COMS E6998-9: Algorithms for Massive Data (Spring'19) April 2, 2019

Lecture 19: Linearity Testing

Instructor: *Alex Andoni* Scribes: *Samrat Phatale*

1 Linearity Testing

A function $f : \{\pm 1\}^n \to \pm 1$, is linear *iff*

$$f(x.y) = f(x).f(y) \qquad \Longrightarrow T_{xy}$$

$$\text{Where } x.y = \sum_{i=1}^{n} x_i y_i$$

Linearity Test

Pick x, y at random and check T_{xy}.

Observation: If f is ϵ far from linear if T_{xy} fails in at least $\Omega(\frac{1}{\epsilon})$ tests.

Claim 1. *if f is ϵ far from linear then*

$$Pr_{x,y}[T_{xy}\, fails] \geq \epsilon$$

Let's look at a tool that will help us do this efficiently

Tool: Fourier Analysis over hyper-cube $\{\pm 1\}^n$

Let $\mathcal{F}$ = set of all functions $f : \{\pm 1\}^n \to \pm 1$
$\mathcal{F}$ represents a vector space of 2^n dimensions, in which every function f is a vector of length 2^n.
Now let's try and find a basis of $\mathcal{F}$.
Basis of $\mathcal{F} = \{f_z\}_{z \in \{\pm 1\}^n}$

$$f_z(x) = \begin{cases} 1 & \text{if x=z} \\ 0 & \text{otherwise} \end{cases}$$

We can see that this is a minimal basis because we can't write any of the f_z's as a linear combination of other f_z's.
Now $\forall f \in \mathcal{F}$, $\exists$ coefficients $\{\alpha_z\}_{z \in \{\pm 1\}^n}$, such that

$$f = \sum_z \alpha_z f_z$$

where $\alpha_z = f_z(x) \forall x \in \{\pm 1\}^n$.
So we have

$$f(x) = \sum_z f_z(x)\, f_z$$

Now let's take a look at another basis for $\mathcal{F}$, the Fourier basis,

$$\chi_S \in \mathcal{F} \quad \text{where} \quad S \subseteq [n]$$

$$\chi_S(x) = \prod_{i \in S} x_i$$

$$\chi_\phi(x) = 1$$

Claim 2. *All χ_S's are linear.*

Proof.

$$\chi_S(x.y) = \prod_{i \in S} x_i y_i$$

$$\chi_S(x.y) = \prod_{i \in S} x_i \prod_{i \in S} y_i$$

$$\chi_S(x.y) = \chi_S(x).\chi_S(y)$$

□

Now let's define inner product in this space.

$$< f, g >= \mathbb{E}_{x \in \{\pm 1\}^n} f(x).g(x)$$

Points to note

1. All basis elements have 'norm' = 1.

$$< \chi_S, \chi_S >= \mathbb{E}_x[chi_S(x) chi_S(x)] = \mathbb{E}_x[\prod_{i \in S} x_i \prod_{i \in S} x_i] = \mathbb{E}_x[1] = 1$$

2. All basis elements are normal to each other, i.e. $\forall S \neq T$, $< \chi_S, \chi_T >= 0$

$$< \chi_S, \chi_T >= \mathbb{E}_x[chi_S(x) chi_T(x)] = \mathbb{E}_x[\prod_{i \in S} x_i \prod_{i \in T} x_i]$$

For i that belong to both S and T, $\mathbb{E}x_i = 1$, since they will be same so,

$$= \mathbb{E}_x[\prod_{i \in S \triangle T} x_i]$$

Since all x_i are independent of each other,

$$= \prod_{i \in S \triangle T} \mathbb{E}_x[x_i] = 0$$

So, χ_S's form an ortho-normal basis.

Fourier Decomposition

Since we have a orthonormal basis, we can decompose any given function as a linear combination of all possible linear functions χ_S.

$$\forall f : \{\pm 1\}^n \rightarrow \pm 1 \quad \exists \quad \{\hat{f}_S\}_{S \subseteq [n]}$$

such that

$$f = \sum_{S \subseteq [n]} \hat{f}_S \, \chi_S$$

Theorem 3. ***Plancherel's equality:***

$$< f, g >= \sum_{S \subseteq [n]} \hat{f}_S . \hat{g}_S$$

This follows intuitively from the fact that $\hat{f}_S$ and $\hat{g}_S$ are coefficients of the underlyting basis vectors.

Theorem 4. ***Parseval's equality:***

$$< f, f >= \sum_{S \subseteq [n]} \hat{f}_S . \hat{f}_S = 1$$

Example 5. ***Examples of Fourier Decomposition:***

1. $f(x) = 1$
   $\hat{f}_\phi = 1, \forall S \neq \phi \hat{f}_S = 0$

2. $f(x) = x_i$
   $\hat{f}_{\{x_i\}} = 1, \hat{f}_{else} = 0$

3. $f(x) = \chi_S(x)$
   $\hat{f}_S = 1, \ \forall T \neq S, \ \hat{f}_T = 0$

4. $f(x) = AND(x_1, x_2) = \begin{cases} -1 & if x_1 = x_2 = -1 \\ 1 & otherwise \end{cases}$
   $f(x) = \frac{1}{2} + \frac{1}{2}\chi_{\{1\}} + \frac{1}{2}\chi_{\{2\}} - \frac{1}{2}\chi_{\{1,2\}}$

