6.842 Randomness and Computation April 13, 2022

Homework 4

Yuchong Pan *MIT ID: 911346847*

1. *Collaborators and sources:* Guanghao Ye, Zixuan Xu.

   Proof. Let L be a subset of the left vertices such that $|L| \leq n/2$. If $L = \emptyset$, then the result trivially holds. Hence, assume that $|L| \geq 1$. Let R_0 be the set of the right vertices. Then

$$\begin{aligned}
\mathbb{P}[|N(L)| < (1+\varepsilon)|L|] &\leq \mathbb{P}\left[\exists R \subset R_0, |R| = \lfloor (1+\varepsilon)|L| \rfloor, N(L) \subset R\right] \\
&\leq \sum_{\substack{R \subset R_0 \\ |R| = \lfloor (1+\varepsilon)|L| \rfloor}} \mathbb{P}[N(L) \subset R] && \text{(union bound)} \\
&= \sum_{\substack{R \subset R_0 \\ |R| = \lfloor (1+\varepsilon)|L| \rfloor}} \left(\frac{|R|}{n} \cdot \frac{|R|-1}{n-1} \cdots \frac{|R|-|L|+1}{n-|L|+1} \right)^3 \\
&\leq \binom{n}{\lfloor (1+\varepsilon)|L| \rfloor} \left(\frac{\lfloor (1+\varepsilon)|L| \rfloor}{n} \right)^{3|L|} && \text{(for } \varepsilon \leq 1\text{)} \\
&\leq \left(\frac{en}{\lfloor (1+\varepsilon)|L| \rfloor} \right)^{\lfloor (1+\varepsilon)|L| \rfloor} \left(\frac{\lfloor (1+\varepsilon)|L| \rfloor}{n} \right)^{3|L|} && \text{(Stirling's approximation)} \\
&\leq \left(\frac{en}{\lfloor (1+\varepsilon)|L| \rfloor} \right)^{(1+\varepsilon)|L|} \left(\frac{\lfloor (1+\varepsilon)|L| \rfloor}{n} \right)^{3|L|} && \text{(for } \varepsilon \leq 2e-1\text{)} \\
&= \left(e^{1+\varepsilon} \left(\frac{\lfloor (1+\varepsilon)|L| \rfloor}{n} \right)^{2-\varepsilon} \right)^{|L|} \\
&\leq \left(e^{1+\varepsilon} \left(\frac{(1+\varepsilon)|L|}{n} \right)^{2-\varepsilon} \right)^{|L|} \\
&\leq \left(e^{1+\varepsilon} \left(\frac{(1+\varepsilon)}{2} \right)^{2-\varepsilon} \right)^{|L|}. && \text{(since } |L| \leq n/2\text{)}
\end{aligned}$$

   Let $\varepsilon = 1/2$. Then $0 < e^{1+\varepsilon}((1+\varepsilon)/2)^{2-\varepsilon} < 1/2$. Since $|L| \geq 1$, then

$$\begin{aligned}
\mathbb{P}[|N(L)| \geq (1+\varepsilon)|L|] &\geq 1 - \mathbb{P}[|N(L)| < (1+\varepsilon)|L|] \\
&\geq 1 - \left(e^{1+\varepsilon} \left(\frac{(1+\varepsilon)}{2} \right)^{2-\varepsilon} \right)^{|L|} \\
&\geq 1 - e^{1+\varepsilon} \left(\frac{(1+\varepsilon)}{2} \right)^{2-\varepsilon} \\
&> 1 - \frac{1}{2} = \frac{1}{2}.
\end{aligned}$$

   This completes the proof. $\square$

2. (a) *Collaborators and sources:* none.

Proof. Note that $\mathbb{1}_{\text{test accepts}} = (1 + f(x)f(y)f(z))/2$. By the Fourier transform of f and by linearity of expecation,

$$\begin{aligned}
\mathbb{E}[f(x)f(y)f(z)] &= \mathbb{E}\left[\left(\sum_{S\subset[n]} \hat{f}(S)\chi_S(x)\right)\left(\sum_{T\subset[n]} \hat{f}(T)\chi_T(y)\right)\left(\sum_{U\subset[n]} \hat{f}(U)\chi_U(z)\right)\right] \\
&= \sum_{S,T,U\subset[n]} \hat{f}(S)\hat{f}(T)\hat{f}(U)\,\mathbb{E}\left[\chi_S(x)\chi_T(y)\chi_U(x\circ y\circ w)\right].
\end{aligned}$$

