CS 395T Computational Complexity of Machine Learning

Lecture 12: March 1, 2005 | Lecturer: Adam Klivans

Scribe: Dongmin

12.1 Fourier Learning

12.1.1 Basis

Consider a boolean function $f : \{0,1\}^n \to \{-1,1\}$. Let $s \subset [n]$ be a set of indices, $\chi_s : \{0,1\}^n \to \{-1,1\}$ be a parity function. Then, $f(x)$ can be represented as:

$$f(x) = \sum_{s \subset [n]} \hat{f}(s)\chi_s(x).$$

Here, $\hat{f}(s)$'s are Fourier coefficients.

We can define an inner product and norm for this function space:

$$\langle f, g \rangle = \frac{1}{2^n} \sum_{x \in \{0,1\}^n} f(x) \cdot g(x),$$

$$\|f\| = \sqrt{\langle f, f \rangle}.$$

Moreover, these 2^n χ_s functions form a orthonormal basis.

$$\forall s, \quad \|\chi_s\| = \sqrt{\langle \chi_s, \chi_s \rangle} = \sqrt{\frac{1}{2^n} \sum_{x \in \{0,1\}^n} \chi_s(x) \cdot \chi_s(x)} = 1$$

Since the basis is orthonormal, the Fourier coefficients are:

$$\hat{f}(s) = \langle f, \chi_s \rangle = E_{x \in \{0,1\}^n} \left[f(x) \cdot \chi_s(x) \right].$$

Each $\hat{f}(s)$ measures the correlation between f and a specific parity fucntion χ_s.

12.1.2 Cross Correlation

For any boolean function f and g,

$$\begin{aligned}
E_x\left[f(x)\cdot g(x)\right] &= E_x\left[\left(\sum_{s_1\subset[n]}\hat{f}(s_1)\chi_{s_1}\right)\left(\sum_{s_2\subset[n]}\hat{g}(s_2)\chi_{s_2}\right)\right] \\
&= E_x\left[\sum_{s_1,s_2\subset[n]}\hat{f}(s_1)\hat{g}(s_2)\chi_{s_1}\chi_{s_2}\right] \\
&= \sum_{s\subset[n]}\hat{f}(s)\hat{g}(s)\cdot E_x\left[\chi_s^2\right] \\
&= \sum_{s\subset[n]}\hat{f}(s)\hat{g}(s). \qquad (12.1)
\end{aligned}$$

Note that the basis is orthonormal, therefore $E\left[\chi_{s_1}\cdot\chi_{s_2}\right]=0$, for $s_1\neq s_2$, and $E\left[\chi_s^2\right]=1$.

Using this result; a special case of Parseval's theorem,

$$\|f\|^2 = E_x\left[f^2(x)\right] = \sum_{s\subset[n]}\hat{f}^2(s) = 1.$$

12.1.3 AND Function

Consider a boolean function $AND:\{0,1\}^n\rightarrow\{0,1\}$.

Firstly, assume that we have AND function over some $T\subset[n]$ of the input variables. For example, $T=\{x_1,\cdots,x_{10}\}$. Let's think of a Fourier coefficient for $\chi_{\{x_1,\cdots,x_{11}\}}$.

$$\begin{aligned}
\hat{f}(\{x_1,\cdots,x_{11}\}) &= \frac{1}{2^n}\sum_{x\in\{0,1\}^n}AND(x_1,\cdots,x_{10})\cdot\chi_{\{x_1,\cdots,x_{11}\}} \\
&= \frac{1}{2^n}\sum_{x\in\{0,1\}^n}AND(x_1,\cdots,x_{10})\cdot\left(\chi_{\{x_1,\cdots,x_{10},0\}}+\chi_{\{x_1,\cdots,x_{10},1\}}\right)=0.
\end{aligned}$$

Since for all s such that $\{x_1,\cdots,x_{10}\}\subset s$, $\hat{f}(s)=0$, we only need to consider s such that $s\subset\{x_1,\cdots,x_{10}\}$.

