Lecture 26: Influence and KKL Theorem

Influence

- Let $f(x_1, \dots, x_n)$ be a boolean function
- The *influence of* x_i represents the probability (over random choice of all other variables) that f is sensitive to the value of x_i
- Alternately,

$$\mathrm{Inf}_i(f) := \mathop{\mathbb{E}}_{x_{[n]\setminus\{i\}} \xleftarrow{\$} \{0,1\}^{n-1}} [f(x) \neq f(x + e_i)],$$

where e_i is the element with 1 exactly at the i-th position and 0 everywhere else

Influence

- Let $J \subseteq [n]$ be a subset of indices
- Influence of J is represented by:

$$\mathrm{Inf}_J(f) := \mathop{\mathbb{E}}_{x_{\overline{J}} \xleftarrow{\$} \{0,1\}^{\overline{J}}} [f(x) \text{ is not constant}]$$

Examples

- AND_n is the function that outputs the AND of its n inputs. The $\mathrm{Inf}_i(\mathrm{AND}_n)$, for any i, is the probability that AND_n is sensitive to x_i. That happens exactly when all others input bits are 1, i.e. with probability $2^{-(n-1)}$.
- OR_n also has identical influence.
- MAJ_n, for odd n, outputs the majority bit of n input bits. The $\mathrm{Inf}_i(\mathrm{MAJ}_n)$ is exactly the probability that the number of 0s and 1s in the remaining inputs bits is equal. This happens with probability $2^{-(n-1)}\binom{n-1}{(n-1)/2} \approx n^{-1/2}$.
- Note that sensitivity of AND_n is very low; but it cannot be *high* because the output of AND_n is constant with ≈ 1 probability. But, when the output of the function is nearly balanced does some variable have high influence?

Example

- Note that MAJ_n has balanced output and its variables have $\approx n^{-1/2}$ influence.
- In fact, any J with size $\approx n^{1/2}$ has constant influence.
- Think: Are there balanced functions with lesser influence?
- $\mathrm{TRIBES}_{s,w}$ is the OR_s of AND_w of $n = sw$ input bits. That is, interpret the input as s blocks of w bits each. Apply AND_w on each block and output the OR_s of the ANDs.
- Think: For what values of s and w is $\mathrm{TRIBES}_{s,w}$ balanced?
- Think: For these values of s and w, what is the influence of any variable?

Target Result

Theorem (Kahn-Kalai-Linial)

For every balanced function f, there exists a variable with influence at least $\approx \log n/n$.

- We will show a result by Talagrand (presented next slide)
- Prove the KKL result using that result
- Think: This is asymptotically tight!

Talagrand's Result

Lemma

Let g be a function with $\|g\|_2 \neq \|g\|_{3/2}$, then:

$$\sum_{S \neq \emptyset} \frac{\widehat{g}(S)^2}{|S|} \leqslant \frac{2.5 \, \|g\|_2^2}{\log \|g\|_2 \, / \, \|g\|_{3/2}}$$

- We apply Hypercontractivity with $p = 3/2$, $q = 2$ and $\rho^2 = 1/2$

$$\|g\|_{3/2}^2 \geqslant \|T_\rho(g)\|_2^2 = \sum_S \widehat{g}(S)^2 / 2^{|S|} \geqslant \sum_{S\colon |S|=k} g(S)^2 / 2^k$$

That is, for any $k > 0$, we have:

$$\sum_{S\colon |S|=k} \frac{g(S)^2}{|S|} \leqslant \frac{2^k}{k} \|g\|_{3/2}^2$$

Proof

- Therefore, for any m, we have:

$$\begin{aligned}\sum_{S\neq\emptyset}\frac{\widehat{g}(S)^2}{|S|} &= \sum_{1\leqslant k\leqslant m}\sum_{S\colon |S|=k}\frac{\widehat{g}(S)^2}{|S|} + \sum_{S\colon |S|>m}\frac{\widehat{g}(S)^2}{|S|}\\ &\leqslant \left(\sum_{1\leqslant k\leqslant m}\frac{2^k}{k}\right)\|g\|_{3/2}^2 + \sum_{S\colon |S|>m}\frac{\widehat{g}(S)^2}{(m+1)}\\ &\leqslant \left(\sum_{1\leqslant k\leqslant m}\frac{2^k}{k}\right)\|g\|_{3/2}^2 + \frac{1}{(m+1)}\|g\|_2^2\end{aligned}$$

