Notes for Lecture 27

Learning Bounded-Depth Circuits

In this lecture we discuss a quasi-polynomial time algorithm to learn bounded-depth circuits with queries under the uniform distribution. As in Lecture 15 and 16, we denote bits using the set $\{-1, 1\}$ instead of $\{0, 1\}$.

The main result is

Theorem 1 (Main) *Given oracle access to a function $f : \{-1, 1\}^n \to \{-1, 1\}$ computed by a circuit of size S and depth d, and given parameters ϵ and δ, in time polynomial in $n^{O((14 \log S/\epsilon)^{d-1})}$ and $\log 1/\delta$ we can output a circuit that, with probability at least $1 - \delta$, is ϵ-close to f.*

As in Lectures 15 and 16, we divide the proof in two parts. We show that every function whose Fourier coefficients satisfy certain properties can be efficiently learned, and then we show that functions computed by bounded depth circuits satisfy these properties.

Lemma 2 *Given oracle access to $f : \{-1, 1\}^n \to \{-1, 1\}$ such that*

$$\sum_{S:|S|\leq t} |\hat{f}_S| \leq m$$

and

$$\sum_{S:|S|>t} \hat{f}_S^2 \leq \alpha$$

and given parameters $\epsilon, \delta > 0$, it is possible to construct in time polynomial in n, m, $1/\epsilon$, $\log 1/\delta$, a circuit C that, with probability at least $1 - \delta$, is $\epsilon + \alpha$-close to f.

Lemma 3 *If $f : \{-1, 1\}^n \to \{-1, 1\}$ is computed by a circuit of size S and depth d, then*

$$\sum_{S:|S|>t} \hat{f}_S^2 \leq 2\alpha$$

for $t := 28 \cdot (14 \log(S/\alpha))^{d-1}$.

Given the two Lemmas, we just need to observe that $\sum_{|S|\leq t} |\hat{f}_S| \binom{n}{t} < n^t$ for every boolean function f, and the proof of Theorem 1 follows.

1 More on Learning and Fourier Analysis

In this section we prove Lemma 2. First recall the version of the Goldreich-Levin algorithm we presented in Lecture 15.

Lemma 4 *There is a probabilistic algorithm that given oracle access to a function $f : \{-1,1\}^n \to \{-1,1\}$, a threshold parameter τ, a confidence parameter δ and an accuracy parameter γ, runs in time polynomial in n, γ^{-1}, τ^{-1} and $\log \delta^{-1}$ and outputs a list L of $O(\tau^{-2} \log \tau^{-1} \delta^{-1})$ sets, and a value $\bar{f}_S$ for every set $S \in L$, such that, with probability at least $1-\delta$ the following conditions hold:*

- *Every set S such that $|\hat{f}_S| \geq \tau$ is in the list;*
- *For every set S in the list, $|\hat{f}_S - \bar{f}_S| \leq \gamma$.*

We fix $\tau = \epsilon/2m$. If $\ell = O(\tau^{-2} \log(\tau^{-1}\delta^{-1})$ is an upper bound to the size of the list returned by the Goldreich-Levin algorithm with threshold τ and confidence δ, then we fix $\gamma = \sqrt{\epsilon/2\ell}$ and we run the Goldreich-Levin algorithm with threshold τ, confidence δ and accuracy γ. We find a list L of sets and values $\bar{f}_S$ for each set in the list such that, with probability $\geq 1-\delta$ over the internal coin tosses of the algorithm, we have:

- Every set S such that $|\hat{f}_S| \geq \tau$ is in the list;
- For every set S in the list, $|\hat{f}_S - \bar{f}_S| \leq \gamma$.

Then we define the function $h(x) = \sum_{S \in L,\ |S| \leq t} \bar{f}_S u_S$. The Fourier coefficients of the difference $d(x) := f(x) - h(x)$ are as follows.

- If $|S| > t$, then $\hat{h}_S = 0$ and so $\hat{d}_S = \hat{f}_S$.
- If $|S| \leq t$ and $S \notin L$, then $\hat{h}_S = 0$ and so $\hat{d}_S = \hat{f}_S$, and also $|\hat{d}_S| = |\hat{f}_S| \leq \tau$.
- If $|S| \leq t$ and $S \in L$, then $|\hat{h}_S| \leq \gamma$.

