Recall:

Lecture 2

The Fundamental Theorem of Boolean Functions:

Every ~~Boolean~~ function $f:\{\pm1\}^n \to \mathbb{R}$ can be uniquely represented as a multilinear polynomial over $\mathbb{R}$:

$$f(x) = \sum_{S \subseteq \{1,\dots,n\}} \hat{f}(S) \cdot \chi_S(x)$$

where $\hat{f}(S) \in \mathbb{R}$ is called the S-Fourier coef. $\chi_S(x) = \prod_{i \in S} x_i$ is called the S-Fourier character.

$$\langle f, g \rangle \triangleq \mathbb{E}_{x \in_R \{\pm1\}^n} [f(x) \cdot g(x)]$$

$\deg(f) = \max\{|S| : \hat{f}(S) \neq 0\}$

Inversion Formula: $\hat{f}(S) = \langle f, \chi_S \rangle$.

Plancherel: $\langle f, g \rangle = \sum_{S \subseteq [n]} \hat{f}(S)\hat{g}(S)$.

Parseval: $\mathbb{E}_{x \in \{\pm1\}^n}[f(x)^2] = \langle f, f \rangle = \sum_{S \subseteq [n]} \hat{f}(S)^2$.

New

Note: if f is Boolean, then $\sum_{S \subseteq [n]} \hat{f}(S)^2 = 1$.

This naturally defines a probability dist D_f over subsets of $[n]$, where S is sampled w.p. $\hat{f}(S)^2$.

Def'n: The total influence is $I[f] \triangleq \mathbb{E}_{S \sim D_f}[|S|] = \sum_{S \subseteq [n]} \hat{f}(S)^2 \cdot |S|$.

$\forall k \in \{0, \dots, n\}$ the k-Fourier weight is

$$W^k[f] \triangleq \Pr_{S \sim D_f}[|S| = k] = \sum_{S : |S| = k} \hat{f}(S)^2$$

Alternative way to define $I[f]$:

For $i = 1, \dots, n$ define the discrete derivative of f in direction i as

$$D_i f(x) \triangleq \frac{f(x|_{x_i=1}) - f(x|_{x_i=-1})}{2} = \sum_{S \ni i} \hat{f}(S) \cdot \prod_{j \in S \setminus \{i\}} x_j .$$

Def'n: If f is Boolean

$$I_i[f] \triangleq \mathop{\mathbb{E}}_{x \in_R \{\pm 1\}^n}[D_i f(x)^2] \overset{\downarrow}{=} \Pr_x\left[f(x) \neq f(x^{\oplus i})\right]$$

$x^{\oplus i}$: x with ith coordinate flipped.

Note:

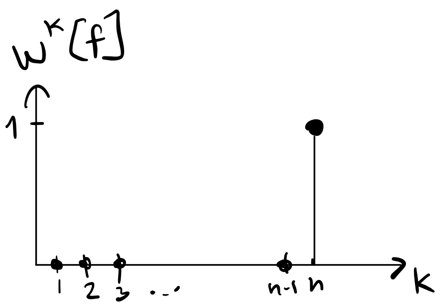
$$I_i[f] \underset{\text{Parseval}}{=} \sum_{S : S \ni i} \hat{f}(S)^2 = \Pr_{S \sim \mathcal{D}_f}[i \in S] .$$

$$I[f] = \sum_{i=1}^{n} I_i[f] = \sum_{i=1}^{n} \sum_{S : S \ni i} \hat{f}(S)^2 = \sum_{S \subseteq [n]} \hat{f}(S)^2 \cdot |S| .$$

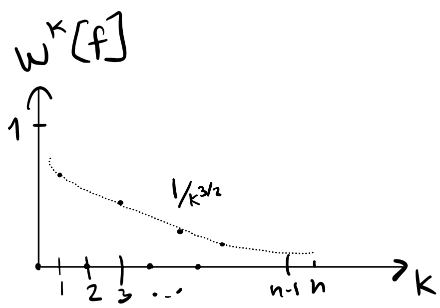
$I[f] \leq \deg(f$

Fourier Concentration and Fourier Tails

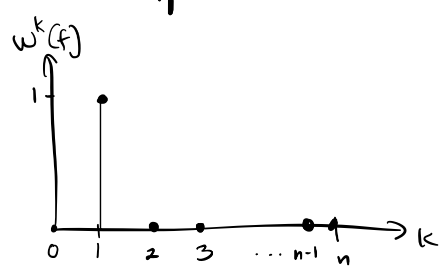
Examples: $f = \text{Parity}_n = \chi_{[n]}$



$f = \text{Majority}_n$



$f = x_i$ (dictator)



Def'n: The Fourier tail of level k is defined as $W^{\geq k}[f]$

$$W^{\geq k}[f] \triangleq \sum_{S:|S|\geq k} \hat{f}(S)^2.$$

Claim: $\forall f \quad W^{\geq k}[f] \leq \frac{I[f]}{k}$.

