6.842 Randomness and Computation	March 21, 2012

Lecture 13

Lecturer: Ronitt Rubinfeld — *Scribe: Jieming Mao*

1 Today

- Examples of Fourier representations for basic functions
- Learning via Fourier representations("low degree algorithm")

2 Two Examples of Fourier representation of basic functions

2.1 $\overline{AND}$ on $T \subseteq [1..n]$ such that $|T| = k$

Definition 1 ($\overline{AND}$ function)

$$\overline{AND}(x) = \begin{cases} 1 & \textit{if } \forall \in T, x_i = -1 \\ -1 & \textit{otherwise} \end{cases}$$

Define

$$f(x) = \begin{cases} 1 & \text{if } \forall \in T, x_i = -1 \\ 0 & \text{otherwise} \end{cases} = \frac{1 - x_{i_1}}{2} \cdot \frac{1 - x_{i_2}}{2} \cdots \frac{1 - x_{i_k}}{2} = \sum_{S \subseteq T} \frac{(-1)^{|S|}}{2^k} \chi_S$$

Then we have

$$\overline{AND}(x) = 2f(x) - 1 = (-1 + \frac{2 \cdot 1}{2^k}) + \sum_{S \subseteq T, S \neq \emptyset} \frac{(-1)^{|S|}}{2^{k-1}} \chi_S$$

2.2 Decision Trees

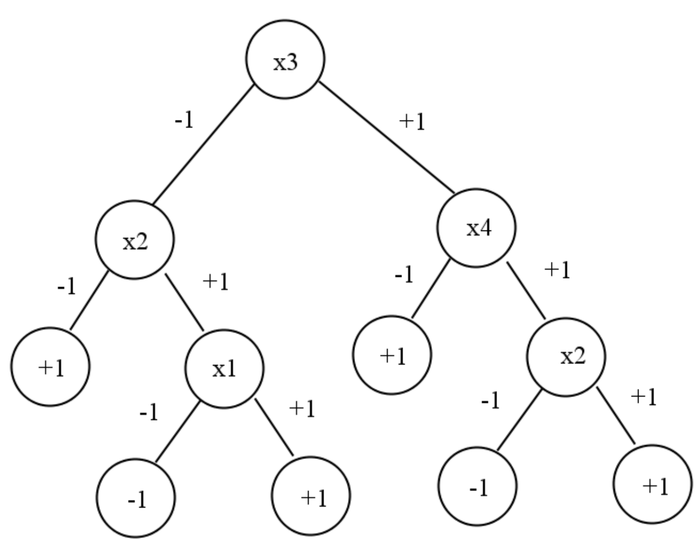

Figure 1: Decision Tree

Definition 2 (path functions)

$$\begin{aligned} f_l(x) &= \prod_{i \in V_l} (1 \pm x_i) \text{ \textit{(Sign depends on whether the path go left or right)}} \\ &= \frac{1}{2^{|V|}} \sum_{S \subseteq V_l} (\pm 1)^{\#\text{ of left turns in the path}} \chi_S \\ &= \begin{cases} 1 & \text{if } x \text{ takes path } l \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

Notice that no input reach more than one leaf, so we can define the decision tree as

$$f(x) = \sum_l f_l(x) \cdot value(l)$$

3 Learning via Fourier Representation

3.1 Fourier Concentration

Definition 3 *$f : \{-1\}^n \to \mathbb{R}$ has $\alpha(\epsilon, n)$-Fourier concentration if*

$$\sum_{S \subseteq [n], |S| > \alpha(\epsilon, n)} \hat{f}(s)^2 \leq \epsilon$$

Remark For boolean function f, by Parseval's theorem, this implies

$$\sum_{S \subseteq [n], |S| \leq \alpha(\epsilon, n)} \hat{f}(s)^2 \geq 1 - \epsilon$$

Observe 1 *If f doesn't depend on x_i, then all $\hat{f}(S)$ for which $i \in S$ satisfy $\hat{f}(S) = 0$.*

Observe 2 *Any function depends on most k variables has*

$$\sum_{S, |S| > k} \hat{f}(S)^2 = 0$$

which implies k-Fourier concentration.