We now want to estimate $\mathbf{E}[(f(x) - h(x))^2]$. We have

$$
\begin{aligned}
\mathbf{E}[(f(x) - h(x))^2] &= \mathbf{E}[d^2(x)] \\
&= \sum_S \hat{d}_S^2 \\
&= \sum_{|S|>t} \hat{d}_S^2 + \sum_{|S| \leq t,\ S \notin L} \hat{d}_S^2 + \sum_{|S| \leq t,\ S \in L} \hat{d}_S^2 \\
&\leq \sum_{|S|>t} \hat{f}_S^2 + \tau \sum_{|S| \leq t, S \notin L} |\hat{d}_S| + \sum_{|S| \leq t,\ S \in L} \gamma^2 \\
&\leq \alpha + \tau m + |L| \gamma^2 \\
&\leq \alpha + \epsilon
\end{aligned}
$$

Define $g : \{-1,1\}^n \to \{-1,1\}$ such that $g(x) = 1$ if $h(x) \geq 0$ and $g(x) = -1$ if $h(x) < 0$. We see that

$$\mathbf{Pr}[g(x) \neq f(x)] \leq \mathbf{E}[(f(x) - h(x))^2] \leq \alpha + \epsilon$$

We output the circuit that computes g.

2 The Fourier Spectrum of Functions Computed by Bounded-Depth Circuits

In this section we prove Lemma 3.

We state without proof the following result of Linial, Mansour and Nisan. Denote by $DTD(f)$ the depth of the shallowest decision tree that computes f.

Lemma 5 (Linial-Mansour-Nisan) *Let R_p be the distribution over random restrictions with parameter p, and let $f : \{-1,1\}^n \to \{-1,1\}$ be a function.*

$$\sum_{|S|>t} \hat{f}_S^2 \leq 2\mathbf{Pr}_{\rho\sim R_p}[DTD(f_\rho) \geq tp/2]$$

In other words, if f is a function that is very likely to be computable by a shallow decision tree after a random restriction, then f has low weight on large coefficient.

Lemma 6 *If f can be computed by a circuit of size C and depth d, then for $p = 1/196 \cdot (\log S/\alpha)^{d-2}$*

$$\mathbf{Pr}_{\rho\sim R_p}[DTD(f_\rho) \geq \log(S/\alpha)] \leq \alpha$$

Proof: Define $k := (\log S/\alpha)$, and let $s_1, \ldots, s_d$ be the number of gates at depth $1, \ldots, d$ in C. First apply a random restriction with parameter $1/14$, and consider the s_1 top gates, that we can think of as being 1-DNF (respectively, 1-CNF). Then, after the random restriction, each gate has probability at least $1 - 2^{-k}$ of being expressible as a decision tree of depth k, and so as a k-CNF (respectively, k-DNF). Substitute such an expression in each gate, and obtain a circuit of depth $d+1$ with top gates of fan-in k and two consecutive layers of AND gates (respectively, OR gates). By using associativity, we get a circuit of depth d, with top gates of fan-in k, and with $s_2, \ldots, s_d$ gates at levels $2, \ldots, d$. (With probability $\leq s_1/2^k$, this first step fails.) The circuit has $n/14$ inputs

Then we apply a random restriction with parameter $1/(14k)$, and consider the s_2 level-2 gates. Each of them computed a k-CNF (respectively, k-DNF) before the restriction. After the restriction, each such gate computes a function computable by a depth-k decision tree, except with probability at most 2^{-k}. By replacing each gate with a k-DNF (respectively, k-CNF) and collapsing level 2 with level 3, we get a circuit of depth $d-1$, with top gates of fan-in k, and with $s_3, \ldots, s_d$ gates at levels $2, \ldots, d_1$. (with probability $\leq s_2/2^k$, this second step fails.) The circuit has now $n/(14 \cdot (14k))$ inputs.

After $d-3$ steps like the one above, we get a circuit of depth 2, with top gates having fan-in $\leq k$. The circuits has $n/(14 \cdot (14k)^{d-2})$ inputs. There is a probability at most $((s_1 + \cdots + s_{d-1})/2^k)$ that the construction failed at any point. We apply one more random restriction with $p = 1/14$, and the circuit finally becomes a depth-k decision tree, except with probability at most 2^{-k}. Putting every together, and recalling $S = s_1 + \cdots + s_{d-1} + 1$ and $2^k = S/\alpha$, we have that for $p = 1/(196 \cdot (14k)^{d-2}$

$$\mathbf{Pr}_{\rho\sim R_p}[DTD(f_\rho) \geq k] \leq \alpha$$

as desired. □

In order to prove Lemma 3, it now remains to fix $t = 28 \cdot (14 \log(S/\alpha)^{d-1})$ and $p = 1/(196(14 \log(S/\alpha))^{d-2})$, and deduce

$$\sum_{|S|>t} \hat{f}_S^2 \leq 2\mathbf{Pr}_{\rho \sim R_p}[DTD(f_\rho) \geq tp/2] = 2\mathbf{Pr}_{\rho \sim R_p}[DTD(f_\rho) \geq \log S/\alpha] \leq 2\alpha$$

3 References

These results are due to Linial, Mansour and Nisan [LMN93].

References

[LMN93] N. Linial, Y. Mansour, and N. Nisan. Constant depth circuits, fourier transform and learnability. *Journal of the ACM*, 40(3):607–620, 1993. 4