Lecture 11: Fourier Basics for Boolean functions. Linearity testing.

Lecturer: Ronitt Rubinfeld

Spring 2014

6.842: Randomness and Computation

Why all the fuss about Boolean functions?

- Truth table of a function (complexity theory)
- Concept to be learned (machine learning)
- Subset of the Boolean cube (coding theory, combinatorics,…)
- Etc.

Why Fourier/Harmonic Analysis?

- Study “structural properties” of Boolean functions
  - Low complexity
  - Depends on few inputs (dictator, junta)
  - “fair” (no variable has too much influence)
  - Homomorphism
  - Spread out/concentrated

The Boolean function

A “new” representation!

-

$$\cancel{f:\{0,1\}^n \rightarrow \{0,1\}}$$

$$\cancel{(x_1, x_2, \ldots, x_n) \oplus (y_1, y_2, \ldots, y_n) = (x_1 \oplus y_1, \ldots, x_n \oplus y_n)}$$

$$f:\{\pm 1\}^n \rightarrow \{\pm 1\}$$

$$(x_1, x_2, \ldots, x_n) \odot (y_1, y_2, \ldots, y_n) = (x_1 \cdot y_1, \ldots, x_n \cdot y_n)$$

The slick (notational) trick:

- $$0 \to +1$$
  $$1 \to -1$$

$$\begin{array}{c|cc} \oplus & 0 & 1 \\ \hline 0 & 0 & 1 \\ 1 & 1 & 0 \end{array} \quad \to \quad \begin{array}{c|cc} \times & +1 & -1 \\ \hline +1 & +1 & -1 \\ -1 & -1 & +1 \end{array}$$

The set of functions and inner product

- $G = \{g | g: \{\pm 1\}^n \to \mathbb{R}\}$ (all *n*-bit fctns into Reals)
  - A vector space of dimension 2^n
    - For any set of basis functions of size 2^n, every $g \in G$ is a linear combination of basis functions.
  - Which basis to use?

Which basis?

- $G = \{g | g: \{\pm 1\}^n \to \mathbb{R}\}$ (all *n*-bit fctns into Reals)
  - A "natural" basis: indicator functions
    - $e_a(x) = \begin{cases} 1 \; if \; x = a \\ 0 \;\; o.w. \end{cases}$
    - Orthonormal
    - Used to describe function via "truth table"

$$f(x) = \Sigma_a f(a) e_a(x)$$

A very useful basis:

- $G = \{g | g: \{\pm 1\}^n \to \mathbb{R}\}$ (all *n*-bit fctns into Reals)
  - Parity functions
    - For S $\subseteq [n]$, $\chi_S(x) = \prod_{i \in S} x_i$
    - Let's agree that $\chi_\emptyset(x) = 1 \; \forall x$

A useful property:

- Fact 0: $\chi_S(x) \cdot \chi_T(x) = \chi_{S \Delta T}(x)$

Proof: $\chi_S(x) \cdot \chi_T(x) = \prod_{i \in S} x_i \ \prod_{j \in T} x_j$

$$= \prod_{S \cap T} \underbrace{x_i^2}_{=1} \prod_{i \in S \Delta T} x_i$$

Inner product

correlation – not same as in probability theory

- $< f, g > = \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} f(x) g(x)$

- Note:

$< \chi_S, \chi_S > = \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} \left(\chi_{S(x)}\right)^2 = 1$

Always 1

Orthogonal:

$S \Delta T$
is nonempty since $S \neq T$
Pick $j \in S \Delta T$

- If $S \neq T$:

$$< \chi_s , \chi_T > = \frac{1}{2^n} \Sigma_x \chi_S(x) \chi_T(x)$$

$$= \frac{1}{2^n} \Sigma_x \chi_{S \Delta T}(x)$$

$$= \frac{1}{2^n} \Sigma_x \prod_{i \in S \Delta T} x_i$$

$x^{\oplus j} = x$ with j[th] bit flipped

$$= \frac{1}{2^n} \Sigma_{pairs\ x, x^{\oplus j}} \left(\prod_{i \in S \Delta T} x_i + \prod_{i \in S \Delta T} (x^{\oplus j})_i\right)$$

= 0 since each pair sums to 0:

$$x_j \left(\Pi_{(i \in S \Delta T \setminus \{j\})}\ x_i\right) - x_j \left(\Pi_{(i \in S \Delta T \setminus \{j\})}\ x_i\right) = 0$$

So we have an orthonormal basis!