Let $S, T, U \subset [n]$. For all $i \in [n]$, since $x_i, y_i \in \{\pm 1\}$, then $x_i^2 = y_i^2 = 1$. Hence,

$$\begin{aligned}
\chi_S(x)\chi_T(y)\chi_U(x\circ y\circ w) &= \left(\prod_{i\in S} x_i\right)\left(\prod_{i\in T} y_i\right)\left(\prod_{i\in U} x_i y_i w_i\right) \\
&= \left(\prod_{i\in S\cap U} x_i^2\right)\left(\prod_{i\in T\cap U} y_i^2\right)\left(\prod_{i\in S\triangle U} x_i\right)\left(\prod_{i\in T\triangle U} y_i\right)\left(\prod_{i\in U} w_i\right) \\
&= \chi_{S\triangle U}(x)\chi_{T\triangle U}(y)\chi_U(w).
\end{aligned}$$

If $S = T = U$, since $w_1, \ldots, w_n$ are all chosen independently and since $\mathbb{E}[w_i] = (-1)\cdot\delta + 1\cdot(1-\delta) = 1 - 2\delta$ for all $i \in [m]$, then

$$\mathbb{E}\left[\chi_{S\triangle U}(x)\chi_{T\triangle U}(y)\chi_U(w)\right] = \mathbb{E}\left[\prod_{i\in S} w_i\right] = \prod_{i\in S}\mathbb{E}\left[w_i\right] = (1-2\delta)^{|S|}.$$

Now, suppose that either $S \neq U$ or $T \neq U$. WLOG assume that $S \neq U$. Then $S\triangle U \neq \emptyset$. Let $j \in S\triangle U$. For $x \in \{\pm 1\}^n$, let $x^{\oplus j}$ be the vector obtained by flipping the j^{th} bit in x. Then we can partition $\{\pm 1\}^n$ into (unordered) pairs $(x, x^{\oplus j})$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[\chi_{S\triangle U}(x)\right] &= \frac{1}{2^n}\sum_{x\in\{\pm1\}^n}\chi_{S\triangle U}(x) = \frac{1}{2^n}\sum_{\text{pairs }(x,x^{\oplus j})}\left(\chi_{S\triangle U}(x) + \chi_{S\triangle U}\left(x^{\oplus j}\right)\right) \\
&= \frac{1}{2^n}\sum_{\text{pairs }(x,x^{\oplus j})}\left(x_j\prod_{i\in(S\triangle U)\setminus\{j\}} x_i + (-x_j)\prod_{i\in(S\triangle U)\setminus\{j\}} x_i\right) = 0.
\end{aligned}$$

Since x, y and w are chosen independently, then for all $S, T, U \subset [n]$ such that either $S \neq U$ or $T \neq U$,

$$\mathbb{E}\left[\chi_{S\triangle U}(x)\chi_{T\triangle U}(y)\chi_U(w)\right] = \mathbb{E}\left[\chi_{S\triangle U}(x)\right]\mathbb{E}\left[\chi_{T\triangle U}(y)\right]\mathbb{E}\left[\chi_U(w)\right] = 0.$$

Therefore,

$$\begin{aligned}
\mathbb{P}[\text{test accepts}] &= \mathbb{E}\left[\mathbb{1}_{\text{test accepts}}\right] = \mathbb{E}\left[\frac{1 + f(x)f(y)f(z)}{2}\right] = \frac{1}{2} + \frac{1}{2}\mathbb{E}[f(x)f(y)f(z)] \\
&= \frac{1}{2} + \frac{1}{2}\sum_{S,T,U\subset[n]}\hat{f}(S)\hat{f}(T)\hat{f}(U)\,\mathbb{E}\left[\chi_{S\triangle U}(x)\chi_{T\triangle U}(y)\chi_U(w)\right] \\
&= \frac{1}{2} + \frac{1}{2}\sum_{S\subset[n]}(1-2\delta)^{|S|}\hat{f}(S)^3.
\end{aligned}$$

This completes the proof. $\square$

(b) *Collaborators and sources:* none.