Lemma 1 *$f = \overline{AND}$ on $T \subseteq [1...n]$ has $\log(\frac{4}{\epsilon})$-Fourier concentration.*

Proof Let $k = |T|$

- If $k \leq \log(\frac{4}{\epsilon})$, we've done by the previous observation.
- If $k > \log(\frac{4}{\epsilon})$, we will show f has 0-Fourier concentration. Notice

  $$\hat{f}(\emptyset)^2 = (-1 + \frac{2}{2^k})^2 > 1 - \epsilon$$

  So

  $$\sum_{S, |S| > 0} \hat{f}(S)^2 \leq \epsilon$$

  which implies f has 0-Fourier concentration.

■

3.2 Low Degree Algorithm

- Given degree d , accuracy τ, confidence δ
- Take $m = O(\frac{n^d}{\tau} \ln \frac{n^d}{\delta})$ samples
- For each S such that $|S| \leq d$, $C_S \leftarrow$ estimate of $\hat{f}(S)$
- Output $h(x) = \sum_{|S| \leq d} C_S \chi_S(x)$
- Use $sign(h(x))$ for hypothesis

3.3 Approximating Functions with Low Fourier Degree

Claim 2 *With probability $\geq 1-\delta$, $\forall S$ such that $|S| \leq d$, $|C_S - \hat{f}(S)| \leq \gamma$ for $\gamma \leftarrow \sqrt{\frac{\tau}{n^d}}$.*

Proof Since samples are taken randomly, this claim can be proved by Hoeffling Bound and Union Bound. ∎

Theorem 3 *If f has $d = \alpha(\epsilon, n)$-Fourier concentration, then h satisfies $E_x[(f(x) - h(x))^2] \leq \epsilon + \tau$ with probability $\geq 1 - \delta$.*

Proof Define $g(x) = f(x) - h(x)$. Then we have $\hat{g}(S) = \hat{f}(S) - \hat{h}(S)$. By definition, $\forall S$ such that $|S| > d$, $\hat{h}(S) = 0 \Rightarrow \hat{g}(S) = \hat{f}(S)$. By claim, $\forall S$ such that $|S| \leq d$, $\hat{h}(S) = C_S \Rightarrow |\hat{g}(S)| \leq \gamma$. Thus,

$$E_x[(f(x) - g(x))^2] = E_x[g(x)^2] = \sum_S \hat{g}(S)^2 = \sum_{|S| \leq d} \gamma^2 + \sum_{|S|>d} \hat{f}(S)^2 \leq \tau + \epsilon$$

∎

3.4 $sign(h)$ is useful for prediction

Theorem 4 *Let $f : \{\pm 1\}^n \rightarrow \{\pm 1\}$ and $h : \{\pm 1\}^n \rightarrow \mathbb{R}$, then $Pr[f(x) \neq sign(h(x))] \leq E[(f(x) - h(x))^2]$.*

Proof

$$E[(f(x) - h(x))^2] = \frac{1}{2^n} \sum (f(x) - h(x))^2$$

$$Pr[f(x) \neq sign(h(x))] = \frac{1}{2^n} \sum 1_{sign(h(x)) \neq f(x)}$$

Notice that $(f(x) - h(x))^2 \geq 1_{f(x) \neq sign(h(x))}$. This is because if $f(x) = sign(h(x))$, $1_{f(x) \neq sign(h(x))} = 0 \leq (f(x) - h(x))^2$. If $f(x) \neq sign(h(x))$, $1_{f(x) \neq sign(h(x))} = 1 \leq (f(x) - h(x))^2$. Then we can directly prove this theorem. ∎

Theorem 5 *If C has Fourier concentration $d = \alpha(\epsilon, h)$. There is a $q = O(\frac{n^d}{\epsilon})$ sample uniform distribution learning algorithm for C which outputs hypothesis h' such that $Pr[f(x) \neq h'(x)] \leq 2\epsilon$.*

Proof Run low degree with $\tau = \epsilon$ and outputs h such that $E_x[(f(x) - h(x))^2] \neq 2\epsilon$. Let $h' = sign(h)$, then $Pr[f(x) \neq h'(x)] \leq 2\epsilon$. ∎