COMS 6998: Unconditional Lower Bounds and Derandomization	Spring 2024

Lecture 11: April 2, 2024

Lecturer: Rocco Servedio *Scribe: Walter McKelvie*

1 Preliminaries

Recall the definition

Definition 1 (LTF)**.** *A Linear Threshold Function is a function of the form*

$$sign(w \cdot x - \theta)$$

for some $(w, \theta) \in \mathbb{R}^n \times \mathbb{R}$*.*

This corresponds to a mapping of whether vertices on the hypercube are on one side or the other of a particular hyperplane. Also recall that in general computing $|f^{-1}(1)|$ (the number of vertices on one side of a hyperplane) is #P-hard.

Consider a new concept: defining $\mathbf{x}$ as a random variable uniform over $\{\pm 1\}^n$, take the distribution of the linear form $w \cdot \mathbf{x}$ corresponding to a particular LTF $\text{sign}(w \cdot x - \theta)$. We can illustrate two possible distributions of $w \cdot \mathbf{x}$:

1. If f is a majority function, then $w = [1, 1, 1, ...]$ and the distribution is a binomial distribution (as a sum of independent Bernoulli distributions).
2. If f is a decision list, then $w = [1, 2, 4, ..., 2^{n-1}]$ (up to permutation) and the distribution is uniform over the odd integers between $1 - 2^n$ and $2^n - 1$

We see that the distribution of $w \cdot \mathbf{x}$ looks different depending on w—we will argue that the second case (taking $w = [1, 1, 1, ..., 1]$) is the "nicest" of such distributions to analyze.

To see this, suppose that instead of being distributed uniformly over $\{\pm 1\}^n$, $\mathbf{x}_i$ are each independently a Gaussian $N(0, 1)$. Then for any weight vector $w = (w_1, ..., w_n)$ with $\|w\|_2 = 1$, we can see that $w \cdot \mathbf{x} \sim N(0, 1)$ in distribution. This is because the sum of independent Gaussians is Gaussian, and by linearity of variance for independent variables. Succinctly

$$N(0,\sigma_1^2) + N(0,\sigma_2^2) \sim N(0,\sigma_1^2+\sigma_2^2).$$

Recall that $N(0,1)$ has a "bell curve" distribution of the form

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right)$$

and that the tails shrink very quickly with area $\leq \exp\left(-t^2/2\right)$ (i.e., a Chernoff bound). If our distribution over each $\mathbf{x}_i$ was independently $N(0,1)$ rather than uniform over ± 1, then our weight vector wouldn't matter (besides its squared 2-norm which determines the variance). So, the "nicest" LTF is the majority function

$$\text{sign}\left(\frac{x_1 + x_2 + ... + x_n}{\sqrt{n}}\right)$$

which has distribution $w \cdot \mathbf{x}$ with $\mathbf{x} \sim \{\pm 1\}^n$ that "looks most like" $N(0,1)$. Now we will define a notion that corresponds to "looking like" $N(0,1)$.

Definition 2 (ϵ-regularity)**.** *We say that an LTF $f = sign(w \cdot x - \theta)$ is ϵ-regular if $\|w\|_2 = 1$ and $\|w\|_\infty \leq \epsilon$.*

We note that MAJ in fact has the best regularity of any function for given n, with $\epsilon = \frac{1}{\sqrt{n}}$. More generally, we can connect the ϵ-regularity with the intuition above by the so-called "Berry-Esseen Theorem". This is a quantitative form of the central limit theorem, which we recall roughly says (for $\mathbf{X}_i$ iid with unit variance)

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \mathbf{X}_i \underset{\substack{N\to\infty \\ \text{in distribution}}}{\longrightarrow} N(0,1)$$

Definition 3. *Denoting the respective CDFs of random variables $\mathbf{X}$ and $\mathbf{Y}$ as*