- Every function can be written as a linear combination of these χ_S 's

Fourier Coefficients

- Theorem:
  $\forall f, f(x) = \Sigma_S \hat{f}(S)\chi_S(x)$ where

$$\hat{f}(S) = < f, \chi_S > = \frac{1}{2^n} \Sigma_{x\in\{\pm 1\}^n} f(x)\chi_S(x)$$

Some examples:

Function	Fourier Representation
$f(x)=1 = \chi(\emptyset)$	1
$f(x)= x_i = \chi(\{i\})$	x_i
$f(x)= AND(x_1 , x_2)$	$½ + ½\, x_1 + ½\, x_2 - ½\, x_1x_2$

Fourier coefficients of parity functions:

- Fact 1: f is a parity function
  iff $f = \chi_S(x)$
  iff (1) $\hat{f}(S) = 1$ and
  (2) for all $T \neq S$,
  $\hat{f}(T) = < \chi_S, \chi_T > = 0$

By orthogonality

Agreement with parity function vs. max Fourier coefficient

- Fact 2: $\hat{f}(S) = 1 - 2 \Pr_{x \in \pm 1^n} [f(x) \neq \chi_S(x)]$

Proof:

$$\hat{f}(\mathrm{S}) = \frac{1}{2^n} \Sigma_x f(x) \chi_S(x)$$

$$= \frac{1}{2^n} \Sigma_{\mathrm{x\ s.t.\ f(x)} = \chi_{S(\mathrm{x})}} (+1) + \frac{1}{2^n} \Sigma_{x\ s.t. f(x) \neq \chi_{S(x)}} (-1)$$

$$= (1 - \Pr_{x \in \pm 1^n} [f(x) \neq \chi_S(x)]) - \Pr_{x \in \pm 1^n} [f(x) \neq \chi_S(x)]$$

Distance between parity functions

- Fact 3: if $S \neq T$ then $\Pr_{x \in \{\pm 1\}^n}[\chi_S(x) = \chi_T(x)] = 1/2$

Proof: Let $f = \chi_T$, then

$$\hat{f}(S) = 0 \ \text{(fact 1)}$$

$$= 1 - 2\Pr[\chi_T(x) \neq \chi_S(x)] \ \text{(fact 2)}$$

Plancherel's Theorem

- Theorem: For $f, g: \{\pm 1\}^n \to \Re$ we have

$$< f, g > \equiv E_{\{\pm 1\}^n}[f(x)g(x)] = \Sigma_{S \subseteq [n]} \hat{f}(S) \cdot \hat{g}(S)$$

Proof:

$$< f, g > = < \Sigma_S \hat{f}(S)\, \chi_s , \Sigma_T \hat{g}(T)\chi_T > \qquad \text{(def)}$$

$$= \Sigma_s \Sigma_T \hat{f}(S)\, \hat{g}(T) < \chi_s, \chi_T > \qquad \text{(bilinearity)}$$

$$= \Sigma_S \hat{f}(S)\, \hat{g}(S) \qquad \text{(orthogonality)}$$

Parseval's Theorem

- Corollary: For $f:\{\pm 1\}^n \to \Re$ we have
  $< f, f > \equiv E_{\{\pm 1\}^n}[f^2(x)] = \Sigma_{S\subseteq[n]} \hat{f}(S)^2$

Boolean Parseval's: For $f:\{\pm 1\}^n \to \{\pm 1\}$

$\Sigma_{S\subseteq[n]} \hat{f}(S)^2 = E_{\{\pm 1\}^n}[f^2(x)] = 1$

=1 for all x

More useful facts:

Plancherel

- Fact 4: $E[f] = E\left[f(x) \cdot 1\right] = E\left[f(x)\chi_\phi(x)\right]$
  $= \Sigma \hat{f}(S)\hat{\chi}_\phi(S) = \hat{f}(\phi) \cdot \hat{\chi}_\phi(\phi) = \hat{f}(\phi)$

we know these

Fact 5: (corollary to fact 4 and to fact 1)

$$E[\chi_S(x)] = \begin{cases} 1 \; if \; S = \phi \\ 0 \; o.w. \end{cases}$$

Linearity (homomorphism) testing

$$\forall x,y \;\; f(x) + f(y) = f(x+y)$$

Linearity Property

- Want to quickly test if a function over a group is linear , that is

$$\forall x,y \;\; f(x) + f(y) = f(x+y)$$

  - Useful for
    - Checking correctness of programs computing matrix, algebraic, trigonometric functions
    - Probabilistically Checkable Proofs

      Is the proof of the right format?
  - In these cases, enough for f to be close to homomorphism

What do we mean by ``close''?

