1 Decision Tree

Instead of bounding the probability that $f|_{\rho}$ can be written as some $s$-DNF, we estimate the probability that $f|_{\rho}$ can be computed by a decision tree of height $s$, which is a stronger statement.

**Definition 1.1.** A decision tree is a binary tree, where each node has exactly two children. Each non-leaf node is associated with one variable $x_i$, and the left child corresponds to $x_i = 0$, the right corresponds to $x_i = 1$; and each leaf has a number, either 0 or 1, which represents the output in the path (from root).

The height of the decision tree is the length of the longest path from root to some leaf. Every decision tree computes a Boolean function.

We leave it as an exercise to prove the following.

**Exercise 1.2.** Every decision tree of height $s$ can be written as an $s$-DNF or $s$-CNF. The converse is not true.

Given some CNF (or DNF similarly), there is a natural way to draw the decision tree, that is, start from the first clause $C$, query the first variable in this clause, if $C$ becomes a constant, move to the next clause, otherwise keep querying free variables. It is called canonical decision tree, defined as follows.

**Definition 1.3** (Canonical Decision Tree). Given some CNF $f$, first, fix an order of all clauses and variables, say $f = C_1 \land C_2 \land \ldots \land C_m$. Pick the first variable in $C_1$, say $x_1$, as the root of the canonical decision tree. In the left child of the root, $x_1 = 0$,
• if $C_1|_{x_1=0} = 1$, then move to the next free variable in $C_1$, and repeat the same argument. That is, repeat the same argument to $C_2 \land \ldots \land C_m$ to build the subtree;
• if $C_1|_{x_1=0} = 0$, then $f = 0$. Thus terminate, and give the leaf a label 0;
• if the value of $C_1|_{x_1=0}$ is undetermined, i.e., $x_1$ (not its negation) appears in $C_1$, repeat the same argument to the next literal in $C_1$.

For the right child of $x_1$, it is similar.

Suppose we have already restricted all the free variable up to clause $C_i$, variable $x_j$, where restriction $\rho : x_1, \ldots, x_{j-1} \rightarrow \{0, 1, \ast\}$ corresponds to the path in our decision tree. We are constructing the subtree under $\rho$ as follows (as we did to the root $x_1$).
• if $C_i|_{\rho \cup \{x_j=0\}} = 1$, then move to the next clause, and repeat the same argument. That is, apply the same argument to $C_{j+1} \land \ldots \land C_m$ to build the subtree;
• if $C_i|_{\rho \cup \{x_j=0\}} = 0$, then $f = 0$. Thus terminate, and give the leaf a label 0.
• if the value of $C_i|_{\rho \cup \{x_j=0\}}$ is undetermined, which implies $x_j$ (not its negation) appears in $C_i$, repeat the same argument to the next literal in $C_i$.

Finally, we will exhaust all clauses and all free variables. It is easy to see the decision tree we have constructed computes the function $f$.

Given Boolean function $f$, let $DT_c(f)$ denotes the height of the canonical decision tree, where the subscript $c$ emphasizes that it is the canonical decision tree, which in general, may not be the shortest decision tree computing $f$.

In order to prove Håstad’s Switching Lemma, we prove a stronger statement.

**Lemma 1.4** (Håstad’s Switching Lemma, decision tree version). Let $f$ be any $t$-CNF. Then
\[
\Pr_\rho[DT_c(f|_\rho) > s] < (5pt)^s,
\]
where $\rho \in \mathcal{R}(p, 1/2)$, and $DT_c(f|_\rho)$ denotes the height of the canonical decision tree computing $f|_\rho$.  

2
The proof presented in the next section is slightly different from Håstad’s original proof, which goes through the following stronger statement, which says, for any $t$-CNF,
\[
\Pr_{\rho}[DT_c(f|_\rho) > s | g|_\rho \equiv 0] < (5pt)^s,
\]
where $g$ is any Boolean function. It is a generally useful trick to make the proof easier by proving something stronger. However, we find this trick unnecessary here, and we will be able to prove it directly by induction on the number of clauses.