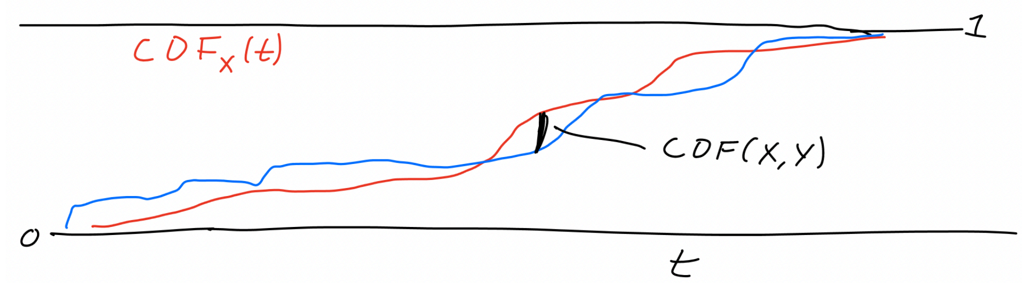
$$CDF_{\mathbf{X}}(t) = \mathbb{P}(\mathbf{X} \leq t)$$

$$CDF_{\mathbf{Y}}(t) = \mathbb{P}(\mathbf{Y} \leq t)$$

the CDF distance between $\mathbf{X}$ and $\mathbf{Y}$ is defined as

$$CDF(\mathbf{X},\mathbf{Y}) = \max_t |CDF_{\mathbf{X}}(t) - CDF_{\mathbf{Y}}(t)|$$

Intuitively, $CDF(\mathbf{X}, \mathbf{Y}) = \|CDF_{\mathbf{X}} - CDF_{\mathbf{Y}}\|_\infty$ (this characterization also allows us to immediately see that CDF distance is a pseudometric). An illustration is given below:



Theorem 4 (Berry-Esseen Theorem)**.** *Let* $\mathbf{S} = \mathbf{X}_1 + ... + \mathbf{X}_n$*, where* $\mathbf{X}_i$*'s are independent real random variables with* $\mathbb{E}[\mathbf{X}_i] = 0$ *and* $\sum Var(\mathbf{X}_i) = 1$*. Suppose each* $\mathbf{X}_i$ *has* $|\mathbf{X}_i| \leq \tau$ *almost surely. Then*

$$CDF(\mathbf{S}, N(0,1)) \leq \tau$$

At this point it should be clear that ϵ-regular LTFs are nice: supposing that I give you an ϵ-regular LTF

$$f(x) = \text{sign}(w \cdot x - \theta)$$

then I can just output $\mathbb{P}(N(0,1) \leq \epsilon)$ and that is $\pm\epsilon$ additively close to $\mathbb{P}(f(x) = 1)$ by BE Theorem.

Our goal for the rest of the lecture will be to make steps toward proving the following theorem, closely following [DGJ+10]:

Theorem 5. *Any* $\tilde{O}\left(\frac{1}{\epsilon^2}\right)$*-wise independent distribution over* $\{\pm 1\}^n$ ϵ*-fools all LTFs.*

In fact we can do better, but not much better.

1. $\frac{1}{\epsilon^2}$ turns out to be optimal up to constant (and possibly logarithmic) factors.
2. It is possible to hand-craft a different PRG of seed length $O\left(\log n + \log^2 \frac{1}{\epsilon}\right)$.

2 Fooling ϵ-regular LTFs

We will focus first on a special case of the "nice" LTFs from the last section.

Lemma 6. $\tilde{O}\left(\frac{1}{\epsilon^2}\right)$*-wise independent distribution over* $\{\pm 1\}^n$ ϵ*-fools all* ϵ*-regular LTFs.*

First recall the main way to know that k-wise independence fools something, via sandwiching polynomials:

Lemma 7. *$f : \{\pm 1\}^n \to \{\pm 1\}$ is ϵ-fooled by any k-wise independent distribution $\mathcal{D}$ if $\exists$ ϵ-sandwiching polynomials q_ℓ, q_u such that:*

