6.895 Randomness and Computation	March 1, 2006
Lecture 7	
Lecturer: Ronitt Rubinfeld	*Scribe: Jelani Nelson*

1 Recap of Previous Lectures

1. Goldreich-Levin Algorithm:

   Given θ and oracle access to $f : \{\pm 1\}^n \to \{\pm 1\}$, output a list L of subsets of $[n]$ such that $\forall S \subseteq [n]$, S will be output if $|\hat{f}(S)| > \theta$, and will not be output if $|\hat{f}(S)| < \theta/2$. Both of these guarantees are with high probability, and the algorithm runs in time poly$(n, \frac{1}{\theta})$.

2. Sampling Theorem:

   $O(k/\delta^2)$ samples allows us to produce an estimate $\tilde{f}(S)$ of $\hat{f}(S)$ such that $Pr_S[|\tilde{f}(S) - \hat{f}(S)| > \delta] \leq e^{-k}$.

2 Decision Trees

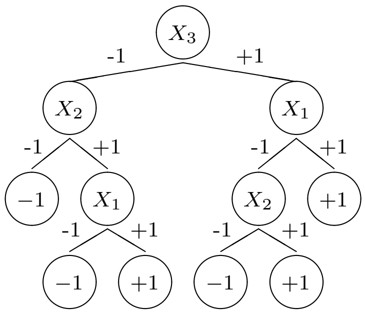

Figure 1: A decision tree for computing the majority function on three variables.

A decision tree is simply a way to represent a function $f : \{\pm 1\}^n \to \{\pm 1\}$. It is a rooted binary tree, where each node is a leaf or has exactly two children. Leaves contain the values $\{\pm 1\}$, signifying that all inputs whose prefixes match the edges taken from the root path are mapped by the function to the value in the leaf. Nodes containing exactly two children contain a variable x_i and have two edges labeled $+1$ and -1, dictacting which way to branch based on the value of x_i in the input.

Definition 1 *The L_1 norm of a function $L_1(f)$ is defined as $\sum_{S \subseteq [n]} |\hat{f}(S)|$.*

Theorem 2 *If f is computable by an m-node decision tree, then $L_1(f) \leq m$.*

Proof For each leaf ℓ define a function $g_\ell : \{\pm 1\}^n \to \{0, 1\}$. We would like to define g_ℓ so that $g_\ell(x)$ is 1 if following x down the tree causes us to arrive at ℓ and 0 otherwise. Let V_ℓ be the set of variables visited on the path from the tree's root to ℓ. We define

$$g_\ell(x) = \prod_{i \in V_\ell} \frac{1 \pm x_i}{2}$$

The sign is a plus if x_1 should be $+1$ to arrive at ℓ and is -1 otherwise. Notice that $g_\ell(x)$ can also be written as

$$\frac{1}{2^{|V_\ell|}} \sum_{S \subseteq V_\ell} (\pm 1) \prod_{i \in S} x_i$$

The sign within the sum is negative iff the number of times -1 taken is odd. Also note that $\prod_{i \in S} x_i$ is just the definition of $\chi_S(x)$, so the above sum is simply $\frac{1}{2^{|V_\ell|}} \cdot 2^{|V_\ell|} = 1$.

Now, $f(X) = \sum_\ell g_\ell(x) \cdot val(\ell)$ (where $val(\ell) \in \{\pm 1\}$). So finally we have

$$\begin{aligned} L_1(f) &= \sum_T |\hat{f}(T)| \\ &= \sum_T \left| \sum_\ell \hat{g}_\ell(T) val(\ell) \right| \\ &\leq \sum_T \sum_\ell |\hat{g}_\ell(T)| \\ &= \sum_\ell \sum_T |\hat{g}_\ell(T)| \\ &= \#leaves \end{aligned}$$

Since the number of leaves is certainly at most the number of total nodes, the claim follows. ■

3 Weak Learning of Monotone Functions

First we will define the partial order $\preceq$ on $\{\pm 1\}^n$. For any $x, y \in \{\pm 1\}^n$, we say that $x \preceq y$ iff $\forall i \; x_i \leq y_i$.

Definition 3 *We say that* $f : \{\pm 1\}^n \to \{\pm 1\}$ *is* monotone *if* $x \preceq y \Rightarrow f(x) \leq f(y)$.

