

Learning and Fourier Analysis

Grigory Yaroslavtsev

<http://grigory.us>



CIS 625: Computational Learning Theory

Fourier Analysis and Learning

- Powerful tool for PAC-style learning under **uniform distribution** over $\{0,1\}^n$
- Sometimes requires **queries** of the form $f(x)$
- Works for learning many classes of functions, e.g:
 - Monotone, DNF, decision trees, low-degree polynomials
 - Small circuits, halfspaces, k-linear, juntas (depend on small # of variables)
 - Submodular functions (analog of convex/concave)
 - ...
- Can be extended to **product** distributions over $\{0,1\}^n$, i.e. $D = D_1 \times D_2 \times \cdots \times D_n$ where \times means that draws are independent

Boolean Functions

- Book: “Analysis of Boolean Functions”, Ryan O’Donnell
<http://analysisofbooleanfunctions.org>
- $f(x_1, \dots, x_n): \{0,1\}^n \rightarrow \mathbb{R}$
- Notation switch:
 - $0 \rightarrow 1$
 - $1 \rightarrow -1$
- $f': \{-1,1\}^n \rightarrow \mathbb{R}$
- Example:
 - $f(x_1, \dots, x_n) = x_2 \oplus x_3 \oplus x_n = \text{Lin}(x_2, x_3, x_n)$
 - $f'(x_1, \dots, x_n) = x_2 * x_3 * x_n$

Fourier Expansion

- Example:

$$\mathbf{min}(x_1, x_2) = -\frac{1}{2} + \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_1x_2$$

- For $\mathbf{S} \subseteq [n]$ let **character** $\chi_{\mathbf{S}}(x) = \prod_{i \in \mathbf{S}} x_i$
- **Thm:** Every function $f: \{-1, 1\}^n \rightarrow \mathbb{R}$ can be **uniquely** represented as a **multilinear polynomial**

$$f(x_1, \dots, x_n) = \sum_{\mathbf{S} \subseteq [n]} \hat{f}(\mathbf{S}) \chi_{\mathbf{S}}(x)$$

- $\hat{f}(\mathbf{S}) \equiv$ Fourier coefficient of f on \mathbf{S}

Functions = Vectors, Inner Product

- Functions as vectors form a vector space:

$$\mathbf{f}: \{-1,1\}^n \rightarrow \mathbb{R} \Leftrightarrow \mathbf{f} \in \mathbb{R}^{2^n}$$

- Inner product on functions = “correlation”:

$$\begin{aligned} \langle \mathbf{f}, \mathbf{g} \rangle &= 2^{-n} \sum_{x \in \{-1,1\}^n} \mathbf{f}(x) \mathbf{g}(x) \\ &= \mathbb{E}_{x \sim \{-1,1\}^n} [\mathbf{f}(x) \mathbf{g}(x)] \end{aligned}$$

- For $\mathbf{f}: \{-1,1\}^n \rightarrow \{-1,1\}$:

$$\begin{aligned} \langle \mathbf{f}, \mathbf{g} \rangle &= \Pr[\mathbf{f}(x) = \mathbf{g}(x)] - \Pr[\mathbf{f}(x) \neq \mathbf{g}(x)] = \\ &= 1 - 2 \mathit{dist}(\mathbf{f}, \mathbf{g}) \end{aligned}$$

where $\mathit{dist}(\mathbf{f}, \mathbf{g}) = \Pr[\mathbf{f}(x) \neq \mathbf{g}(x)]$

Orthonormal Basis: Proof

- **Thm:** Characters form an orthonormal basis in \mathbb{R}^{2^n}
$$\langle \chi_S, \chi_T \rangle = 1, \quad \text{if } S = T$$
$$\langle \chi_S, \chi_T \rangle = 0, \quad \text{if } S \neq T$$
- $\chi_S(x)\chi_T(x) = \prod_{i \in S} x_i \prod_{j \in T} x_j =$
 $\prod_{i \in S \Delta T} x_i \prod_{j \in S \cap T} x_j^2 = \prod_{i \in S \Delta T} x_i = \chi_{S \Delta T}(x)$
- $\mathbb{E}_x[\chi_S(x)\chi_T(x)] = \mathbb{E}_x[\chi_{S \Delta T}(x)] = \mathbb{E}_x[\prod_{i \in S \Delta T} x_i] =$
 $\prod_{i \in S \Delta T} \mathbb{E}_x[x_i]$
- Since $\mathbb{E}_x[x_i] = 0$ we have:
 - $\mathbb{E}_x[\chi_S(x)\chi_T(x)] = 0$ if $S \Delta T \neq \emptyset$
 - $\mathbb{E}_x[\chi_S(x)\chi_T(x)] = 1$ if $S \Delta T = \emptyset$
- This proves that χ' 's are an orthonormal basis