1. *$deg(q_\ell), deg(q_u) \leq k$.*
2. *$q_\ell(x) \leq f(x) \leq q_u(x)$ for all $x \in \{\pm 1\}^n$.*
3. *$\mathbb{E}_{\mathbf{x}\sim\mathcal{U}}[q_u(\mathbf{x}) - q_\ell(\mathbf{x})] \leq \epsilon$.*

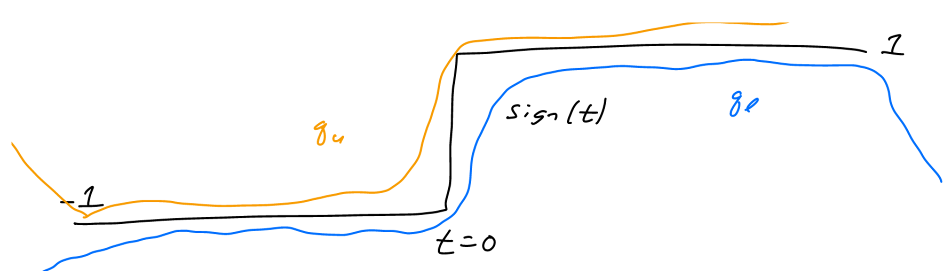
To prove Lemma 6, we will show that for any ϵ-regular LTF $f(x) = \text{sign}(w \cdot x - \theta)$, there is a univariate $\tilde{O}\left(\frac{1}{\epsilon^2}\right)$-degree sandwiching polynomial pair q_ℓ, q_u. We can do this by giving good $\tilde{O}\left(\frac{1}{\epsilon^2}\right)$-degree approximation polynomial for univariate $\text{sign}(t)$ function under $N(0,1)$.

More specifically, fixing any $f(x) = \text{sign}(w \cdot x - \theta)$ which is ϵ-regular, the Berry-Esseen Theorem says that the distribution of $w \cdot \mathcal{U}$ is ϵ-close in CDF distance to $N(0,1)$. Therefore, it suffices

Lemma 8. *There exist univariate degree-$O\left(\frac{1}{\epsilon^2}\right)$ polynomials q_ℓ, q_u such that*

1. *$q_\ell(t) \leq sign(t) \leq q_u(t)$, $\forall t \in \mathbb{R}$.*
2. *$\mathbb{E}_{\mathbf{t}\sim N(0,1)}(q_u(g) - sign(\mathbf{t})) \leq \epsilon/2$*
3. *$\mathbb{E}_{\mathbf{t}\sim N(0,1)}(sign(\mathbf{t}) - q_\ell(g)) \leq \epsilon/2$*

In other words, graphing as a function of $t = \omega \cdot x - \theta$, we want to find polynomials that upper and lower bound a step function:



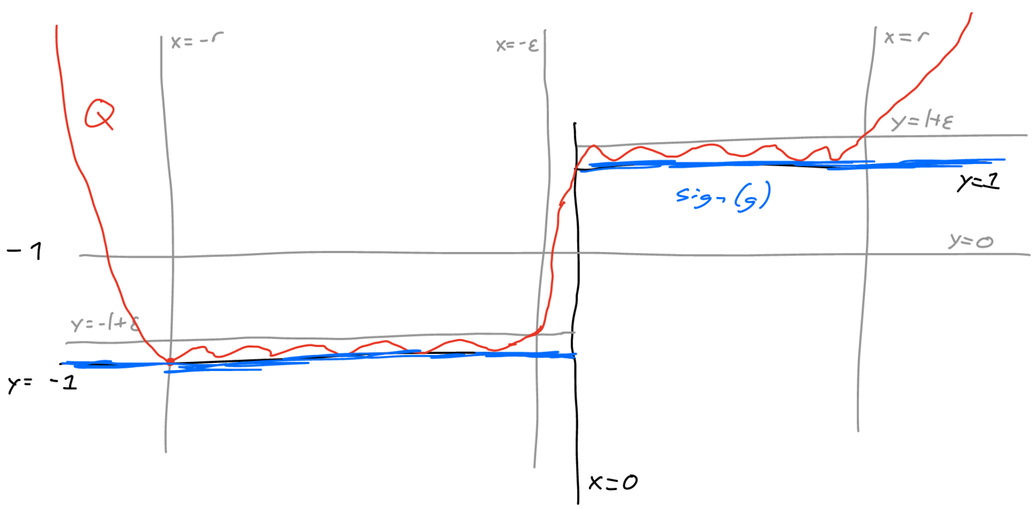
and where the notion of "distance" is weighted by a Gaussian so that more likely values of t closer to $t = 0$ contribute more error.