2 Håstad’s Proof

Proof. We are proving Lemma 1.4 by induction on the number of clauses. Write $t$-CNF $f$ as
\[
f = C_1 \land C_2 \land \ldots \land C_m.
\]

Instead of proving by induction, we try to unveil the discovery of the proof, that is to write down and estimate a recursive formulae. For this purpose, let
\[
T(s, t, m) := \sup_{f \in \mathcal{R}(p, 1/2)} \Pr_{\rho \in \mathcal{R}(p, 1/2)}[DT_c(f|_\rho) > s],
\]
where the supremum is taking over all $t$-CNF with at most $m$ clauses in any number of variables. The goal is to prove
\[
T(s, t, m) < (5pt)^s.
\] (1)

If $m = 1$, (1) is trivially true, because $C_1$ is automatically $s$-DNF for any $s \geq 1$.

If $m > 1$, assuming Lemma 1.4 is true for any $t$-CNF $f$ with number of clauses $< m$, we shall prove it for any $t$-CNF $f$ with $m$ clauses.

Without loss of generality, assume $C_1 = x_1 \lor x_2 \lor \ldots \lor x_t$. (If some literal is a negation in $C_1$, then negate all instances of it.)

Think of $\rho \in \mathcal{R}(p, 1/2)$ as a composition of two restrictions $\rho = \rho_1 \circ \rho_2$, where,
\[
\rho_1 : \{x_1, \ldots, x_t\} \to \{0, 1, *\},
\]
\[
\rho_2 : \{x_{t+1}, \ldots, x_n\} \to \{0, 1, *\},
\]
and $\rho_1, \rho_2 \in R(p, 1/2)$ as well. For convenience, let $g = C_2 \land \ldots \land C_m$, and thus $f = C_1 \land g$, and

$$f|_{\rho} = C_1|_{\rho} \land g|_{\rho} = C_1|_{\rho_1} \land g|_{\rho_1 \circ \rho_2}.$$  

Now let us estimate the height of the canonical decision tree of $C_1|_{\rho_1} \land g|_{\rho_1 \circ \rho_2}$ case by case. Recall the definition of canonical decision tree: we are constructing clause by clause, literal by literal. In this case, we start with $C_1|_{\rho_1}$. The situation is divided into 2 cases.

**Case 1:** there exists some $i \in [t]$, $\rho_1(x_i) = 1$, and thus $C_1|_{\rho_1} = 1$. Recall the definition of canonical decision tree, in which we simply move to the next clause, since $C_1|_{\rho_1} \equiv 1$. Let us denote this event by $A$. It is easy to verify

$$\Pr_{\rho}[A] = 1 - \left(1 + \frac{p}{2}\right)^t$$

and

$$\Pr_{\rho}[DT_c(f|_{\rho}) > s \mid A] \leq T(s, t, m - 1),$$

since $f|_{\rho} = g|_{\rho}$, which is a $t$-CNF with $m - 1$ clauses.

**Case 2:** $\rho_1(x_i) \in \{0, *\}$ for all $i \in [t]$. Let $L := \rho_1^{-1}(*).$ Denote this event by $B_L$, where $L \subseteq [t]$. The union of $B_L$ for all $L \subseteq [t]$ is the complement of $A$. It is easy to see

$$\Pr_{\rho}[B_L] = p^{\ell} \left(1 - \frac{p}{2}\right)^{t-\ell},$$

where $\ell := |L|$.

Recall the construction of canonical decision tree, where we are querying free variables in $L$ one by one until $C_1$ becomes a constant, then either move to $g = C_2 \land \ldots \land C_m$ or terminate (depending on whether $C_1$ becomes 1 or 0). When querying free variables in $L$, there are $2^{\ell}$ results in total, which one-to-one corresponds to restrictions (mappings)

$$\sigma : \{x_i : i \in L\} \rightarrow \{0, 1\}.$$  

If $\sigma \equiv 0$, we simply terminate because $C_1$ becomes 0 and thus $f$ becomes 0. Otherwise, we construct the canonical decision tree for $g|_{\rho_1 \circ \rho_2}$, and append it to $\sigma$, where $\sigma$ represents a path in the decision tree.
The event $DT_c(f|\rho) > s$ implies that there exists some $\sigma \neq 0$ such that $DT(g|\rho_1 \circ \sigma \rho_2) > s - \ell$, for which we can apply a union bound, i.e,