Fourier Coefficients

- Recall linearity of dot products:

$$- \langle \mathbf{f} + \mathbf{g}, \mathbf{h} \rangle = \langle \mathbf{f}, \mathbf{h} \rangle + \langle \mathbf{g}, \mathbf{h} \rangle$$

$$- \langle \alpha \mathbf{f}, \mathbf{g} \rangle = \alpha \langle \mathbf{f}, \mathbf{g} \rangle$$

- **Lemma:** $\hat{\mathbf{f}}(\mathbf{S}) = \langle \mathbf{f}, \chi_{\mathbf{S}} \rangle$

$$\begin{aligned} \langle \mathbf{f}, \chi_{\mathbf{S}} \rangle &= \left\langle \sum_{T \subseteq [n]} \hat{\mathbf{f}}_T \chi_T, \chi_{\mathbf{S}} \right\rangle = \sum_{T \subseteq [n]} \langle \hat{\mathbf{f}}_T \chi_T, \chi_{\mathbf{S}} \rangle = \\ &= \sum_{T \subseteq [n]} \hat{\mathbf{f}}_T \langle \chi_T, \chi_{\mathbf{S}} \rangle = \hat{\mathbf{f}}(\mathbf{S}) \end{aligned}$$

Parseval's Theorem

- **Parseval's Thm:** For any $f: \{-1,1\}^n \rightarrow \mathbb{R}$

$$\langle f, f \rangle = \mathbb{E}_{x \sim \{-1,1\}^n} [f^2(x)] = \sum_{\mathbf{s} \subseteq [n]} \hat{f}(\mathbf{s})^2$$

If $f: \{-1,1\}^n \rightarrow \{-1,1\}$ then $\sum_{\mathbf{s} \subseteq [n]} \hat{f}(\mathbf{s})^2 = 1$

- **Example:**

$$f = \mathit{min}(x_1, x_2) = -\frac{1}{2} + \frac{1}{2}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_1x_2$$

$$\sum_{\mathbf{s} \subseteq [2]} \hat{f}(\mathbf{s})^2 = 4 \times \frac{1}{4} = 1$$

Plancharel's Theorem

- **Plancharel's Thm:** For any $f, g: \{-1,1\} \rightarrow \mathbb{R}$

$$\langle f, g \rangle = \mathbb{E}_{x \sim \{-1,1\}^n} [f(x)g(x)] = \sum_{S \subseteq [n]} \hat{f}(S) \hat{g}(S)$$

- **Proof:**

$$\begin{aligned} \langle f, g \rangle &= \left\langle \sum_{S \subseteq [n]} \hat{f}(S) \chi_S, \sum_{T \subseteq [n]} \hat{g}(T) \chi_T \right\rangle = \\ &= \sum_{S, T \subseteq [n]} \hat{f}(S) \hat{g}(T) \langle \chi_S, \chi_T \rangle = \sum_{S \subseteq [n]} \hat{f}(S) \hat{g}(S) \end{aligned}$$