We will be even more specific with how we construct polynomials that fit the constraints of Lemma 8.

Lemma 9. *Let $r = \tilde{O}\left(\frac{1}{\epsilon}\right)$. There is a polynomial $Q(g)$ of degree $d \leq \tilde{O}\left(\frac{1}{\epsilon^2}\right)$, with the following properties:*

1. $Q(g) \geq sign(g) \geq -Q(-g), \forall g \in \mathbb{R}$
2. $Q(g) \in [sign(g), sign(g) + \epsilon]$ *for all* $g \in [-r, -\epsilon] \cup [0, r]$
3. $Q(g) \in [-1, 1 + \epsilon]$ *for* $g \in [-\epsilon, 0]$.
4. $Q(g) \leq 2 \cdot (4\epsilon g)^d$ *for* $|g| \geq r$.

These are a lot of constraints, but luckily a very pretty picture was drawn by Rocco in lecture:



We will now convince ourselves that Lemma 9 would imply Lemma 8. We will choose $q_u := Q(g)$ and $q_\ell := -Q(-g)$, and the first property of Lemma 8 follows easily from the first property of 9. More difficult is proving the second and third property; it suffices to prove only the second property, because from there the third would follow by reflective symmetry of the Gaussian and our definitions.

We need that

$$\mathop{\mathbb{E}}_{\mathbf{g}\sim N(0,1)}[Q(\mathbf{g}) - \mathrm{sign}(\mathbf{g})] \leq O(\epsilon)$$

and there are three areas which each contribute to the above integral:

1. Most outcomes of $g \in N(0,1)$ are in the region $g \in [-r, -\epsilon] \cup [0, r]$. The pointwise bound on the error of Q implies that the error of this region is $O(\epsilon)$.

2. Tiny regime: if $g \in [-\epsilon, 0]$ we could have pointwise error as large as $O(1)$, which would contribute $O(\epsilon)$ error.

3. If $|g| \geq r$, then the pointwise error $|Q(g) - \mathrm{sign}(g)|$ may be huge. However, it will only grow as a polynomial of degree d while the Gaussian tail bounds are exponentially small.

   Sketch: consider outcomes of g in $[r, r+1]$. We have $\mathbb{P}(g \in [r, r+1]) \leq \mathbb{P}(g > r) \leq \exp(-r^2/2)$.

   On the other hand, for such g, the error of Q is

   $$\leq 2 \cdot (4\epsilon(r+1))^d \approx (\mathrm{polylog}(1/\epsilon))^d \approx 2^{\tilde{O}(1/\epsilon^2)}$$

   By suitable choice of hidden $\log \frac{1}{\epsilon}$ factors in r, we get $e^{-r^2/2} \cdot 2^{\tilde{O}(1/\epsilon^2)} << \frac{\epsilon}{2}$. Similar argument gives $[r+t, r+t+1]$ contributes error $\leq \frac{\epsilon}{2^t}$, so the total is at most $O(\epsilon)$.