Observation 6. *How to compute $\hat{f}_S$ from f*

$\hat{f}_S$ is just a projection of f along the basis vector χ_S

$$\hat{f}_S =< f, \chi_S >= \mathbb{E}_x[f(x).\chi_S(x)]$$

2 Back to Testing Linearity

Fact 7. $\{\chi_S\}_{S\subseteq[n]}$ *are all possible linear functions.*

Let f be ϵ far from linearity.
If f is linear $\implies$, $\exists S \subseteq [n]$ such that $f = \chi_S$.
If f is ϵ far from linearity, then

$$\forall \chi_S \;\; Pr[f(x) = \chi_S(x)] \leq 1 - \epsilon$$

Claim 8. $\forall f : \{\pm 1\}^n \to \pm 1$, *that are* ϵ *far from linearity,* $\forall S \subseteq [n]$, $\hat{f}_S \leq 1 - 2\epsilon$.

Proof.

$$\hat{f}_S = < f, \chi_S > = \mathbb{E}_x[f(x).\chi_S(x)]$$

$$= Pr[f(x) = \chi_S(x)](+1) + Pr[f(x) \neq \chi_S(x)](-1)$$

$$\leq 1 - \epsilon - \epsilon = 1 - 2\epsilon$$

Hence proved.

□

Observation 9. *By Parseval's Equality, we have* $\sum_S \hat{f}_S = \mathbb{E}_x f(x)^2 = 1$

Theorem 10. $Pr_{x,y}[T_{xy}\, \textit{fails}] \geq \epsilon$

Proof.

Observation 11.

$$T_{x,y} \textit{ succeeds} \iff f(x.y) = f(x).f(y)$$

$$\iff f(x.y)f(x)f(y) = 1$$

Let, $\delta = Pr[T_{xy}\, \text{succeeds}]$

$$\delta = Pr_{x,y}[f(x.y)f(x)f(y) = 1]$$

$$\mathbb{E}_{x,y}[f(x.y)f(x)f(y)] = \delta(+1) + (1-\delta)(-1)$$

$$= 2\delta - 1$$

So,

$$\delta = \frac{1}{2} + \frac{1}{2}\mathbb{E}_{x,y}[f(x.y)f(x)f(y)]$$

$$\delta = \frac{1}{2} + \frac{1}{2}\mathbb{E}_{x,y}[(\sum_S \hat{f}_S \chi_S(x.y))(\sum_T \hat{f}_T \chi_T(x))(\sum_U \hat{f}_U \chi_U(y))]$$

Since χ_S is linear, we have $\chi_S(x.y) = \chi_S(x).\chi_S(y)$.

$$= \frac{1}{2} + \frac{1}{2}\mathbb{E}_{x,y}[\sum_{S,T,U} \hat{f}_S \hat{f}_T \hat{f}_U \chi_S(x)\chi_S(y)\chi_T(x)\chi_U(y)]$$

$$= \frac{1}{2} + \frac{1}{2}\sum_{S,T,U} \mathbb{E}_{x,y}[\hat{f}_S \hat{f}_T \hat{f}_U \chi_S(x)\chi_S(y)\chi_T(x)\chi_U(y)]$$

$$= \frac{1}{2} + \frac{1}{2} \sum_{S,T,U} \hat{f}_S \hat{f}_T \hat{f}_U \mathbb{E}_x[\chi_S(x)\chi_T(x)]\mathbb{E}_y[\chi_S(y)\chi_U(y)]$$

if $S = T = U$,

$$= \frac{1}{2} + \frac{1}{2} \sum_S \hat{f}_S^3$$

$$\leq \frac{1}{2} + \frac{1}{2} \sum_S \hat{f}_S^2(1-2\epsilon)$$

$$= \frac{1}{2} + \frac{1}{2}(1-2\epsilon) = 1-\epsilon$$

□

Now we have $Pr[T_{xy}\text{ succeeds}] \leq 1-\epsilon$ and $Pr[T_{xy}\text{ fails}] \geq \epsilon$.

Linearity Testing Algorithm:

1. Draw x, y iid and test T_{xy} for $\mathcal{O}(\frac{1}{\epsilon})$ times.
2. If one test fails, f is not linear. If all pass, f is at least *epsilon* close to linear.

3 Locally Decodable Code

Encoding,

$$C : \{0,1\}^n \to \{0,1\}^m \qquad m > n$$

Decoding,

$$D : \{0,1\}^m \to \{0,1\}^n$$

1. $\forall X \in \{0,1\}^n$, $Y \in \{0,1\}^m$, such that $||y||_1 \leq \epsilon m$
$$D(C(X)+Y) = X$$
2. For any i $\exists$ a procedure (randomized), that queries q positions of $C(X)+Y$ and outputs x_i with $\geq 90\%$
if $q = 1$, impossible
if $q = 2$, $m = 2^{\mathcal{O}(n)}$ possible
if $q = 2$, $m = 2^{n^{\mathcal{O}(1)}}$ possible
$\vdots$
if $q = (\log n)^{\frac{1}{\epsilon}}$, $m = \mathcal{O}(n^{1+O(\epsilon)})$ possible.
There is a trade off between number of queries required and the blowup required to reconstruct the message.