Total Influence (lecture notes) [Edited 2025-10-20]

Course: Analysis of Boolean Functions, Autumn 2025, University of Chicago

Instructor: William Hoza (`williamhoza@uchicago.edu`)

In these notes, we will show that a few interesting classes of functions are concentrated at relatively low degree, hence nontrivially learnable from random examples. The proofs are based on the notion of *total influence*. If $x \in \{\pm 1\}^n$ and $i \in [n]$, let $x^{\oplus i}$ denote x with the i-th bit flipped.

Definition 0.1 (Influence and total influence)**.** If $f : \{\pm 1\}^n \to \{\pm 1\}$, then the *influence* of variable i on f is defined by $\mathrm{Inf}_i[f] = \Pr_x[f(x) \neq f(x^{\oplus i})]$. Furthermore, the *total influence* of f is defined by $\mathrm{I}[f] = \sum_i \mathrm{Inf}_i[f]$.

Total influence is a measure of the "complexity" of f. Besides its application to learning theory, total influence is also interesting for its own sake.

1 Total influence of size-s decision trees

As a warm-up, let's analyze the total influence of decision trees, even though this won't immediately buy us anything in terms of learnability. We use a connection between total influence and *sensitivity*.

Definition 1.1 (Sensitivity)**.** For a function $f : \{\pm 1\}^n \to \{\pm 1\}$ and an input $x \in \{\pm 1\}^n$, define

$$\mathrm{sens}_f(x) = |\{i : f(x) \neq f(x^{\oplus i})\}|.$$

Lemma 1.2. *For any $f : \{\pm 1\}^n \to \{\pm 1\}$, the total influence of f is equal to the average sensitivity of f. That is,* $\mathrm{I}[f] = \mathbb{E}_x[\mathrm{sens}_f(x)]$.

Proof. Linearity of expectation. □

For a decision tree f, let $\mathrm{cost}_f(x)$ denote the number of queries that f makes on x.

Lemma 1.3. *If f is a decision tree, then* $\mathrm{sens}_f(x) \leq \mathrm{cost}_f(x)$.

Proof. If f did not query x_i, then $f(x^{\oplus i}) = f(x)$. □

Lemma 1.4. *If f is a size-s decision tree, then $s = \mathbb{E}_x[2^{\mathrm{cost}_f(x)}]$, and moreover $\mathbb{E}[\mathrm{cost}_f(x)] \leq \log s$.*

Proof. Let L be the set of leaves. For each leaf $u \in L$, let d_u be the depth of u. Then

$$\mathop{\mathbb{E}}_x[2^{\mathrm{cost}_f(x)}] = \sum_{u \in L} \Pr[\text{reach u}] \cdot 2^{d_u} = \sum_{u \in L} 2^{-d_u} \cdot 2^{d_u} = |L| = s.$$

The "moreover" part of the lemma follows from Jensen's inequality. □

Corollary 1.5. *If $f : \{\pm 1\}^n \to \{\pm 1\}$ is a size-s decision tree, then* $\mathrm{I}[f] \leq \log s$.

2 Total influence of width-w DNFs

A DNF formula is a disjunction of *terms*, each of which is a conjunction of literals (variables and their negations). The *width* of a DNF formula is the maximum number of literals in a single term. To bound the total influence of width-w DNFs, we use a modified version of Lemma 1.2.

Lemma 2.1. *For any $f : \{0,1\}^n \to \{0,1\}$, we have* $\mathrm{I}[(-1)^f] = 2\,\mathbb{E}_x[f(x) \cdot \mathrm{sens}_f(x)]$.

Proof. We have

$$\mathrm{Inf}_i[(-1)^f] = \Pr_x[f(x) \neq f(x^{\oplus i})] = 2\Pr_x[f(x) = 1, f(x^{\oplus i}) = 0] = 2\mathop{\mathbb{E}}_x[f(x) \cdot 1[f(x) \neq f(x^{\oplus i})]].$$

Linearity of expectation completes the proof. $\square$

Corollary 2.2. *If $f\colon \{0,1\}^n \to \{0,1\}$ is a width-w DNF, then $\mathrm{I}[(-1)^f] \leq 2w$.*

Proof. For any $x \in \{0,1\}^n$, we have $f(x) \cdot \mathrm{sens}_f(x) \leq w$, because, if $f(x) = 1$, then some term of f is satisfied, hence only variables in that term can be pivotal for f on x. $\square$

It is apparently an open question whether the factor of two in Corollary 2.2 can be eliminated.