Basic Fourier Analysis

- **Mean** $\equiv \mathbb{E}_x[\mathbf{f}(x)] = \langle \mathbf{f}, \mathbf{1} \rangle = \hat{\mathbf{f}}(\emptyset)$
 - For $\mathbf{f}: \{-1, 1\}^n \rightarrow \{-1, 1\}$ we have
$$\mathbb{E}_x[\mathbf{f}(x)] = \Pr[\mathbf{f}(x) = 1] - \Pr[\mathbf{f}(x) = -1]$$
- **Variance** $\equiv \mathbf{Var}(f) \equiv \mathbb{E}_x[(\mathbf{f}(x) - \mathbb{E}_x[\mathbf{f}(x)])^2]$

$$\begin{aligned}\mathbf{Var}(f) &= \mathbb{E}_x[(\mathbf{f}(x) - \mathbb{E}_x[\mathbf{f}(x)])^2] \\ &= \mathbb{E}_x[\mathbf{f}^2(x)] - \mathbb{E}_x[\mathbf{f}(x)]^2 \\ &= \sum_{\mathbf{S} \subseteq [n]} \hat{\mathbf{f}}^2(\mathbf{S}) - \hat{\mathbf{f}}^2(\emptyset) \quad (\text{Parseval's Thm}) \\ &= \sum_{\mathbf{S} \subseteq [n], \mathbf{S} \neq \emptyset} \hat{\mathbf{f}}^2(\mathbf{S})\end{aligned}$$

Convolution

- **Def.:** For $x, y \in \{-1, 1\}^n$ define $x \odot y \in \{-1, 1\}^n$:
$$(x \odot y)_i = x_i y_i$$
- **Def.:** For $f, g: \{-1, 1\}^n \rightarrow \mathbb{R}$ their **convolution**
 $f * g: \{-1, 1\}^n \rightarrow \mathbb{R}$:
$$f * g(x) \equiv \mathbb{E}_{y \sim \{-1, 1\}^n} [f(y) g(x \odot y)]$$
- **Properties:**
 1. $f * g = g * f$
 2. $f * (g * h) = (f * g) * h$
 3. For all $S \subseteq [n]$: $\widehat{f * g}(S) = \widehat{f}(S) \widehat{g}(S)$

Convolution: Proof of Property 3

- **Property 3:** For all $\mathbf{S} \subseteq [n]$: $\widehat{\mathbf{f} * \mathbf{g}}(\mathbf{S}) = \widehat{\mathbf{f}}(\mathbf{S})\widehat{\mathbf{g}}(\mathbf{S})$

- **Proof:**

$$\begin{aligned}\widehat{\mathbf{f} * \mathbf{g}}(\mathbf{S}) &= \mathbb{E}_{x \sim \{-1,1\}^n}[(\mathbf{f} * \mathbf{g})(x)\chi_{\mathbf{S}}(x)] \\ &= \mathbb{E}_x[\mathbb{E}_y[\mathbf{f}(y)\mathbf{g}(x \odot y)]\chi_{\mathbf{S}}(x)] \text{ (def. convolution)} \\ &= \mathbb{E}_y[\mathbf{f}(y)\mathbb{E}_x[\mathbf{g}(x \odot y)\chi_{\mathbf{S}}(x)]] \\ &= \mathbb{E}_y[\mathbf{f}(y)\mathbb{E}_x[\mathbf{g}(x \odot y)\chi_{\mathbf{S}}(x \odot y)\chi_{\mathbf{S}}(x \odot y)\chi_{\mathbf{S}}(x)]] \\ &= \mathbb{E}_y[\mathbf{f}(y)\chi_{\mathbf{S}}(y)\mathbb{E}_x[\mathbf{g}(x \odot y)\chi_{\mathbf{S}}(x \odot y)]] \\ &= \mathbb{E}_y[\mathbf{f}(y)\chi_{\mathbf{S}}(y)\widehat{\mathbf{g}}(\mathbf{S})] \\ &= \mathbb{E}_y[\mathbf{f}(y)\chi_{\mathbf{S}}(y)]\widehat{\mathbf{g}}(\mathbf{S}) \\ &= \widehat{\mathbf{f}}(\mathbf{S})\widehat{\mathbf{g}}(\mathbf{S})\end{aligned}$$