Proof: By Markov's Ineq. $W^{\geq k}[f] = \Pr_{S\sim \mathcal{D}_f}[|S|\geq k] \leq \frac{\mathbb{E}[|S|]}{k} = \frac{I[f]}{k}$

Low Degree Approximations:

Q: Let $f: \{\pm 1\}^n \to \mathbb{R}$.

Which multilinear polynomial $p(x)$ of degree $\leq k$ best approximates f in ℓ_2-distance, i.e. minimizes

$$\mathbb{E}_{x\in_R \{\pm1\}^n}[(f(x)-p(x))^2]$$

A: $f^{\leq k}(x) \triangleq \sum_{S:|S|\leq k} \hat{f}(S) \prod_{i\in S} x_i$ is the minimizer.

Why? Let p be a generic degree $\leq k$ multilinear polynomial.

$$\mathbb{E}_{x\in_R\{\pm1\}^n}[(f(x)-p(x))^2] = \sum_{S\subseteq[n]} (\hat{f}(S)-\hat{p}(S))^2$$

$$= \sum_{S:|S|\leq k} (\hat{f}(S)-\hat{p}(S))^2 + \sum_{S:|S|>k} \hat{f}(S)^2$$

To minimize, pick p with $\hat{p}(S)=\hat{f}(S)$ for all S of size $\leq k$.

Note: The error is exactly $W^{>k}[f]$.

Low Degree Learning Algorithm [Linial-Mansour-Nisan]

Given: random labeled examples of $(x, f(x))$

Want to come up w. a good hypothesis $h(x)$ s.t.

whp h is ε-close to f.

Here, we assume the input distribution is uniform over $\{\pm 1\}^n$.

LMN: A learning algorithm that given n, k and samples $(x, f(x))$ where $W^{>k}[f] \leq \varepsilon$ outputs whp a hypothesis that is $O(\varepsilon)$-close to f in runtime $n^{O(k)}$.

The Low-Degree Learning Algorithm

- For each $S \subseteq [n]$ of size $\leq k$
  - Estimate $\hat{f}(S) \longrightarrow \tilde{f}(S)$
- Output $h = \mathrm{sgn}\left(\sum_{|S| \leq k} \tilde{f}(S) \chi_S\right)$

How to estimate $\hat{f}(S)$?

Recall: $\hat{f}(S) = \mathbb{E}_{x \in_R \{\pm 1\}^n}[f(x) \cdot \chi_S(x)]$ $\Rightarrow$ Compute emperical mean from labeled examples

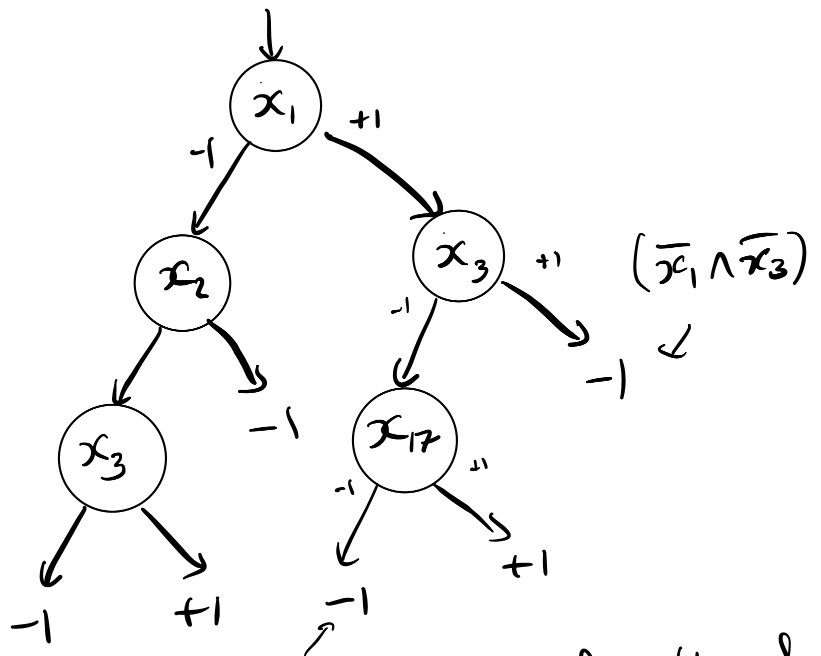
$$\Pr[f(x) \neq h(x)] \leq \mathbb{E}\left[\left(f(x) - \sum_{|S| \leq k} \tilde{f}(S) \chi_S(x)\right)^2\right] \leq W^{>k}[f] + \sum_{S : |S| \leq k} \left(\hat{f}(S) - \tilde{f}(S)\right)^2$$