- Choose largest m such that $2^m\|g\|_{3/2}^2 \leqslant \|g\|_2^2$. Using the maximality property, we have: $(m+1) > 2\log\|g\|_2 / \|g\|_{3/2}$

Proof

- By induction we can prove the following upper bound:

$$\sum_{1 \leqslant k \leqslant m} \frac{2^k}{k} \leqslant \frac{2 \cdot 2^{m+1}}{(m+1)}$$

- So, overall we have:

$$\sum_{S \neq \emptyset} \frac{\widehat{g}(S)^2}{|S|} \leqslant \frac{4 \cdot 2^m}{(m+1)} \|g\|_{3/2}^2 + \frac{1}{(m+1)} \|g\|_2^2$$

$$\leqslant \frac{(4+1)}{(m+1)} \|g\|_2^2 \leqslant \frac{5}{2 \log \|g\|_2 \,/\, \|g\|_{3/2}} \|g\|_2^2$$

- This gives the overall bound of the lemma

Using Talagrand's Result to get KKL Theorem

- For $i \in [n]$, define $g_i(x) := f(x) - f(x + e_i)$ (the '$-$' sign in the definition is subtraction over $\mathbb{R}$ and '$+$' sign in the definition is addition over $\{0,1\}^n$)
- Note that $\widehat{g_i}(S) = 2\widehat{f}(S)$, if $i \in S$; otherwise, $\widehat{g_i}(S) = 0$
- Note that: $\mathbb{E}_{x \xleftarrow{\$} \{0,1\}^n}[|g_i(x)|] = \mathrm{Inf}_i(f)$
- Since f is a boolean function, g_i has output in $\{-1, -1\}$
- Therefore, for $p \geqslant 1$, we have: $\|g_i\|_p^p = \|g_i\|_1 = \mathrm{Inf}_i(f)$
- So, $\|g_i\|_2 = \mathrm{Inf}_i(f)^{1/2}$ and $\|g_i\|_{3/2} = \mathrm{Inf}_i(f)^{2/3}$
- Using this, we can deduce:

$$\|g_i\|_2^2 = \mathrm{Inf}_i(f)$$

$$\log \|g\|_2 \,/\, \|g\|_{3/2} = (1/6) \log 1/\mathrm{Inf}_i(f)$$

Proof Continued

- Use Talagrand's Result on g_i:

$$\sum_{S \neq \emptyset} \frac{\widehat{g_i}(S)^2}{|S|} \leqslant \frac{15\mathrm{Inf}_i(f)}{\log 1/\mathrm{Inf}_i(f)}$$

- Now, let us understand the relation between the left-hand-side using f's Fourier coefficient:

$$\sum_{S \neq \emptyset} \frac{\widehat{g_i}(S)^2}{|S|} = \sum_{S:\, i \in S} \frac{4\widehat{f}(S)^2}{|S|}$$

- Previous two inequalities gives:

$$\sum_{S:\, i \in S} \frac{\widehat{f}(S)^2}{|S|} \leqslant (15/4)\frac{\mathrm{Inf}_i(f)}{\log 1/\mathrm{Inf}_i(f)}$$

Proof Continued

- Summing the previous inequality over all $i \in [n]$, we get:

$$(15/4) \sum_{i\in[n]} \frac{\mathrm{Inf}_i(f)}{\log 1/\mathrm{Inf}_i(f)} \geqslant \sum_{i\in[n]} \sum_{S\colon i\in S} \frac{\widehat{f}(S)^2}{|S|} \overset{(*)}{=} \sum_{S\neq\emptyset} \widehat{f}(S)^2 \overset{(\dagger)}{=} \mathrm{Var}[f]$$

  The $(*)$ equality is because the term $\widehat{f}(S)^2/|S|$ is considered once for every $i \in S$, i.e. $|S|$ times. The $(\dagger)$ equality is because $\mathrm{Var}[f] = \mathbb{E}[f^2] - \mathbb{E}[f]^2 = \sum_{S\neq\emptyset} \widehat{f}(S)^2$.
- A "nearly balanced f" has $\mathrm{Var}[f] = \Omega(1)$
- So, we have:

$$\sum_{i\in[n]} \frac{\mathrm{Inf}_i(f)}{\log 1/\mathrm{Inf}_i(f)} \geqslant \Omega(1)$$

- So, there exists $i = i^* \in [n]$ such that:

$$\frac{\mathrm{Inf}_i(f)}{\log 1/\mathrm{Inf}_i(f)} \geqslant \Omega(1/n)$$

- That is $\mathrm{Inf}_{i^*}(f) \geqslant \Omega(\log n/n)$ (the KKL Result)