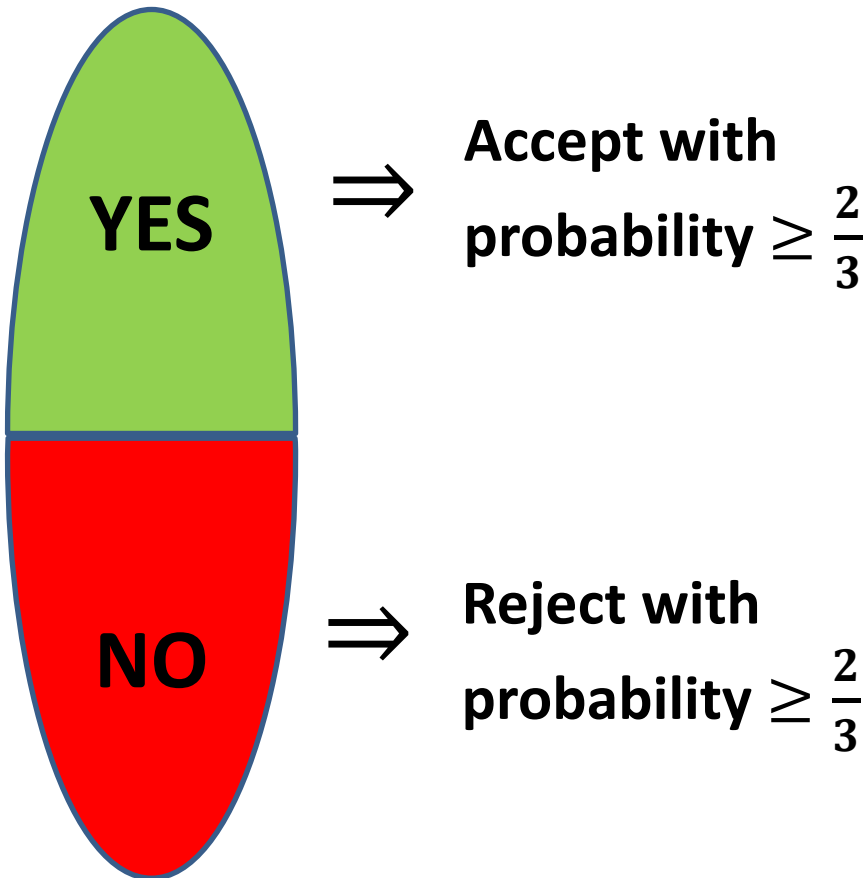
Approximate Linearity

- **Def:** $f: \{0,1\}^n \rightarrow \{0,1\}$ is **linear** if
 - $f(x \oplus y) = f(x) \oplus f(y)$ for all $x, y \in \{0,1\}^n$
 - $\exists \mathcal{S} \subseteq [n]$ such that $f(x) = \bigoplus_{i \in \mathcal{S}} x_i$ for all x
- **Def:** $f: \{0,1\}^n \rightarrow \{0,1\}$ is **approximately linear** if
 1. $f(x \oplus y) = f(x) \oplus f(y)$ for **most** $x, y \in \{0,1\}^n$
 2. $\exists \mathcal{S} \subseteq [n]$ such that $f(x) = \bigoplus_{i \in \mathcal{S}} x_i$ for **most** x
- **Q:** Does **1.** imply **2.**?