[Goldreich-Levin, Kushilevitz-Mansour]:

If f is ε-close in ℓ_2-norm to a polynomial with m monomials

$\Rightarrow$ Can learn f up to error $O(\varepsilon)$ using membership queries to f in $\mathrm{poly}(m/\varepsilon)$ time.

Decision Trees



Fact: If f is a decision tree of depth d then $\deg(f) \le d$ and the Fourier repr of f has $\le 4^d$ monomials.

Idea: write $f(x) = \sum_{\ell \in \text{leaves}} 1_\ell(x) \cdot f(\ell)$.

Cor: DTs can be learned in $\mathrm{poly}(n^d)$ using examples

$\mathrm{poly}(2^d)$ using queries.

Restrictions & random restrictions

$$f(x_1,\dots,x_n) \longrightarrow f(-1,x_2,+1,x_4,x_5,\dots,-1)$$

A restriction is a partial assignment

$J \subseteq [n]$ – the alive variables

$z \in \{\pm 1\}^{\bar{J}}$ – the assignment to the rest.

The restricted function:

$$f_{J,z}(x) = f(y) \qquad y_i = \begin{cases} x_i, & i \in J \\ z_i, & i \notin J \end{cases}$$

Fourier prespective:

$$f_{J,z}(x) = \sum_{T \subseteq [n]} \hat{f}(T) \cdot \prod_{i \in T \cap J} x_i \cdot \prod_{i \in T \setminus J} z_i$$

$$= \sum_{S \subseteq J} \prod_{i \in S} x_i \cdot \underbrace{\sum_{T : T \cap J = S} \hat{f}(T) \cdot \prod_{i \in T \setminus J} z_i}_{\widehat{f_{J,z}}(S)}$$

$$g_{J,S}(z) \triangleq \widehat{f_{J,z}}(S) = \sum_{T : T \cap J = S} \hat{f}(T) \cdot \prod_{i \in T \setminus J} z_i$$

Fix J, choose z uniformly at random in $\{\pm 1\}^{\bar{J}}$.

$$\mathop{\mathbb{E}}_{z \in_R \{\pm 1\}^{\bar{J}}}\left[\widehat{f_{J,z}}(S)\right] = \hat{f}(S) \cdot 1_{S \subseteq J}$$

$$\mathop{\mathbb{E}}_{z \in_R \{\pm 1\}^{\bar{J}}}\left[\widehat{f_{J,z}}(S)^2\right] = \sum_{T} \hat{f}(T)^2 \cdot 1[T \cap J = S]$$

ρ random restrictions

- Pick $J \subseteq_p [n]$: for each $i \in [n]$ indep. include $i \in J$ w.p. p
- Sample $z \in_R \{\pm 1\}^{\bar{J}}$.

$\Big\}$ denoted $(J,z) \sim R_p$

$$\mathop{\mathbb{E}}_{(J,z)\sim R_p}\left[\widehat{f_{J,z}}(S)\right] = \hat{f}(S) \cdot \Pr_J\left[S \subseteq J\right] = \hat{f}(S) \cdot p^{|S|}$$

$$\mathop{\mathbb{E}}_{(J,z)\sim R_p}\left[\widehat{f_{J,z}}(S)^2\right] = \sum_T \hat{f}(T)^2 \cdot \Pr_J\left[T \cap J = S\right]$$

> **Lemma [LMN].**
>
> $$\mathop{\mathbb{E}}_{(J,z)\sim R_p}\left[W^k[f_{J,z}]\right] = \sum_\ell W^\ell[f] \cdot \Pr\left[\mathrm{Bin}(\ell,p)=k\right].$$

Proof:

$$\mathop{\mathbb{E}}_{(J,z)\sim R_p}\left[W^k(f_{J,z})\right] = \mathop{\mathbb{E}}_{(J,z)}\left[\sum_{S:|S|=k} \widehat{f_{J,z}}(S)^2\right]$$