Proof. Let $f : \{\pm 1\}^n \to \{\pm 1\}$ be a dictator function. Then $f = \chi_{\{j\}}$ for some $j \in [n]$. Therefore, $\hat{f}(\{j\}) = 1$ and $\hat{f}(S) = 0$ for all $S \subset [n]$ with $S \neq \{j\}$. By part (a),

$$\begin{aligned}
\mathbb{P}[\text{test accepts}] &= \frac{1}{2} + \frac{1}{2} \sum_{S \subset [n]} (1-2\delta)^{|S|} \hat{f}(S)^3 \\
&= \frac{1}{2} + \frac{1}{2} \left((1-2\delta)^{|\{j\}|} \hat{f}(\{j\})^3 + \sum_{\substack{S \subset [n] \\ S \neq \{j\}}} (1-2\delta)^{|S|} \hat{f}(S)^3 \right) \\
&= \frac{1}{2} + \frac{1}{2} \left((1-2\delta)^1 \cdot 1^3 + \sum_{\substack{S \subset [n] \\ S \neq \{j\}}} (1-2\delta)^{|S|} \cdot 0^3 \right) \\
&= \frac{1}{2} + \frac{1}{2}(1-2\delta) = 1 - \delta.
\end{aligned}$$

This completes the proof. □

(c) *Collaborators and sources:* none.

Proof. Let $f : \{\pm 1\}^n \to \{\pm 1\}$ be such that f passes with probability at least $1 - \varepsilon$ for some $\varepsilon \in (0, 1/2)$. By part (a),

$$1 - \varepsilon \leq \mathbb{P}[\text{test accepts}] = \frac{1}{2} + \frac{1}{2} \sum_{S \subset [n]} (1 - 2\delta)^{|S|} \hat{f}(S)^3.$$

Rearranging the above inequality and applying Parseval's identity yield

$$\begin{aligned} 1 - 2\varepsilon &\leq \sum_{S \subset [n]} (1 - 2\delta)^{|S|} \hat{f}(S)^3 \leq \left(\max_{S \subset [n]} (1 - 2\delta)^{|S|} \hat{f}(S) \right) \sum_{S \subset [n]} \hat{f}(S)^2 \\ &= \left(\max_{S \subset [n]} (1 - 2\delta)^{|S|} \hat{f}(S) \right) \cdot 1 = \max_{S \subset [n]} (1 - 2\delta)^{|S|} \hat{f}(S). \end{aligned}$$

Hence, there exists $S \subset [n]$ such that $(1-2\delta)^{|S|} \hat{f}(S) \geq 1 - 2\varepsilon$. Set $\delta = \varepsilon$ in the test. Then $(1 - 2\varepsilon)^{|S|} \hat{f}(S) \geq 1 - 2\varepsilon$. Since $\varepsilon \in (0, 1/2)$, then $1 - 2\varepsilon \in (0, 1)$, so $(1 - 2\varepsilon)^{|S|} \in (0, 1]$. Therefore,

$$\hat{f}(S) \geq \frac{1 - 2\varepsilon}{(1 - 2\varepsilon)^{|S|}} \geq \frac{1 - 2\varepsilon}{1} = 1 - 2\varepsilon.$$

This completes the proof. $\square$

(d) *Collaborators and sources:* none.

By part (c), if f passes with probability at least $1-\varepsilon$ for some $\varepsilon \in (0, 1/2)$, then there exists $S \subset [n]$ such that $(1-2\varepsilon)^{|S|}\hat{f}(S) \geq 1-2\varepsilon$ by setting $\delta = \varepsilon$ in the test. Since $\operatorname{dist}(f, \chi_S) \in [0,1]$, then $\hat{f}(S) = 1 - 2\operatorname{dist}(f, \chi_S) \in [-1, 1]$. Since $\varepsilon \in (0, 1/2)$, then $1-2\varepsilon \in (0,1)$. If $|S| \geq 2$, then $0 < (1-2\varepsilon)^{|S|} < 1-2\varepsilon$, so $(1-2\varepsilon)^{|S|}\hat{f}(S) < 1-2\varepsilon$, a contradiction. Therefore, one of the following two cases holds:

(i) $|S| = 1$ and $\hat{f}(S) = 1$ (so $\operatorname{dist}(f, \chi_S) = 0$, and $f = \chi_S$ is a dictator function);
(ii) $|S| = 0$ and $\hat{f}(S) \geq 1 - 2\varepsilon$ (so $\operatorname{dist}(f, \chi_\emptyset) \leq \varepsilon$).

Hence, if f is ε-close to $\chi_\emptyset \equiv 1$ (a non-dictator function), then f also passes with probability at least $1-\varepsilon$.

