function called the "one in a box function," and we will show that it requires PTF degree $\Omega(n^{\frac{1}{3}})$. Define an m term DNF $f = T_1 \vee ... \vee T_m$ where each term T_i is the conjunction of $4m^2$ variables: $T_i = \bigwedge_{j=1}^{4m^2} x_{i,j}$. We will define the sets A_i as the variables used in the i'th term: $A_i = \{x_{i,1}...x_{i,4m^2}\}$, and emphasize the fact that the A_i are meant to be disjoint.

Assume f is computed by a PTF q having degree d. For all i, let π_i be a permutation on the variables in A_i. Consider $Q \equiv \sum_{\pi_1...\pi_m} q \circ \pi_1 \circ ... \circ \pi_m$. Q is symmetric with respect to each A_i. Define a new function Q' to exploit this: $Q = Q'(\sum_{x\in A_1} x, ... \sum_{x\in A_n} x)$. Q is positive if and only if $\sum_{x\in A_i} x = 4m^2$ for some i. Consider the following univariate polynomial $S(t) = Q'(4m^2 - (t-(2*1-1))^2, 4m^2 - (t-(2*2-1))^2, 4m^2 - (t-(2*3-1))^2, ...)$. By construction, S has degree $\leq 2d$. The key observation is that $S(0) \leq 0$, $S(1) > 0$, $S(2) \leq 0$, and etc. Therefore, S has at least $2m$ roots. This implies S has degree $\geq 2m$ and $2d \geq 2m \Rightarrow d \geq m$. Since the total number of variables in this DNF is $4m^2 * m = 4m^3$, the required degree as a function of variables is at least $(\frac{number-of-variables}{4})^{\frac{1}{3}}$.

References

[Bshouty 96] N. Bshouty. A subexponential exact learning algorithm for DNF using equivalence queries. Information Processing Letters 59 (1996), 37-39.

[Klivans, Servedio 01] Adam R. Klivans, Rocco A. Servedio. Learning DNF in Time $2^{O(n^{1/3})}$. Proceedings of the 33rd Symposium on Theory of Computing (STOC), 2001.

[Minsky, Papert 68] Marvin L. Minsky and Seymour Papert. Perceptrons. MIT press, Cambridge, MA, 1988. Expanded Edition. The first edition appeared in 1968.