Definition: f, over domain of size N,

is ε-close to linear if can change at most εN values to turn it into one.

Otherwise, ε-far.

What do we mean by ``quick''?

- query complexity measured in terms of domain size N

- Our goal (if possible):
  - constant independent of N?

Linearity Testing

- If f is linear (i.e., $\forall x,y \;\; f(x) + f(y) = f(x+y)$) then test should PASS with probability >2/3
- If f is ε-far from linear then test should FAIL with probability >2/3
- Note: If f not linear, but ε-close, then either output is ok

Linearity Testing for f: $GF(2)^n \to GF(2)$

- $x = (x_1, \ldots, x_n), y = (y_1, \ldots, y_n) \in \{0,1\}^n$
  - $x + y = (x_1 \oplus y_1, \ldots, x_n \oplus y_n)$ ($\oplus$ is "xor")
- $\forall x, y\ f(x) \oplus f(y) = f(x + y)$
- Linear functions are exactly $\{f_a \mid f_a(x) = \Sigma\, a_i \cdot x_i \bmod 2 \text{ for } a \in \{0,1\}^n\}$

Linearity Testing for f: $\{\pm 1\}^n \rightarrow \{\pm 1\}$

- $x = (x_1, \dots, x_n), y = (y_1, \dots, y_n) \in \{\pm 1\}^n$
  - $x \odot y = (x_1 \cdot y_1, \dots, x_n \cdot y_n)$
- $\forall x, y \quad f(x) \cdot f(y) = f(x \odot y)$
- Linear functions are exactly the parity functions $\{\chi_S\}$

Proposed Tester:

- Repeat $r = O(\frac{1}{\rho})$ times:
  - Pick $x, y \in_R \{0,1\}^n$
  - If $f(x)f(y) \neq f(x \odot y)$ output "fail" and halt
- Output "pass"

- Easy to see:
  - If f is linear, then tester passes with probability 1
  - If f is such that $\Pr_{x,y}[f(x)f(y) \neq f(x \odot y)] \geq \rho$ then (constant in O notation can be chosen so that) tester fails with probability at least 2/3

Characterizing "close" to linear

- Suppose $\Pr_{x,y}[f(x)f(y) \neq f(x \odot y)]$ is small... is f close to linear?

Nontriviality [Coppersmith]:

- $f: Z_{3k} \rightarrow Z_{3k-1}$
- $f(3h+d)=h, \text{ for } h < 3^k, \; d \in \{-1,0,1\}$

- f satisfies $f(x)+f(y) \neq f(x+y)$ for only 2/9 of choices of x,y (*i.e.* $\delta_f = 2/9$)
- f is 2/3-far from a linear!

Our goal:

- Theorem: If f is $\epsilon-$ far from linear, then

$$\Pr_{x,y}\underbrace{[f(x)f(y) \neq f(x \odot y)]}_{\boxed{\Pr_{x,y}[f(x)f(y)f(x \odot y) \neq 1]}} \geq \epsilon$$

Call this δ

Main Lemma:

$$1-\delta \equiv \Pr_{x,y}[f(x)f(y)f(x \odot y) = 1] = \frac{1}{2} + \frac{1}{2}\,\Sigma_{S\subseteq[n]}\,\hat{f}(S)^3$$

Lemma → Theorem

- Theorem: If f is $\epsilon-$far from linear, then
$$\Pr_{x,y}[f(x)f(y)f(x \odot y) \neq 1] \geq \epsilon$$

$\equiv \delta$

Proof:

Main Lemma implies $1 - \delta \leq \frac{1}{2} + \frac{1}{2}\, \Sigma_{S \subseteq [n]} \hat{f}(S)^3$

So $\quad 1 - 2\delta \leq \Sigma\, \hat{f}(S)^3$

$$\leq \max_S (\hat{f}(S))\, \Sigma\, \hat{f}(S)^2$$

=1 by Boolean Parseval

$$\leq \max_S (\hat{f}(S))$$

Pick T to maximize

$$\leq \hat{f}\ (T)$$

$$\leq 1 - 2\Pr[f(x) \neq \chi_T(x)]$$

Fact 2

So $\delta \geq \Pr[f(x) \neq \chi_T(x)] \geq \epsilon$

Before the main lemma:

- $\frac{1+f(x)f(y)f(x\odot y)}{2}\begin{cases}= 1\ if\ x,y\ PASS \\ = 0\ if\ x,y\ FAIL\end{cases}$

Indicator variable describing result of test!