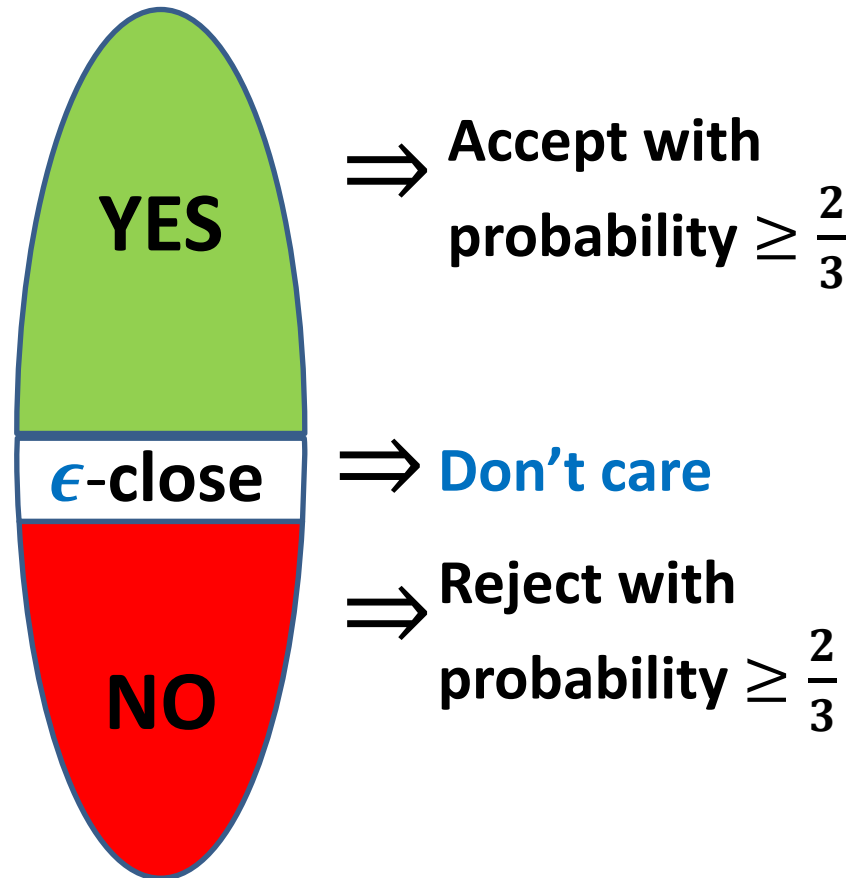
Property Testing

[Goldreich, Goldwasser, Ron; Rubinfeld, Sudan]

Randomized Algorithm



Property Tester

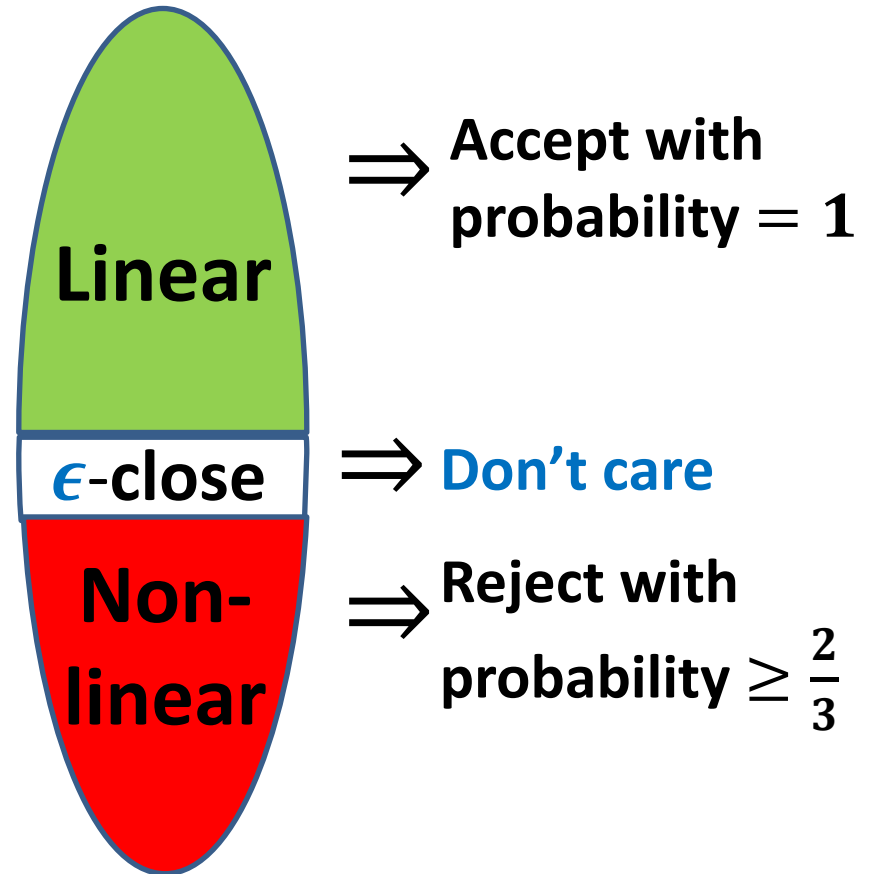


ϵ -close : $\leq \epsilon$ fraction has to be changed to become YES

Linearity Testing

- $f: \{0,1\}^n \rightarrow \{0,1\}$
- \mathcal{P} = class of linear functions
- $dist(f, \mathcal{P}) = \min_{g \in \mathcal{P}} dist(f, g)$
- ϵ -close: $dist(f, \mathcal{P}) \leq \epsilon$