Now that we are satisfied that the total error is $O(\epsilon)$, we will prove Lemma 9. This requires another definition and theorem:

Definition 10. *Suppose we have a continuous function $f : [-1,1] \to \mathbb{R}$. Its modulus of continuity is*

$$\omega_f(\delta) = \sup_{x-y \leq \delta} |f(x) - f(y)|$$

Theorem 11 (Dunham Jackson's Theorem)**.** *Let $f : [-1,1] \to \mathbb{R}$ be bounded, continuous. Let $\ell \geq 1, \ell \in \mathbb{N}$. There exists a polynomial $J(t)$, deg$(J) \leq \ell$, such that*

$$\max_{t \in [-1,1]} |J(t) - f(t)| \leq 6 \cdot \omega_f\left(\frac{1}{\ell}\right)$$

We use these to prove the following lemma:

Lemma 12. *Let* $a = \tilde{O}(\epsilon^2)$*, let*

$$m = \frac{300 \ln \frac{1}{\epsilon}}{a} = \tilde{O}\left(\frac{1}{\epsilon^2}\right)$$

There is a polynomial $q(t)$ *of degree* $\leq m$ *such that*

$$\max_{t \in [-1,-a] \cup [a,1]} |q(t) - sign(t)| \leq \epsilon$$

(i.e., think of $Q(g)$ *as* $Q(g) = q(g/r)$*, i.e.* $Q(r \cdot t) = q(t)$*)*

Proof of Lemma 12. Define $f(t) : [-1,1] \to [-1,1]$ by

$$f(x) = \begin{cases} \text{sign}(x) & |x| \in [a,1] \\ x/a & |x| \leq a \end{cases}$$

We have $\omega_f\left(\frac{1}{\ell}\right) = \frac{1}{a \cdot \ell}$. Take $\ell = \frac{25}{a}$. As the great Dunham Jackson tells us, there exists a polynomial $J(t)$ of degree ℓ such that

$$\max_{a \leq |t| \leq 1} |J(t) - \text{sign}(t)| \leq \max_{|t| \leq 1} |J(t) - f(t)| \leq \frac{6}{a\ell} \leq \frac{1}{4}$$

We want this $\frac{1}{4}$ to instead be ϵ. We could use Jackson with larger ℓ, but we would then need degree $\tilde{O}\left(\frac{1}{\epsilon^3}\right)$ which is paying a little too much. Instead, we use a trick. Define a degree-k "amplifying polynomial"

$$A_k(u) = \sum_{j \geq \frac{k}{2}}^{k} \binom{k}{j} \cdot \left(\frac{1+a}{2}\right)^j \cdot \left(\frac{1-a}{2}\right)^{k-j}$$

This is reminiscent of a binomial distribution, in that

$$A_k(u) = \mathbb{P}[\text{toss } \frac{1+a}{2}\text{-biased coin } k \text{ times and get } \geq \frac{k}{2} \text{ heads}]$$

and we can use a Chernoff bound to get the following facts:

- If $u \in [3/5, 1]$ then $2A_k(u) - 1 \in [1 - 2\exp(-k/6), 1]$
- If $u \in [-1, -3/5]$ then $2A_k(u) - 1 \in [-1, -1 + 2\exp(-k/6)]$

Our final polynomial, then, is

$$q(t) = 2A_k\left(\frac{4}{5}J(t)\right) - 1$$

where $k = 12\log\frac{1}{\epsilon}$. Scale $J(t)$ by 4/5 to ensure that

$$\frac{4}{5}J(t) \in \left[-1, -\frac{3}{5}\right] \cup \left[\frac{3}{5}, 1\right]$$

so $2\exp(-k/6) < \epsilon$. As for our degree, we simply note that

$$\deg(g) \leq \deg(J) \cdot \deg(A_k) \leq \frac{25}{a} \cdot 12\log\frac{1}{\epsilon} = \frac{300}{a}\log\frac{1}{\epsilon} = m$$

as desired. ■

Here we stop to declare a moral victory in fooling ϵ-regular LTFs. Despite several details going unresolved, the above is the most interesting part of the proof.

For those interested, the remainder of the proof begins on page 15 in [DGJ+10]. The polynomial existence is not really constructive—it starts with the best bounded-degree polynomial approximation of sign(g) and then uses this to construct another polynomial. The analysis utilizes Chebyshev's Theorem on polynomial approximations.

