Lecture 07

Earlier...

1 Theorem. If $\mathsf{Disc}_\mu(F) \leq 2^{-V}$, for $v \gg (\log n)^2$, then any majority of s linear threshold functions to compute F must have $s = \frac{2^{\Omega(V)}}{n}$.

---

2 *Gordan's Transposition Theorem.* Let $M \in \mathbb{R}^{m \times n}$. Then exactly one of the following statements must hold:

(i) For some vector u, $Mu \geq \varepsilon$ (all coordinates of Mu are positive);

(ii) $M^T v = 0$ for some non-zero vector $v \geq 0$.

Proof. Let e, e' be all-1 vectors of suitable dimension. Consider the following dual pair of linear programs:

$$\text{P} \left\{ \begin{array}{lrcl} \min & \langle e, z \rangle & & \\ \text{s.t.} & Mu + z & \geq & e' \\ & z & \geq & 0 \end{array} \right. \qquad \text{Q} \left\{ \begin{array}{lrcl} \max & \langle e', v \rangle & & \\ \text{s.t.} & M^T v & = & 0 \\ & v & \leq & e \\ & v & \geq & 0 \end{array} \right.$$

Clearly both P and Q are feasible; $z = e'$ satisfies P, and $v = 0$ satisfies Q. Hence both programs have solutions, regardless of M.

Now suppose that for any u, some coordinate of Mu is ≤ 0. Then any z satisfying the constraints of P must have some coordinate > 0. This is also true for the optimum solution u^*, z^* of P, and hence the optimum value of P is > 0. But this equals the optimum value of Q, and so the optimum solution $v = v^*$ of Q must be non-zero (and have $M^T v = 0$).

Now suppose instead that some u has $Mu \geq \varepsilon$; then we may assume $\varepsilon = 2$ by multiplying u with a sufficiently large number. Then $z = 0$ is a solution to P, and so 0 is the optimum value of both P and Q. But that implies that $v = 0$ is the only solution to Q. ■

3 Corollary. For any f, d, exactly one of the following holds:

(i) $\mathsf{deg}_\pm(f) \le d$;

(ii) there exists ρ such that $\sum_x \rho(x) f(x) \chi_S(x) = 0$ whenever $|S| \le d$.

(proof of thm)

Degree-discrepancy Theorem

The following upper bound on discrepancy will be used

4 Lemma (Discrepancy Bound). $f : X \times Y \to \{1, -1\}$, $\mu \sim X \times Y$.

$$\mathsf{Disc}_\mu(f)^2 \le |X| \sum_{y,y' \in Y} \left| \sum_{x \in X} \mu(x,y)\mu(x,y')f(x,y)f(x,y') \right|$$

The proof in the original paper is sufficiently simple and straightforward. We will not copy it here.

5 Theorem (Degree/Discrepancy theorem). Let $\mathsf{deg}_\pm(f) = d \ge 1$. Let $N \ge n$, and define $F(x, S) = f(x|_S)$ with $x \in \{1, -1\}^N$ and $S \in \binom{N}{n}$. Then

$$\mathsf{Disc}_\mu(F) \le \left(\frac{4en^2}{Nd}\right)^{d/2} \qquad (\le 2^{-d/2} \text{ for } N \ge n^2/d).$$

Proof of Degree/Discrepancy theorem. For any given distribution λ on x, S, the Discrepancy bound lemma says that:

$$\mathsf{Disc}_\lambda(f)^2 \le 2^N \sum_{S,T} \left| \sum_x \lambda(x,S)\lambda(x,T) f(x|_S) f(x|_T) \right|.$$

So let μ be given by §3, and

$$\lambda(x,S) = 2^{-N+n} \binom{N}{n}^{-1} \mu(x|_S),$$

i.e., λ first picks S uniformly, then chooses $x|_S$ according to μ, and picks the remaining x-coordinates uniformly.

Then,

$$\begin{aligned} \mathsf{Disc}_\lambda(f)^2 &\le 2^N \sum_{S,T} \binom{N}{n}^{-2} \left| \sum_x 2^{-2N+2n} \mu(x,S)\mu(x,T) f(x|_S) f(x|_T) \right|. \\ &= 2^{2n} \mathop{E}_{S,T\sim\mathcal{U}} \left[|\Gamma(S,T)| \right], \end{aligned}$$

where $\Gamma(S,T) = E_{x\sim\mathcal{U}} \left[\mu(x,S)\mu(x,T) f(x|_S) f(x|_T) \right]$.

$\Gamma(S,T)$ is sort-of a measure of f-correlation among S and T wrt μ. If $|S \cap T| = k$, then regardless of the distribution μ, we can at least say that $|\Gamma(S,T)| \le 2^{k-2n}$. Indeed, suppose that $S = \{x_1, \ldots, x_n\}$ and $T = \{x_1, \ldots, x_k, x_{n+1}, \ldots, x_{2n-k}\}$. Then:

$$\begin{aligned} |\Gamma(S,T)| &\le \mathop{E}_{x\sim U} |\mu(x|_S)\mu(x|_T) f(x_S) f(x_T)| \\ &= \sum_x 2^{-N} \mu(x|_S)\mu(x|_T) \\ &= \sum_{x|_S} \left(2^{-n}\mu(x|_S) \sum_{x|_{T\setminus S}} 2^{-n+k} \mu(x|_T) \right) \\ &\le \left(\sum_{x|_S} 2^{-n}\mu(x|_S) \right) \left(\max_{x_1,\ldots,x_k} \sum_{x|_{T\setminus S}} 2^{-n+k}\mu(x|_T) \right) \\ &\le 2^{-n} \cdot 2^{-n+k} \end{aligned}$$

(first inequality is the triangle inequality, and the last follows because μ is a distribution).

But when $|S \cap T| = k < d$, we can actually show that $\Gamma(S,T) = 0$, for the special distribution μ of §3. Indeed, assume S and T are given as before; then

$$\begin{aligned}
\Gamma(S,T) &= \sum_x 2^{-N} \mu(x_1,\ldots,x_n) f(x_1,\ldots,x_n) \mu(x|_T) f(x|_T) \\
&= \sum_{x_1,\ldots,x_n} \mu(x_1,\ldots,x_n) f(x_1,\ldots,x_n) \left(\sum_{x|_{[N]\setminus[n]}} 2^{-N} \mu(x|_T) f(x|_T) \right)
\end{aligned}$$

The right-hand parenthesis is actually a function g which depends only on $< d$ (boolean) coordinates $\{x_1,\ldots,x_k\}$, and the whole expression is the measure of f-correlation between f and g with respect to μ; which equals zero, by our choice of μ.