6.895 Randomness and Computation February 21, 2006

Lecture 4

Lecturer: Ronitt Rubinfeld *Scribe: Daniel Dumitran*

1 Fourier Representation

Let us consider the functions $f : \{\pm 1\}^n \to \{\pm 1\}$ and the following inner product $\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} f(x)g(x)$.

Definition 1 *Let $S \in \{\pm 1\}^n$. We define $\chi_S : \{\pm 1\}^n \to \{\pm 1\}$ with $\chi_S(x) = \prod_{i \ st \ s_i = -1} x_i$. Note that throughout the lectures, S is sometimes used as a subset of $[n]$ instead of a vector. If $S \subseteq [n]$ then $\chi_S(x) = \prod_{i \in S} x_i$.*

Notice that functions χ_S form an orthonormal basis under inner product $\langle\rangle$ (i.e. $\langle \chi_S, \chi_T \rangle = \delta_{S,T}$ [1]).

Definition 2 $\forall S \in \{\pm 1\}^n$ *we define* $\hat{f}(S) = \langle f, \chi_S \rangle = \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} f(x)\chi_S(x)$

Theorem 3 $\forall f$ *we have* $f(x) = \frac{1}{2^n} \sum_{z \in \{\pm 1\}^n} \hat{f}(x)\chi_z(x)$

Remark f linear $\Leftrightarrow \exists S \in \{\pm 1\}^n$ st $\forall T \in \{\pm 1\}^n$ we have $\hat{f}(T) = \delta_{S,T}$.

Definition 4 $dist(f, g) = Pr_{x \in \{\pm 1\}^n}[f(x) \neq g(x)]$

Lemma 5 $\forall S \in \{\pm 1\}^n$ *and* $f : \{\pm 1\}^n \to \{\pm 1\}$, *we have* $\hat{f}(S) = 1 - 2 * dist(f, \chi_S)$

Proof of Lemma 5:

$$\begin{aligned} \hat{f}(S) &= \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} f(x)\chi_S(x) \\ &= \frac{1}{2^n}[\sum_{x \ st \ f(x) = \chi_S(x)} f(x)\chi_S(x) + \sum_{x \ st \ f(x) \neq \chi_S(x)} f(x)\chi_S(x)] \end{aligned}$$

However, $f(x)\chi_S(x)$ is 1 for all terms in the first sum and -1 for all terms in the second sum. Therefore

$$\begin{aligned} \hat{f}(S) &= \frac{1}{2^n}[2^n - 2 \sum_{x \ st \ f(x) \neq \chi_S(x)} -1] \\ &= 1 - 2Pr[f(x) \neq \chi_S(x)] \\ &= 1 - dist(f, \chi_S(x)) \end{aligned}$$

■

[1] $\delta(S, T)$ is 1 if $S = T$ and 0 otherwise.

Let $S \neq T$ be two elements in $\{\pm 1\}^n$. We have

$$\begin{aligned} dist(\chi_S, \chi_T) &= \frac{1 - \hat{\chi_T}(S)}{2} \\ &= \frac{1 - \langle \chi_S, \chi_T \rangle}{2} \\ &= \frac{1}{2} \end{aligned}$$

What this tells us is that two different linear functions agree on EXACTLY half of their inputs.

2 Parseval's Identity (for Boolean functions only)

Lemma 6 *$\forall f : \{\pm 1\}^n \rightarrow \{\pm 1\}$ we have* $\sum_{S \in \{\pm 1\}^n} [\hat{f}(S)]^2 = 1$.
(For the general case, we have $< f, f >= \sum_{S \in \{\pm 1\}^n} [\hat{f}(S)]^2$*.)*

We are going to prove the lemma for Boolean functions only.
Proof of Lemma 6: If f is Boolean, we have $< f, f >= 1$ because

$$\begin{aligned} < f, f > &= \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} f^2(x) \\ &= \frac{1}{2^n} \sum_{x \in \{\pm 1\}^n} 1 \\ &= \frac{1}{2^n} 2^n \\ &= 1 \end{aligned}$$

However, we also have

$$\begin{aligned} < f, f > &= < \sum_{S \in \{\pm 1\}^n} \hat{f}(S)\chi_S, \sum_{T \in \{\pm 1\}^n} \hat{f}(T)\chi_T > \\ &= \sum_{S,T} \hat{f}(S)\hat{f}(T) < \chi_S, \chi_T > \\ &= \sum_{S,T} \hat{f}(S)\hat{f}(T)\delta_{S,T} \\ &= \sum_{S} [\hat{f}(S)]^2 * 1 \\ &= \sum_{S} [\hat{f}(S)]^2 \end{aligned}$$

Therefore $\sum_{S} [\hat{f}(S)]^2 = 1$ ■

3 More Linearity Testing

We have

$$\begin{aligned} f(xy) = f(x)f(y) \quad &\Leftrightarrow \quad f(x)f(y)f(xy) = 1 \\ &\Leftrightarrow \quad \frac{1 - f(x)f(y)f(xy)}{2} = 0 \end{aligned}$$

and

$$\begin{aligned} f(xy) \neq (x)f(y) \quad &\Leftrightarrow \quad f(x)f(y)f(xy) = -1 \\ &\Leftrightarrow \quad \frac{1 - f(x)f(y)f(xy)}{2} = 1. \end{aligned}$$