3 Fourier concentration from total influence bounds

In this course, we will develop several methods for using total influence bounds to prove Fourier concentration and learnability bounds. The simplest version is a bound that says every Boolean function f is ε-concentrated on degree up to $\mathrm{I}[f]/\varepsilon$. The proof is based on *discrete derivatives.*

Definition 3.1 (Discrete derivatives)**.** If $f\colon \{\pm 1\}^n \to \mathbb{R}$, then

$$(D_i f)(x) = \frac{f(x^{(i\mapsto +1)}) - f(x^{(i\mapsto -1)})}{2}.$$

Let us compute the Fourier coefficients of $D_i f$. We have $D_i \chi_S = \chi_{S\setminus\{i\}}$ if $i \in S$, and $D_i \chi_S = 0$ if $i \notin S$. (Just like partial derivatives from calculus class!) Therefore, by linearity,

$$D_i f = \sum_{S\subseteq[n], i\in S} \widehat{f}(S) \cdot \chi_{S\setminus\{i\}}.$$

Lemma 3.2 (Fourier formula for total influence)**.** *For any $f\colon \{\pm 1\}^n \to \{\pm 1\}$, we have $\mathrm{I}[f] = \mathbb{E}_{S\sim\mathcal{S}_f}[|S|]$.*

Proof.

$$\mathrm{I}[f] = \sum_{i=1}^n \mathrm{Inf}_i[f] = \sum_{i=1}^n \mathop{\mathbb{E}}_x[(D_i f)(x)^2] = \sum_{i=1}^n \sum_{S\subseteq[n]} \widehat{D_i f}(S)^2 = \sum_{i=1}^n \sum_{S\subseteq[n], i\in S} \widehat{f}(S)^2 = \sum_{S\subseteq[n]} |S| \cdot \widehat{f}(S)^2. \quad \square$$

Corollary 3.3. *Every $f\colon \{\pm 1\}^n \to \{\pm 1\}$ is ε-concentrated up to degree $\mathrm{I}[f]/\varepsilon$.*

Proof. This is Markov's inequality applied to the random variable $|S|$ where $S \sim \mathcal{S}_f$. $\square$

For example, width-w DNFs are ε-concentrated up to degree $O(w/\varepsilon)$, hence learnable from random examples in time $n^{O(w/\varepsilon)}$. We will improve these bounds in later classes.

4 Total influence of unate functions

Definition 4.1. A Boolean function $f\colon \{\pm 1\}^n \to \{\pm 1\}$ is *monotone* if, for every $x, y \in \{\pm 1\}^n$, we have $x \leq y \implies f(x) \leq f(y)$. More generally, we say that f is *unate* if it can be written in the form $f(x) = g(x \circ a)$, where g is a monotone function, $a \in \{\pm 1\}^n$, and $x \circ a$ denotes coordinatewise multiplication.

If $x \in \{\pm 1\}^n$, $i \in [n]$, and $b \in \{\pm 1\}$, let $x^{(i\mapsto b)}$ denote x with b in place of the i-th coordinate. We use $\widehat{f}(i)$ as a shorthand for $\widehat{f}(\{i\})$.

Lemma 4.2. *Let $f\colon \{\pm 1\}^n \to \{\pm 1\}$. If f is monotone, then $\mathrm{Inf}_i[f] = \widehat{f}(i)$. If f is unate, then $\mathrm{Inf}_i[f] = |\widehat{f}(i)|$.*

Proof. We have $\widehat{f}(i) = \mathbb{E}_x[f(x) \cdot x_i] = \frac{1}{2} \cdot \mathbb{E}_x\left[f(x^{(i\mapsto +1)}) - f(x^{(i\mapsto -1)})\right] = \mathbb{E}_x[(D_i f)(x)]$. If f is monotone, the latter quantity is equal to $\mathrm{Inf}_i[f]$. If f is unate, it is $\pm\,\mathrm{Inf}_i[f]$. □

Lemma 4.3. *For any* $f\colon \{\pm 1\}^n \to \{\pm 1\}$, *we have* $\sum_{i=1}^n |\widehat{f}(i)| \le \sqrt{n}$.