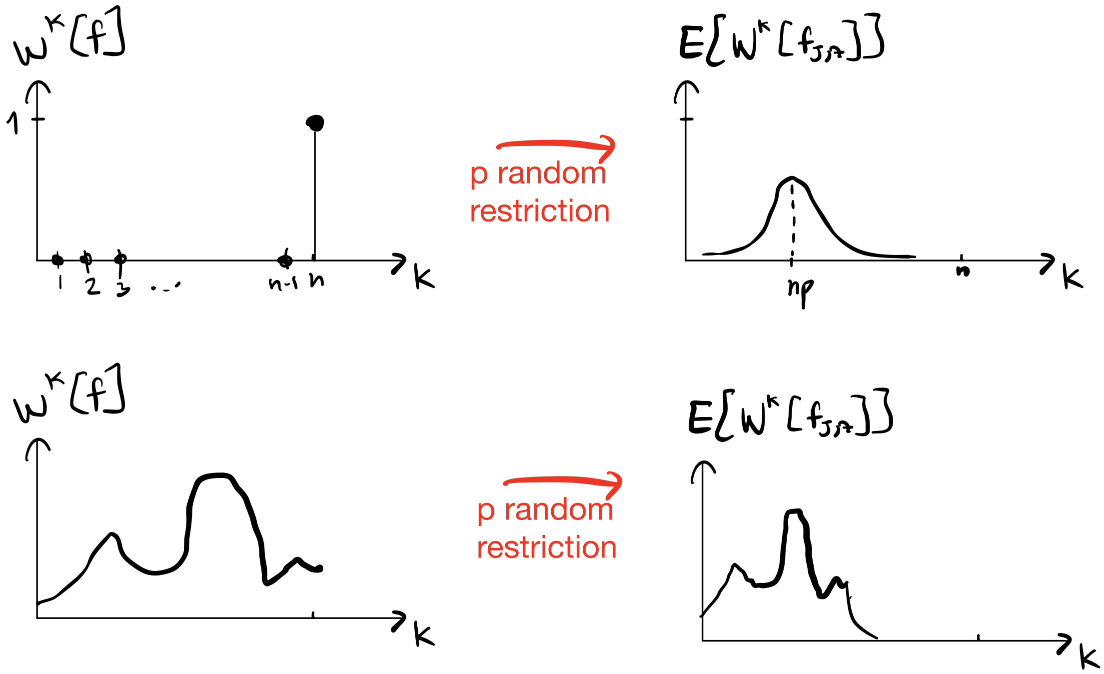
$$= \sum_{S:|S|=k} \mathop{\mathbb{E}}_{(J,z)}\left[\widehat{f_{J,z}}(S)^2\right]$$

$$= \sum_{S:|S|=k} \sum_T \hat{f}(T)^2 \cdot \Pr_{J\subseteq_p[n]}\left[T \cap J = S\right]$$

$$= \sum_{T\subseteq[n]} \hat{f}(T)^2 \cdot \underbrace{\sum_{|S|=k} \Pr_{J\subseteq_p[n]}\left[T \cap J = S\right]}_{\Pr_{J\subseteq_p[n]}\left[|T\cap J| = k\right]}.$$

$\blacksquare$

Examples $f = \text{Parity}_n$



$D_{f,p}$: the expected Fourier dist. of $f_{J,z}$

- Sample $S \sim D_f$
- Sample $J \subseteq_p [n]$, independently.
- Output $S \cap J$

Beyound Decision Trees

a width-w DNF is a function of the form

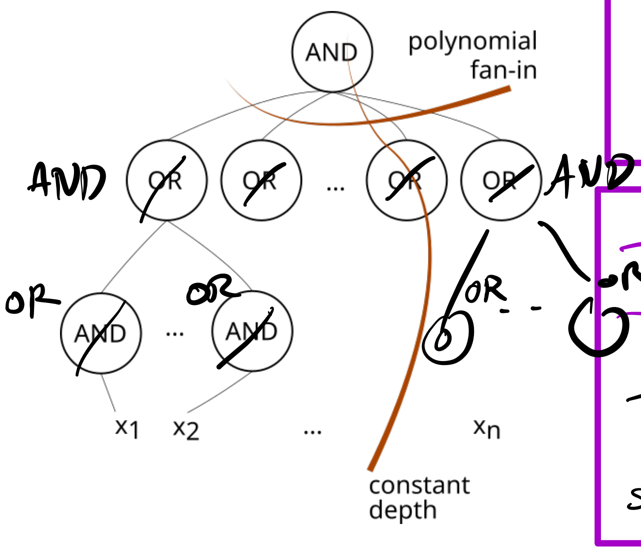
$$f = (x_{i_1} \wedge \overline{x_{i_2}} \wedge \dots \wedge x_{i_w}) \vee (\wedge \_ \wedge) \vee \dots \vee ($$

each term has $\le w$ literals

a width-w CNF is ...