Theorem 4 $\forall f$ *monotone, there is a function* $g \in \{-1, 1, x_1, x_2, \ldots, x_n\}$ *such that* $Pr_x[f(x) = g(x)] \geq \frac{1}{2} + \Omega(\frac{1}{n})$.

Before we prove Theorem 4, we will introduce the notion of the influence of a variable.

In the following, let $u_i = (1, 1, \ldots, 1, -1, 1, \ldots, 1)$ be the vector that has a -1 in the *ith* location and 1's elsewhere.

Definition 5 *The* influence of the variable x_i *on the function* $f : \{\pm 1\}^n \to \{\pm 1\}$ *to be*

$$Inf_i(f) = Pr_x[f(x) \neq f(x \cdot u_i)]$$

where the multiplication $x \cdot u_i$ *is component-wise. We then define the* influence of f *to be*

$$Inf(f) = \sum_{i=1}^{n} Inf_i(f)$$

We will also refer to edges between points in $\{\pm 1\}^n$. There will be an edge between two points iff all coordinates except for one match exactly. We say that the edge crosses the ith direction if the ith coordinate is the one that differs. Also, we refer to the set of points where f is $+1$ as red points, points where it is -1 as blue, and edges as red-blue if they are incident on one red and one blue point.

Then

$$Inf_i(f) = \frac{\text{total \# red-blue edges in } i\text{th direction}}{\text{total \# edges in } i\text{th direction}} = \frac{\text{total \# red-blue edges in } i\text{th direction}}{2^{n-1}}$$

and

$$Inf(f) = \frac{\text{total \# red-blue edges}}{\text{total \# edges}} = \frac{\text{total \# of red-blue edges}}{n2^{n-1}}$$

Lemma 6 *If f is monotone, then $Inf_i(f) = \hat{f}(\{i\})$.*

Proof of Theorem 4 Assume $Pr[f(x) = 1] \in [\frac{1}{4}, \frac{3}{4}]$, else one of $\{\pm 1\}$ agrees with f with probability at least $\frac{3}{4}$. Now we will use a canonical path argument. Suppose x is red and y is blue. A canonical path from x to y is computed by scanning the bits left to right, following an edge where $x_i \neq y_i$. For example, the following is a path from $(+1, +1, -1, -1)$ to $(-1, -1, +1, +1)$:

+1	+1	-1	-1
-1	+1	-1	-1
-1	-1	-1	-1
-1	-1	+1	-1
-1	-1	+1	+1

Now, if $Pr[f(x) = 1] \in [\frac{1}{4}, \frac{3}{4}]$, then the numbers of red and blue nodes are each at least $\frac{1}{4}2^n$. So the number of red-blue pairs is at least $\frac{1}{16}2^{2n}$. Now for any given edge, let us analyze the number of x, y pairs whose canonical paths traverse that edge. We assume the edge crosses the ith direction. The length $(n-i)$ suffix of x must match that of the two vertices the edge is incident upon, and the length $i-1$ prefix of y must match the prefixes of those two vertices. Furthermore, the ith element of y is constrained to be the opposite of that of x. Thus we have 2^i choices for x and 2^{n-i} choices for y, for a total of at most 2^n x, y pairs whose canonical paths cross a given edge. Each canonical path between a red-blue pair must cross at least one red-blue edge. So, there are at least $\frac{1}{16}2^n$ red-blue edges, and thus there are at least $\frac{1}{16n}2^n$ red-blue edges in direction i, for some i. So now we have

$$\hat{f}(\{i\}) = Inf_i(f) \geq \frac{\frac{1}{16n}2^n}{2^{n-1}} = \frac{1}{8n}$$

Since $\hat{f}(S) = 2Pr_x[f(x) = \chi_S(x)]$, the above leads to $Pr_x[f(x) = x_i] \geq \frac{1}{2} + \frac{1}{16n}$. ■

In fact, one can do better using the Kruskal-Katona theorem to show that one of the functions $\{-1, +1, \textit{majority function}\}$ has $\frac{1}{2} + \Omega(\frac{1}{\sqrt{n}})$ agreement with any monotone function. Imagine that the points of $\{\pm 1\}^n$ are placed at vertices of a hypercube, rotated so that $(+1, +1, \ldots, +1)$ is at the very top. We then say that a point is at level i if it contains exactly i $+1$'s. Kruskal-Katona tells us about the rate at which $p_k = Pr[f(x) = 1 | x \text{ at level } k]$ changes as k decreases. We will not go into the proof details here though.