$$\Pr[DT_c(f|\rho) > s | B_L] \leq (2^\ell - 1)T(s - \ell, t, m - 1).$$

Putting two cases together,

$$\Pr_{\rho}[DT_c(f|\rho) > s] = \Pr_{\rho}[A] \Pr_{\rho}[DT_c(f|\rho) > s | A] + \sum_{L \subseteq [t]} \Pr_{\rho}[B_L] \Pr_{\rho}[DT_c(f|\rho) > s - \ell | B_L]$$

$$= (1 - \left(\frac{1 + p}{2}\right)^t)T(s, t, m - 1) + \sum_{\ell = [t]} \sum_{L \subseteq [t]} p^\ell \left(\frac{1 - p}{2}\right)^{t-\ell}(2^\ell - 1)T(s - \ell, t, m - 1).$$

The remaining is calculation. Instead of plugging in $(5pt)^*$, we put in $(c_Hpt)^*$, where the constant $c_H$ is to be optimized later. It suffices to prove, there exists some $c_H$ such that

$$(c_Hpt)^* \geq (1 - \left(\frac{1 + p}{2}\right)^t)(c_Hpt)^* + \sum_{L} p^\ell \left(\frac{1 - p}{2}\right)^{t-\ell}(2^\ell - 1)(c_Hpt)^{*-\ell}$$

holds for any integer $s, t \geq 1$. Dividing both sides by $(c_Hpt)^*$,

$$1 \geq (1 - \left(\frac{1 + p}{2}\right)^t) + \sum_{L} p^\ell \left(\frac{1 - p}{2}\right)^{t-\ell}(2^\ell - 1)(c_Hpt)^{-\ell}$$

$$\Leftrightarrow (1 - \frac{1 + p}{2})^t \geq \sum_{L} p^\ell \left(\frac{1 - p}{2}\right)^{t-\ell}(2^\ell - 1)(c_Hpt)^{-\ell}.$$

Dividing both sides by $((1 - p)/2)^t$,

$$\left(\frac{1 + p}{1 - p}\right)^t \geq \sum_{L} p^\ell \left(\frac{1 - p}{2}\right)^{-\ell}(2^\ell - 1)(c_Hpt)^{-\ell}$$

$$= \sum_{L} \left(\frac{4}{c_Ht(1-p)}\right)^\ell - \left(\frac{2}{c_Ht(1-p)}\right)^\ell$$

$$= \left(\frac{4}{c_Ht(1-p)} + 1\right)^t - \left(\frac{2}{c_Ht(1-p)} + 1\right)^t$$

$$\geq \left(\frac{4}{c_Ht} + 1\right)^t - \left(\frac{2}{c_Ht} + 1\right)^t$$
where the second last step is by binomial theorem. It is not difficult to check that if $c_H = 5$, $(\frac{4}{c_H t} + 1)^t - (\frac{2}{c_H t} + 1)^t < 1$ for any $t > 0$, which finishes our proof.

In many applications of switching lemma, $t = s$ and $p$ is taking to be $\Theta(1/s)$, in which case we can get better constant $c_H$. The following qualitatively “stronger” version of Håstad’s Switching Lemma can be easily extracted from the last step of our proof.

**Lemma 2.1** (Håstad’s Switching Lemma, improved constant). Let $f$ be any $t$-CNF. Then

$$\Pr_{\rho}[DT_c(f|\rho) > s] < (c_H pt)^s,$$

where $\rho \in \mathcal{R}(p, 1/2)$, and $DT_c(f|\rho)$ denotes the height of the canonical decision tree computing $f|\rho$, and $c_H$ is any constant such that

$$(\frac{1+p}{1-p})^t \geq (\frac{4}{c_H t} + 1)^t - (\frac{2}{c_H t} + 1)^t.$$

**Exercise 2.2.** Use the above lemma to improve the multiplicative constant in the $2^{\Omega(n^{1/(d-1)})}$ lower bound for \textsc{Parity}_n, where $d$ is the depth as usual.

**Remark 2.3.** The improved constant version appears in Håstad’s dissertation, which is due to Ravi Boppana.

**References**