Main Lemma:

$$1-\delta \equiv \Pr_{x,y}[f(x)f(y)f(x\odot y)=1] = \frac{1}{2}+\frac{1}{2}\Sigma_{S\subseteq[n]}\hat{f}(S)^3$$

- Proof: $1-\delta = E_{x,y}\left[\frac{1+f(x)f(y)f(x\odot y)}{2}\right]$

$$= \frac{1}{2}+\frac{1}{2}\underbrace{E_{x,y}[f(x)f(y)f(x\odot y)]}_{\text{Focus here}}$$

$$E_{x,y}[f(x)f(y)f(x\odot y)]$$
$$= E\left[\left(\Sigma_S \hat{f}(S)\chi_S(x)\right)\left(\Sigma_T \hat{f}(T)\chi_T(y)\right)\left(\Sigma_U \hat{f}(U)\chi_U(x\odot y)\right)\right.$$
$$= \Sigma_{S,T,U}\ \hat{f}(S)\hat{f}(T)\hat{f}(U)\underbrace{E[\chi_S(x)\,\chi_T(y)\chi_U(x\odot y)]}_{\text{What is this?}}$$

A final calculation:

- $E[\chi_S(x)\,\chi_T(y)\chi_U(x \odot y)]$

$$= E[\,\Pi_{i\in S}\, x_i\, \Pi_{j\in T}\, y_j\, \Pi_{k\in U}\,(x_k \cdot y_k)]$$
$$= E[\Pi_{i\in S\,\triangle\, U} x_i\, \Pi_{j\in T\,\triangle\, U} y_j]$$
$$= \underbrace{E[\Pi_{i\in S\,\triangle\, U} x_i\,]}_{\substack{1 \text{ if } S\,\triangle\, U = \phi \\ 0 \text{ o.w.}}}\underbrace{E[\Pi_{j\in T\,\triangle\, U} y_j]}_{\substack{1 \text{ if } T\,\triangle\, U = \phi \\ 0 \text{ o.w.}}}$$

= 1 if S=T=U and 0 otherwise

Main Lemma:

$1 - \delta \equiv$

$$\Pr_{x,y}[f(x)f(y)f(x \odot y) = 1] = \frac{1}{2} + \frac{1}{2}\Sigma_{S \subseteq [n]} \hat{f}(S)^3$$

- Proof: $1 - \delta = E_{x,y}\left[\frac{1 + f(x)f(y)f(x \odot y)}{2}\right]$

$$= \frac{1}{2} + \frac{1}{2}\underbrace{E_{x,y}[f(x)f(y)f(x \odot y)]}_{\text{Focus here}}$$

$$E_{x,y}[f(x)f(y)f(x \odot y)]$$
$$= E\left[\left(\Sigma_S \hat{f}(S)\chi_S(x)\right)\left(\Sigma_T \hat{f}(T)\chi_T(y)\right)\left(\Sigma_U \hat{f}(U)\chi_U(x \odot y)\right)\right]$$
$$= \Sigma_{S,T,U}\ \hat{f}(S)\hat{f}(T)\hat{f}(U)E[\chi_S(x)\ \chi_T(y)\chi_U(x \odot y)]$$
$$= \Sigma_S \hat{f}(S)^3$$

1 if S=T=U

0 otherwise

Linearity tests over other domains

- Still constant, even for general nonabelian groups
- Slightly weaker relationship between parameters

Self-correction

- Given program P computing linear f that is correct on at least 7/8 of the inputs (BUT YOU DON'T KNOW WHICH ONES!)
- Can you correctly compute f on each input?
  - To compute f(x), can't just call P on x...

Self-corrector:

- Repeat $r = O(\frac{1}{\rho})$ times:
  - Pick $y \in_R \{0,1\}^n$
  - Let $\text{guess}(x) \leftarrow P(y) \cdot P(x \odot y)$
- Output most common guess

- If P correct on both calls, then guess is correct
- What is probability of this?
  - Observe: Since y uniformly distributed, so is $x \odot y$
  - $\Pr[P\ wrong\ on\ either\ y\ or\ x \odot y] \leq \frac{1}{4}$