Proof. By Cauchy-Schwarz, we have $\sum_{i=1}^n |\widehat{f}(i)| \le \sqrt{n \cdot \sum_{i=1}^n \widehat{f}(i)^2}$. By Parseval, $\sum_{i=1}^n \widehat{f}(i)^2 = 1$. □

Corollary 4.4. *If* $f\colon \{\pm 1\}^n \to \{\pm 1\}$ *and* f *is unate, then* $\mathrm{I}[f] \le \sqrt{n}$.

Thus, unate functions are ε-concentrated up to degree $O(\sqrt{n}/\varepsilon)$, hence learnable from random examples in time $n^{O(\sqrt{n}/\varepsilon)}$, which is slow but highly nontrivial.

5 Total influence of size-s unate decision trees

Theorem 5.1. *Let* $f\colon \{\pm 1\}^n \to \{\pm 1\}$ *be unate and computable by a size-s decision tree. Then* $\mathrm{I}[f] \le \sqrt{\log s}$.

Note that every function can be computed by a decision tree of size $s = 2^n$, hence Theorem 5.1 strengthens the result from the previous section that unate functions have $\mathrm{I}[f] \le \sqrt{n}$.

Proof. Assume first that f is monotone. Then $\mathrm{Inf}_i = |\widehat{f}(i)|$. Sample $x \in \{\pm 1\}^n$ uniformly at random. Define $y \in \{\pm 1\}^n$ by

$$y_i = \begin{cases} x_i & \text{if } f \text{ queries } x_i \text{ on input } x \\ 0 & \text{otherwise.} \end{cases}$$

The outcome $f(x)$ is determined by y. Abusing notation, we can write $f(x) = f(y)$. Then we have

$$\widehat{f}(i) = \mathop{\mathbb{E}}_x[f(x) \cdot x_i] = \mathop{\mathbb{E}}_x[f(y) \cdot x_i] = \mathop{\mathbb{E}}_y\left[f(y) \cdot \mathop{\mathbb{E}}_{x|y}[x_i]\right] = \mathbb{E}[f(y) \cdot y_i].$$

Therefore,

$$\mathrm{I}[f] = \sum_{i=1}^n \widehat{f}(i) = \mathbb{E}\left[f(y) \cdot \sum_{i=1}^n y_i\right] \le \mathbb{E}\left[\left|\sum_{i=1}^n y_i\right|\right] \le \sqrt{\mathbb{E}\left[\left(\sum_{i=1}^n y_i\right)^2\right]} = \sqrt{\mathbb{E}\left[\sum_{i=1}^n y_i^2\right] + \sum_{i\neq j} \mathbb{E}[y_i y_j]}.$$

We analyze the second term first. If $i \neq j$, then

$$\mathbb{E}[y_i y_j] = \mathop{\mathbb{E}}_y\left[\mathop{\mathbb{E}}_{x|y}[x_i] \cdot \mathop{\mathbb{E}}_{x|y}[x_j]\right] = \mathop{\mathbb{E}}_y\left[\mathop{\mathbb{E}}_{x|y}[x_i x_j]\right] = \mathbb{E}[x_i x_j] = \mathbb{E}[x_i] \cdot \mathbb{E}[x_j] = 0 \cdot 0 = 0.$$

Therefore,

$$\mathrm{I}[f] \le \sqrt{\mathbb{E}\left[\sum_{i=1}^n y_i^2\right]} = \sqrt{\mathbb{E}[\mathrm{cost}_f(x)]} \le \sqrt{\log s}$$

by Lemma 1.4. Finally, suppose more generally that f is unate, say $f(x) = g(x \circ a)$ for some monotone g. Then $\mathrm{Inf}_i[f] = \mathrm{Inf}_i[g]$, and g can be computed by a size-s decision tree, so $\mathrm{I}[f] = \mathrm{I}[g] \le \sqrt{\log s}$. □

Consequently, size-s monotone decision trees are ε-concentrated up to degree $\sqrt{\log s}/\varepsilon$, hence learnable from random examples in time $n^{O(\sqrt{\log s})}$, which is faster than the previous $n^{O(\log s)}$ algorithm we saw for general decision trees. By more sophisticated techniques, one can show that size-s monotone decision trees are learnable from random examples in $\mathrm{poly}(n, s)$ time.