6.842 Randomness and Computation	Nov 22, 2017
Lecture 21	
Lecturer: Ronitt Rubinfeld	*Scribe: Yuzhou Gu*

1 Monotone functions

We will talk about weak learning of monotone functions.

Definition 1 *Fix a partial order $\preceq$ on the domain Ω. On the product domain Ω^n, define $\preceq$ so that $x \preceq y$ iff $x_i \le y_i$ for all i. A function $f : \Omega^n \to \mathbb{R}$ is **monotone** if $x \preceq y \Rightarrow f(x) \le f(y)$.*

Now fix $\Omega = \{\pm 1\}$.

Question: How many monotone functions are there?

Lower bound: Consider the set A of points with $\lfloor \frac{n}{2} \rfloor$ -1's. For each assignment of these points to $\{0, 1\}$, there is at least one way to extend this assignment to a monotone function. It is known that $|A| = \Theta(\frac{2^n}{\sqrt{n}})$. Therefore the number of monotone functions is at least $2^{\Theta(\frac{2^n}{\sqrt{n}})}$.

In homework you will prove that one can learn monotone functions over the uniform distribution in $2^{\Theta(\sqrt{n})}$ samples.

Today, you are given random samples and can weakly learn monotone functions on uniform distribution much faster.

Theorem 1 *For all monotone f, there exists $g \in \{\pm 1, x_1, \ldots, x_n\} := S$ such that*

$$\Pr_{X \in U}[f(x) = g(x)] \ge \frac{1}{2} + \Omega(\frac{1}{n}).$$

Corollary 2 *Algorithm for weakly learning monotone functions: for each $g \in S$, estimate agreement with f to within $\Theta(\frac{\epsilon}{n})$.*

Proof [Proof of Theorem 1] The easy case is when f weakly agrees with ± 1. Suppose this does not happen. Then

$$\Pr[f(x) = +1] \in [\frac{1}{2} - \Theta(\frac{1}{n}), \frac{1}{2} + \Theta(\frac{1}{n})] \subseteq [\frac{1}{4}, \frac{3}{4}].$$

Remaining of proof is conducted in Section 2. ■

Definition 2 ***Influence of i-th variable** on $f : \{\pm 1\}^n \to \{\pm 1\}$ is*

$$\mathbf{Inf}_i(f) = \Pr_x[f(x) \ne f(x^{\oplus i})],$$

*where $x^{\oplus i}$ is x with i-th coordinate flipped. **Total influence** is*

$$\mathbf{Inf}(f) = \sum_{1 \le i \le n} \mathbf{Inf}_i(f).$$

You will prove the following results in homework.

Theorem 3 *If f is monotone, then $\mathbf{Inf}_i(f) = \hat{f}(\{i\})$.*

Theorem 4 *Majority function $f(x) = \mathrm{sgn}(\sum x_i)$ maximizes $\mathbf{Inf}(f)$ among all monotone functions.*

How to understand influence of a monotone function: Think about the Hasse diagram of the poset $\{\pm 1\}^n$. Let f be a monotone function $\{\pm 1\}^n \to \{\pm 1\}$. We can think of f as a coloring of the vertices, where red means $f(x) = +1$ and blue means $f(x) = -1$. Monotonicity of f means there are no blue vertices above red vertices. We have

$$\mathbf{Inf}_i(f) = \frac{\text{number of red-blue edges in } i\text{-th direction}}{2^{n-1}}$$

and

$$\mathbf{Inf}(f) = \frac{\text{number of red-blue edges}}{2^n}.$$

2 Canonical path argument

Now let us prove Theorem 1 in the case $\Pr[f(x) = +1] \in [\frac{1}{4}, \frac{3}{4}]$.

Plan:

1. Define a "canonical path" between every pair of red-blue nodes. (Note: must cross ≥ 1 red-blue edge.)
2. Show upper bound on the number of canonical paths passing through any edge (in particular any red-blue edge).
3. Conclude lower bound on the number of red-blue edges.

Part 1 Canonical path from x to y is by flipping bits left to right, where each flip is a step in path.

How many red-blue x-y pairs? At least $\frac{2^n}{4} \cdot \frac{3 \cdot 2^n}{4} = \frac{3}{16} 2^{2n}$.

Part 2 Consider a red-blue edge a-b. Let k be the coordinate where $a_k \neq b_k$. An x-y pair with canonical path going through it must have $y_i = b_i$ for $i \leq k$ and $x_i = a_i$ for $i \geq k$. Therefore there are at most 2^{n-1} such x-y pairs.

Part 3 Each red-blue canonical path uses ≥ 1 red-blue edge. So

$$(\text{\# red-blue edges}) \cdot (\text{max \# of c.p.s per edge}) \geq \text{\# red-blue c.p.s.}$$

So

$$\text{\# red-blue edges} \geq \frac{\frac{3}{16} 2^{2n}}{2^{n-1}} = \frac{3}{8} 2^n.$$

So there exists i such that

$$\text{\# red-blue edges in direction } i \geq \frac{3}{8n} 2^n.$$

So

$$\mathbf{Inf}_i(f) \geq \frac{\frac{3}{8n} 2^n}{2^{n-1}} \geq \frac{3}{4n}.$$

Then

$$\frac{3}{4n} \leq \mathbf{Inf}_i(f) = \hat{f}(\{i\}) = 2 \Pr[f(x) = x_i] - 1.$$

So

$$\Pr[f(x) = x_i] \geq \frac{1}{2} + \frac{3}{8n}.$$