If we represent the function AND as follows:

$$AND(x_1,\cdots,x_{10})=\prod_{i=1}^{10}\left(\frac{1-\chi_{\{x_i\}}}{2}\right), \quad \chi_{\{x_i\}}=\begin{cases}1, & \text{if } x_i=0,\\ -1, & \text{otherwise}\end{cases},$$

Then the Fourier coefficients are:

$$\hat{f}(s)=\begin{cases}\frac{(-1)^{|s|}}{2^{10}}, & \text{if } s\subset\{x_1,\cdots,x_{10}\},\\ 0, & \text{otherwise}\end{cases}.$$

12.2 Approximation of A Function via Its Spectrum

Assume that we have a function f and its Fourier coefficient $\hat{f}(\beta) = 0.9$ for some β.

Let $g(x) = 0.9\chi_\beta(x)$. Then the error between f and g is:

$$\begin{aligned} E_x\left[|f-g|^2\right] &= \sum_{\alpha\neq\beta} \hat{f}^2(\alpha) + \hat{f}^2(\beta) - \hat{f}^2(\beta) \\ &= 1 - 0.81 = 0.19 \end{aligned}$$

Therefore, with some large Fourier coefficient, we can get somewhat tight bound for the approximation error.

Now, assume f is a boolean function, g is defined as follows:

$$T \subset S, \quad g(x) = \sum_{t\subset T} \hat{f}(t)\chi_t(x).$$

Take the sign of g as $Sign(g)$, then the following inequality holds.

$$E_x\left[f(x) \neq Sign(g(x))\right] \leq E_x\left[|f-g|^2\right]$$

Proof: Let $I\left(f(x) \neq Sign(g(x))\right) = 1, I\left(f(x) = Sign(g(x))\right) = 0$. Then,

$$\begin{aligned} I\left(f(x) \neq Sign(g(x))\right) &\leq |f(x) - g(x)| \\ \Rightarrow I\left(f(x) \neq Sign(g(x))\right) &\leq |f(x) - g(x)|^2 \\ \Rightarrow E\left[I\left(f(x) \neq Sign(g(x))\right)\right] &\leq E\left[|f(x) - g(x)|^2\right] \\ \Rightarrow Pr_x\left[f(x) \neq Sign(g(x))\right] &\leq E\left[|f(x) - g(x)|^2\right] \\ \Rightarrow Pr_x\left[f(x) \neq Sign(g(x))\right] &\leq \sum_{s\subset[n]} \left(\hat{f}(s) - \hat{g}(s)\right)^2. \end{aligned}$$

■

12.3 Compute $\hat{f}(s)$ from Examples

12.3.1 SAMPLE Algorithm

The algorithm consists of two steps.

- Take m samples $x_1, \cdots, x_m$ uniformly at random.
- Output

$$a_s = \frac{1}{m}\sum_{i=1}^{m} f(x_i)\cdot\chi_s(x_i).$$

a_s is an estimated value of $\langle f, \chi_s \rangle$ or $E_{x \in \{0,1\}^n}[f \cdot \chi_s]$.

Theorem 1 Chernoff Bound

Let $x_1, \cdots, x_m$ are i.i.d random variables such that $E[x_i] = p$. Then,

$$Pr\left[\left|\frac{1}{m}\sum x_i - p\right| > \lambda\right] \le e^{-\frac{d^2 m}{2}}.$$

Therefore, if $m > \frac{2\ln\left(\frac{2}{\delta}\right)}{\lambda^2}$, a_s is within λ of $\hat{f}(s)$ with the probability $\ge 1 - \delta$.

12.3.2 Approach 1: Low Degree Algorithm

Definition 1 *f has (α, d) degree, if*

$$\sum_{|s| \le d} \hat{f}^2(s) \ge 1 - \alpha.$$

If a function f has (α, d) degree, then there exists an $n^{O(d)}$ time algorithm for learning f with the accuracy $\ge 1 - \alpha$.

12.3.3 Approach 2

Definition 2 *f is t-sparse, if f has at most t non-zero Fourier coefficients.*

We can use queries to find these t coefficients.

References

[1] N. Linial, Y. Mansour, and N. Nisan, "Constant Depth Circuits, Fourier Transfrom and Learnability," *Journal of the ACM*, 40(3):607–620, 1993.