3 Fooling all LTFs

There is still a big piece missing from what we were promised at the beginning: not every LTF is ϵ-regular. Like, for instance,

$$\text{sign}(2^n x_1 + 2^{n-1}x_2 + ... + x^1 x_n - \theta)$$

is only $\Theta(1)$-regular—it is really not close to a Gaussian at all. However, this specific LTF is not really difficult to deal with. In fact, it is really a decision list which is ϵ-close to a $\log\frac{1}{\epsilon}$-junta. Maybe functions which are not regular are somehow like juntas.

Say your LTF is not ϵ-regular, but still has constraint $\|w\|_2 = 1$. Without loss of generality, say $|w_1| \geq |w_2| \geq ... \geq |w_n|$. Since it is not regular, we know

$$|w_1| \geq \epsilon$$

Consider the process of throwing away the first weight and renormalizing the remaining weights, and seeing whether our new $|w_1| \geq \epsilon$.

Definition 13. *Fix $f(x) = sign(w \cdot x - \theta)$, and denote $w^\ell = (w_\ell, w_{\ell+1}, ..., w_n)$. The ϵ-critical index of f is the minimum value ℓ such that $(w_\ell, w_{\ell+1}, ..., w_n)$ is ϵ-regular, i.e.*

$$|w_\ell| \leq \epsilon \|w^\ell\|_2$$

Fact 14. *If $\ell(\epsilon)$ is the ϵ-critical index of $(w_1, ..., w_n)$, then*

$$\|w_\ell\|_2^2 = \sum_{j=\ell(\epsilon)}^{n} w_j^2 \leq \left(1 - \epsilon^2\right)^{\ell(\epsilon)-1}$$

Given any $f = \text{sign}(w \cdot x - \theta)$, consider three cases based on $\ell(\epsilon) = \epsilon$-critical index of f:

1. $\ell(\epsilon) = 1$: then f is ϵ-regular and we are done
2. $\ell(\epsilon) \leq \frac{K}{\epsilon^2}$: then $w \cdot x$ has a "junta part" for the first few variables and a "regular part" for the remaining variables.
3. $\ell(\epsilon) > \frac{K}{\epsilon^2}$: by Fact 14, $\|w^\ell\|_2^2 \leq (1-\epsilon^2)^{k/\epsilon^2} \leq e^{-K}$. Take $K = 100 \log \frac{1}{\epsilon}$, we can show that f is very close to a $\frac{K}{\epsilon^2}$-junta. The proof of this part is also omitted.

We can prove along these lines a "structure theorem" for LTFs:

Theorem 15. *Fix $\epsilon > 0$ and $f(x) = sign(w \cdot x - \theta)$. Then there exists a set $H \subseteq [n]$ of $\tilde{O}\left(\frac{1}{\epsilon^2}\right)$ variables of f (the ones with the largest $|w_i|$) such that either*

1. *$f \restriction \rho$ is ϵ-regular for every restriction ρ fixing variables in H (1,2 above)*
2. *f is ϵ-close to an H-junta.*

This structure theorem can be used to show that $\tilde{O}\left(\frac{1}{\epsilon^2}\right)$-wise independence fools all LTFs, not just ϵ-regular ones:

1. $\tilde{O}\left(\frac{1}{\epsilon^2}\right)$-wise independence fools all H-juntas and another $\tilde{O}\left(\frac{1}{\epsilon^2}\right)$-wise independence fools every ϵ-regular $f \restriction p$.
2. $\tilde{O}\left(\frac{1}{\epsilon^2}\right)$-wise independence fools any H-junta

Next time: PTFs? Harder, but we can also do some things.

References

[DGJ+10] Ilias Diakonikolas, Parikshit Gopalan, Ragesh Jaiswal, Rocco A. Servedio, and Emanuele Viola. Bounded independence fools halfspaces. *SIAM Journal on Computing*, 39(8):3441–3462, 2010. 1, 2