Linearity Tester



Linearity Testing [Blum, Luby, Rubinfeld]

- **BLR Test** (given access to $f: \{0,1\}^n \rightarrow \{0,1\}$):
 - Choose $x \sim \{0,1\}^n$ and $y \sim \{0,1\}^n$ independently
 - Accept if $f(x) \oplus f(y) = f(x \oplus y)$
- **Thm:** If BLR Test accepts with prob. $1 - \epsilon$ then f is ϵ -close to being linear
- **Proof:**
 - Apply notation switch $0 \rightarrow -1, -1 \rightarrow 1$
 - BLR accepts if $f(x)f(y) = f(x \odot y)$
- $\frac{1}{2} + \frac{1}{2} f(x)f(y)f(x \odot y) = 1$ if $f(x)f(y) = f(x \odot y)$
- $\frac{1}{2} + \frac{1}{2} f(x)f(y)f(x \odot y) = 0$ if $f(x)f(y) \neq f(x \odot y)$

Linearity Testing: Analysis

$$\begin{aligned}1 - \epsilon &= \Pr[BLR \text{ Accepts } f] \\&= E_{x,y} \left[\frac{1}{2} + \frac{1}{2} f(x)f(y)f(x \odot y) \right] \\&= \frac{1}{2} + \frac{1}{2} E_{x,y} [f(x)f(y)f(x \odot y)] \\&= \frac{1}{2} + \frac{1}{2} E_x [f(x) E_y [f(y)f(x \odot y)]] \\&= \frac{1}{2} + \frac{1}{2} E_x [f(x) \cdot (f * f)(x)] \text{ (def. of convolution)} \\&= \frac{1}{2} + \frac{1}{2} \sum_{S \subseteq [n]} \hat{f}(S) (\widehat{f * f})(S) \text{ (Plancharel's thm)} \\&= \frac{1}{2} + \frac{1}{2} \sum_{S \subseteq [n]} \hat{f}^3(S)\end{aligned}$$

Linearity Testing: Analysis Continued

- $1 - 2\epsilon = \sum_{\mathbf{S} \subseteq [n]} \hat{f}^3(\mathbf{S})$
- $\leq \max_{\mathbf{S} \subseteq [n]} \{\hat{f}(\mathbf{S})\} \sum_{\mathbf{S} \subseteq [n]} \hat{f}^2(\mathbf{S})$
- $= \max_{\mathbf{S} \subseteq [n]} \{\hat{f}(\mathbf{S})\}$
- Recall $\hat{f}(\mathbf{S}) = \langle \mathbf{f}, \chi_{\mathbf{S}} \rangle = 1 - 2\text{dist}(\mathbf{f}, \chi_{\mathbf{S}})$
- Let $\mathbf{S}^* = \underset{\mathbf{S} \subseteq [n]}{\text{argmax}} \{\hat{f}(\mathbf{S})\}$: $1 - 2\epsilon \leq 1 - 2\text{dist}(\mathbf{f}, \chi_{\mathbf{S}^*})$
- $\text{dist}(\mathbf{f}, \chi_{\mathbf{S}^*}) \leq \epsilon$

Local Correction

- Learning linear functions takes n queries
- **Lem:** If f is ϵ -close to linear function χ_S then for every x one can compute $\chi_S(x)$ w.p. $1 - 2\epsilon$ as:
 - Pick $y \sim \{0,1\}^n$
 - Output $f(y) \oplus f(x \oplus y)$

- **Proof:**

$$\Pr[f(y) \neq \chi_S(y)] = \Pr[f(x \oplus y) \neq \chi_S(x \oplus y)] = \epsilon$$

By union bound:

$$\Pr[f(y) = \chi_S(y), f(x \oplus y) = \chi_S(x \oplus y)] \geq 1 - 2\epsilon$$

$$\text{Then } f(y) \oplus f(x \oplus y) = \chi_S(y) \oplus \chi_S(x \oplus y) = \chi_S(x)$$

Thanks!