Note that for any dictator function, say $\chi_{\{j\}}$ for some $j \in [n]$,

$$\mathop{\mathbb{P}}_{x \in \{\pm 1\}^n}\left[\chi_{\{j\}}(x) = 0\right] = \mathop{\mathbb{P}}_{x \in \{\pm 1\}^n}\left[x_j = 0\right] = \frac{|\{x \in \{\pm 1\}^n : x_j = 0\}|}{2^n} = \frac{2^{n-1}}{2^n} = \frac{1}{2}.$$

In other words, any dictator function equals 0 for half of the inputs, and 1 for the other half. We give a simple fix to the test by applying the following new test before the original test. For any sufficiently small $\eta > 0$, we independently and uniformly sample $\Theta(\log(1/\eta))$ random inputs from $\{\pm 1\}^n$, and reject if and only if more than $3/4$ of the values are 1. If f is ε-close to $\chi_\emptyset \equiv 1$ for some $\varepsilon \in (0, 1/8)$, then by the Chernoff bound,

$$\mathbb{P}[\text{new test rejects } f] = 1 - \mathbb{P}[\leq 3/4 \text{ of the values are } 1] \geq 1 - e^{-\Theta(\log(1/\eta))} = 1 - \Theta(\eta).$$

On the other hand, if f is a dictator function, then by the Chernoff bound,

$$\mathbb{P}[\text{new test accepts } f] = 1 - \mathbb{P}[> 3/4 \text{ of the values are } 1] \geq 1 - e^{-\Theta(\log(1/\eta))} = 1 - \Theta(\eta).$$

Hence, if f passes the combination of the new test and the original test with probability at least $1-\varepsilon$ and with $\delta = \varepsilon$ in the original test for some sufficiently small $\varepsilon > 0$, then f is a dictator function with probability at least $1 - \Theta(\eta)$; on the other hand, if f is a dictator function, then the union bound implies that f passes the combined test with probability at least $1 - \Theta(\eta) - \delta$. This shows that the combined test is a dictator test.

3. *Collaborators and sources:* Guanghao Ye.

Proof. Let $\mathcal{A}$ be a PAC learning algorithm for a class C that runs in $\mathrm{poly}(\log n, 1/\varepsilon, 1/\delta)$ time. We denote by $\mathrm{error}_{\mathcal{D}}(h) = \mathbb{P}_{x\sim\mathcal{D}}[h(x) \neq f(x)]$ the error of a hypothesis h with respect to f with inputs drawn from distribution $\mathcal{D}$. We denote by $\mathrm{error}_S(h) = |\{x \in S : h(x) \neq f(x)\}|/|S|$ the error of h in the sample set S. We give a PAC learning algorithm in Algorithm 1 with running time $\mathrm{poly}(\log n, 1/\varepsilon, \log(1/\delta))$. Let $\mathcal{D}$ be the distribution of inputs.

$\ell \leftarrow \lceil \log_2(3/\delta) \rceil$
foreach $i \leftarrow 1, \dots, \ell$ **do**
    run $\mathcal{A}$ with accuracy $\varepsilon/2$ and confidence $1/2$, obtaining a hypothesis h_i
$m \leftarrow \lceil (12/\varepsilon^2) \log(6\ell/\delta) \rceil$
foreach $j \leftarrow 1, \dots, m$ **do**
    draw $x_j \sim \mathcal{D}$
$S \leftarrow \{x_1, \dots, x_m\}$
$i^* \leftarrow \arg\min_{i\in[\ell]}(\mathrm{error}_S(h_i))$
return h_{i^*}

Algorithm 1: A PAC learning algorithm with running time $\mathrm{poly}(\log n, 1/\varepsilon, \log(1/\delta))$, given accuracy $\varepsilon > 0$ and confidence $\delta > 0$.

Since each call to $\mathcal{A}$ runs in $\mathrm{poly}(\log n, 1/\varepsilon)$ time, then the running time of Algorithm 1 is

$$O\left(\log\frac{1}{\delta}\right)\mathrm{poly}\left(\log n, \frac{1}{\varepsilon}\right) + O\left(\frac{1}{\varepsilon^2}\log\frac{\log\frac{1}{\delta}}{\delta}\right) = \mathrm{poly}\left(\log n, \frac{1}{\varepsilon}, \log\frac{1}{\delta}\right).$$

First, since $\mathbb{P}[\mathrm{error}_{\mathcal{D}}(h_i) \leq \varepsilon/2] \geq 1 - 1/2 = 1/2$ for each $i \in [\ell]$, then

$$\mathbb{P}\left[\exists i \in [\ell], \mathrm{error}_{\mathcal{D}}(h_i) \leq \varepsilon/2\right] \geq 1 - \left(1 - \frac{1}{2}\right)^{\ell} = 1 - \left(\frac{1}{2}\right)^{\lceil \log_2\left(\frac{3}{\delta}\right)\rceil} \geq 1 - \left(\frac{1}{2}\right)^{\log_2\left(\frac{3}{\delta}\right)} = 1 - \frac{\delta}{3}.$$