It is therefore a natural choice to use the indicator variable $I[\frac{1-f(x)f(y)f(xy)}{2}]$ in order to measure the probability of group law failure of a function f.

$$\begin{aligned} E_{x,y}[f(x)f(y)f(xy)] &= E_{x,y}\{[\sum_S \hat{f}(S)\chi_S(x)][\sum_T \hat{f}(T)\chi_T(y)][\sum_U \hat{f}(U)\chi_U(xy)]\} \\ &= E_{x,y}[\sum_{S,T,U} \hat{f}(S)\hat{f}(T)\hat{f}(U)\chi_S(x)\chi_T(y)\chi_U(xy)] \\ &= \sum_{S,T,U} \{\hat{f}(S)\hat{f}(T)\hat{f}(U)E_{x,y}[\chi_S(x)\chi_T(y)\chi_U(xy)]\} \end{aligned}$$

Let us first compute $E_{x,y}[\chi_S(x)\chi_T(y)\chi_U(xy)]$. There are two cases to analyze: $(S = T = U)$ and $(S \neq U$ or $T \neq U)$.
If $S = T = U$, we have

$$\begin{aligned} E_{x,y}[\chi_S(x)\chi_T(y)\chi_U(xy)] &= E_{x,y}[\chi_S(x)\chi_S(y)\chi_S(xy)] \\ &= E_{x,y}[\prod_{i\in S} x_i \prod_{i\in S} y_i \prod_{i\in S}(x_i y_i)] \\ &= E_{x,y}[\prod_{i\in S}(x_i y_i)^2] \\ &= E_{x,y}[\prod_{i\in S} 1] = 1. \end{aligned}$$

If $S \neq U$ or $T \neq U$, then

$$\begin{aligned} E_{x,y}[\chi_S(x)\chi_T(y)\chi_U(xy)] &= E_{x,y}[\prod_{i\in S} x_i \prod_{j\in T} y_j (\prod_{k\in U} x_k \prod_{l\in U} y_l)] \\ &= E_{x,y}[\prod_{i\in S\Delta U} x_i \prod_{j\in T\Delta U} y_j] \\ &= E_x[\prod_{i\in S\Delta U} x_i] * E_y[\prod_{j\in T\Delta U} y_j] \\ &= 0 \end{aligned}$$

because either $E_x[\prod_{i\in S\Delta U} x_i]$ or $E_y[\prod_{j\in T\Delta U} y_j]$ is 0.
Having computed $E_{x,y}[\chi_S(x)\chi_T(y)\chi_U(xy)]$, we come back to $E_{x,y}[f(x)f(y)f(xy)]$:

$$
\begin{aligned}
E_{x,y}[f(x)f(y)f(xy)] &= \sum_{S,T,U} \{\hat{f}(S)\hat{f}(T)\hat{f}(U)E_{x,y}[\chi_S(x)\chi_T(y)\chi_U(xy)]\} \\
&= \sum_S [\hat{f}(S)]^3 \leq max[\hat{f}(S)\sum_S \hat{f}^2(S)] \\
&= max[\hat{f}(s)](due\ to\ Parseval's\ identity) \\
&= 1 - 2min[dist(f, \chi_S)].
\end{aligned}
$$

Therefore, we know that $Pr[group\ law\ failure] \geq min_S[dist(f, \chi_S)]$.

4 Learning functions with Sparse Fourier Representation

Definition 7 *Let* $f : \{\pm 1\}^n \to \{\pm 1\}$ *and* $g : \{\pm 1\}^n \to \Re$.
We say that g ϵ-approximates f *(in* L_2*-norm) if* $E_x[(f(x) - g(x))^2] \leq \epsilon$.

We will use the sign of g to predict the values of f (we are not interested in the magnitude of g; just its sign). If $f(x) \neq sign(g(x))$, we have a *prediction error*.

Claim 8 $Pr[f(x) \neq sign(g(x))] \leq E_x[(f(x) - g(x)^2]$

Proof of Claim 8: We will analyze $I[f(x) \neq sign(g(x))] = 1 - \delta_{f(x),sign(g(x))}$.
Let us denote the indicator variable above by I. There are two cases to analyze depending if $f(x)$ is equal or not to $sign(g(x))$.
If $f(x) = sign(g(x))$ then obviously we have $I = 0$. We also know that $(f(x) - g(x))^2 \geq 0$, therefore $I \leq (f(x) - g(x))^2$.
If $f(x) \neq sign(g(x))$ then $I = 1$; however, in this case, $(f(x) - g(x))^2 \geq 1$. This means that $I \leq (f(x) - g(x))^2$.
We have seen that $I \leq (f(x) - g(x))^2$ regardless of x.

$$
\begin{aligned}
\forall x\ I[f(x) \neq sign(g(x))] \leq (f(x) - g(x))^2 &\Rightarrow E_x[I[f(x) \neq sign(g(x))]] \leq E_x[(f(x) - g(x))^2] \\
&\Leftrightarrow Pr_x[f(x) \neq sign(g(x))] \leq E_x[(f(x) - g(x))^2]
\end{aligned}
$$

■