Fact: Any DT of depth d can be written as a width-d DNF/CNF.

AC^0 Circuits (constant depth circuits)



Thm [FSS, Ajtai, Yao, Håstad]

Parity $\notin AC^0$.

Thm [LMN, T]: If $f \in AC^0(m, d)$ (m: size, d: depth)

then $\forall \varepsilon \geq 0 \ \exists k = O(\log m)^{d-1} \cdot \log(1/\varepsilon)$

s.t. $W^{\geq k}[f] \leq \varepsilon$.

Cor: Any $f \in AC^0(m,d)$ can be learned in $n^{O(\log m)^{d-1}}$ time.

Proof Idea: Simplification under random restrictions.

The Switching Lemma [Håstad]:

If f is a width-w DNF, then

$\forall p, k$: $\Pr_{(J,z)\sim R_p}[f_{J,z}$ is a ~~width-k CNF~~ depth-k DT$] \geq 1-(5pw)^k$

Pick $w = k = \log(m/\varepsilon)$ $\quad p = \frac{1}{10w}$

w.p. $\geq 1-\varepsilon$ all width-w DNFs $\xRightarrow{\text{random rest.}}$ width-w CNFs

which reduces the depth by 1.

Apply $d-1$ iterations $\Rightarrow$ A simple function w.p. $\geq 1-\varepsilon$.

Håstad ⇒ LMN:

$f \in AC^0(m,d)$

The composition of $d-1$ random restrictions $\equiv R_q$ with $q = p^{d-1}$

w.p. $1-\varepsilon$, $f_{J,z}$ for $(J,z)\sim R_q$ is a DT of depth k,

i.e., $W^{>k}[f_{J,z}] = 0$.

$$\Rightarrow \mathbb{E}\left[W^{>k}[f_{J,z}]\right] \le \varepsilon \Longrightarrow W^{\ge \underbrace{k/q}_{O(\log(m/\varepsilon))^d}}[f] \le 2\varepsilon.$$

How to Prove the Switching Lemma:

Let $\ell \in \mathbb{N}$ think $\ell \approx np$.

R_ℓ - restrictions leaving exactly ℓ vars alive.

$$BAD = \{(J,z) \in R_\ell \mid DT(f_{J,z}) \ge k\}$$

$$E : BAD \xrightarrow{1:1} R_{\ell-k} \times X \qquad |X| \le O(w)^k.$$

$$\Pr_{(J,z)\sim R_\ell}[(J,z) \in BAD] = \frac{|BAD|}{|R_\ell|} \le \frac{|R_{\ell-k}|\cdot|X|}{|R_\ell|} = \frac{\binom{n}{\ell-k}\cdot 2^{n-(\ell-k)}\cdot O(w)^k}{\binom{n}{\ell}\cdot 2^{n-\ell}}$$

$$\approx \left(\frac{\ell}{n}\right)^k O(w)^k$$

Encoding for width-w DNFs, $k=1$.

<u>WTS:</u> $\Pr[DT(f_{J,z}) \geq 1] \leq O\left(\frac{\ell}{n} \cdot w\right)$

$$f = \underbrace{(x_{i_1} \wedge \ldots \wedge \overline{x_{i_w}})}_{T_1} \vee \underbrace{(\wedge \ldots \wedge)}_{T_2} \vee \ldots \vee \underbrace{(x_j)}_{T_i}$$

Let $(J,z) \in BAD$. Then, $f_{J,z}$ is non-constant.

Then $\exists$ a first i s.t. $(T_i)_{J,z} \not\equiv$ false.

since $(J,z) \in BAD$ $\Rightarrow$ $(T_i)_{J,z} \not\equiv$ true.

Find a literal $x_j / \overline{x_j}$ not fixed in T_i.

Assign x_j consistently with its literal in $T_i \Rightarrow (J',z')$

<u>Extra Information:</u> position of $x_j / \overline{x_j}$ in T_i.

Enc$((J,z))$ outputs a pair $((J',z'), \text{index}) \in R_{\ell-1} \times [w]$.

<u>Injective:</u>

Given $((J',z'), \text{index})$, we decode (J,z) as follows:

- Find first term T_i s.t. $(T_i)_{(J',z')} \neq$ false.
- Let $x_j / \overline{x_j}$ the literal at position index inside T_i.
- Make x_j alive $\rightarrow$ (J,z).