Second,

$$\begin{aligned}
&\mathbb{P}\left[\left|\mathrm{error}_{\mathcal{D}}(h_i) - \mathrm{error}_S(h_i)\right| < \frac{\varepsilon}{4}\ \forall i \in [\ell]\right] \\
&= 1 - \mathbb{P}\left[\exists i \in [\ell], \left|\mathrm{error}_{\mathcal{D}}(h_i) - \mathrm{error}_S(h_i)\right| \geq \frac{\varepsilon}{4}\right] \\
&\geq 1 - \sum_{i=1}^{\ell}\mathbb{P}\left[\left|\mathrm{error}_{\mathcal{D}}(h_i) - \mathrm{error}_S(h_i)\right| \geq \frac{\varepsilon}{4}\right] && \text{(union bound)} \\
&= 1 - \sum_{i=1}^{\ell}\mathbb{P}\left[\left|\mathrm{error}_{\mathcal{D}}(h_i) - \mathrm{error}_S(h_i)\right| \geq \frac{\varepsilon}{4\,\mathrm{error}_{\mathcal{D}}(h_i)}\cdot\mathrm{error}_{\mathcal{D}}(h_i)\right] \\
&\geq 1 - \sum_{i=1}^{\ell} 2\exp\left(-\frac{1}{3}m\,\mathrm{error}_{\mathcal{D}}(h_i)\cdot\left(\frac{\varepsilon}{4\,\mathrm{error}_{\mathcal{D}}(h_i)}\right)^2\right) && \text{(Chernoff bound)} \\
&= 1 - \sum_{i=1}^{\ell} 2\exp\left(-\frac{m\varepsilon^2}{12\,\mathrm{error}_{\mathcal{D}}(h_i)}\right)
\end{aligned}$$

$$
\begin{aligned}
&\geq 1 - \ell \cdot 2 \exp\left(-\frac{\left\lceil \frac{12}{\varepsilon^2} \log \frac{6\ell}{\delta} \right\rceil \cdot \varepsilon^2}{12 \cdot 1}\right) \\
&\geq 1 - 2\ell \exp\left(-\frac{\frac{12}{\varepsilon^2} \log \frac{6\ell}{\delta} \cdot \varepsilon^2}{12}\right) \\
&= 1 - 2\ell \exp\left(-\log \frac{6\ell}{\delta}\right) \\
&= 1 - 2\ell \cdot \frac{\delta}{6\ell} \\
&= 1 - \frac{\delta}{3}.
\end{aligned}
$$

Since i^* minimizes $\mathrm{error}_S(h_i)$ over $i \in [\ell]$, then by the union bound,

$$
\begin{aligned}
\mathbb{P}\left[\mathrm{error}_S\left(h_{i^*}\right) < \frac{3\varepsilon}{4}\right] &\geq \mathbb{P}\left[\exists i \in [\ell], \mathrm{error}_S\left(h_i\right) < \frac{3\varepsilon}{4}\right] \\
&\geq \mathbb{P}\left[\left(\left|\mathrm{error}_{\mathcal{D}}\left(h_i\right) - \mathrm{error}_S\left(h_i\right)\right| < \frac{\varepsilon}{4}\ \forall i \in [\ell]\right) \wedge \left(\exists i \in [\ell], \mathrm{error}_{\mathcal{D}}\left(h_i\right) \leq \frac{\varepsilon}{2}\right)\right] \\
&\geq 1 - \left(\frac{\delta}{3} + \frac{\delta}{3}\right) \\
&= 1 - \frac{2\delta}{3}.
\end{aligned}
$$

By the union bound again,

$$
\begin{aligned}
\mathbb{P}\left[\mathrm{error}_{\mathcal{D}}\left(h_{i^*}\right) < \varepsilon\right] &\geq \mathbb{P}\left[\left(\left|\mathrm{error}_{\mathcal{D}}\left(h_i\right) - \mathrm{error}_S\left(h_i\right)\right| < \frac{\varepsilon}{4}\ \forall i \in [\ell]\right) \wedge \left(\mathrm{error}_S\left(h_{i^*}\right) < \frac{3\varepsilon}{4}\right)\right] \\
&\geq 1 - \left(\frac{\delta}{3} + \frac{2\delta}{3}\right) \\
&= 1 - \delta.
\end{aligned}
$$

This completes the proof. $\square$