6.842 Randomness and Computation April 16, 2022

Homework 4

Yuchong Pan *MIT ID: 911346847*

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

Proof. Note that if $n \in \{1,2\}$, then the statement holds trivially. Therefore, WLOG assume that $n \geq 2$. Let L_0 be the set of the left vertices in the graph. For each $L \subset L_0$, let $N_1(L)$, $N_2(L)$ and $N_3(L)$ be the neighborhoods of L in the subgraphs of the graph induced by the three random permutations, respectively. Then

$$\begin{aligned}
&\mathbb{P}\left[\exists L \subset L_0, |L| \leq \frac{n}{2}, |N(L)| < (1+\varepsilon)|L|\right] \\
&\leq \sum_{\substack{L\subset L_0\\ 0<|L|\leq \frac{n}{2}}} \mathbb{P}[|N(L)| < (1+\varepsilon)|L|] && \text{(union bound)} \\
&= \sum_{\substack{L\subset L_0\\ 0<|L|\leq \frac{n}{2}}} \mathbb{P}\left[\exists R \subset R_0 \setminus N_1(L), |R| = \lceil \varepsilon|L|\rceil - 1, N_2(L)\cup N_3(L) \subset R\right] \\
&\leq \sum_{\substack{L\subset L_0\\ 0<|L|\leq \frac{n}{2}}} \sum_{\substack{R\subset R_0\setminus N_1(L)\\ |R|=\lceil\varepsilon|L|\rceil-1}} \mathbb{P}\left[N_2(L)\cup N_3(L) \subset R\right] && \text{(union bound)} \\
&= \sum_{\ell=1}^{\frac{n}{2}} \binom{n}{\ell}\binom{n-\ell}{\lceil\varepsilon\ell\rceil-1}\left(\frac{\binom{\ell+\lceil\varepsilon\ell\rceil-1}{\ell}}{\binom{n}{\ell}}\right)^2 \\
&\leq \sum_{\ell=1}^{\frac{n}{2}} \binom{n}{\ell}\binom{n-\ell}{\lfloor\varepsilon\ell\rfloor}\left(\frac{\binom{\ell+\lfloor\varepsilon\ell\rfloor}{\ell}}{\binom{n}{\ell}}\right)^2 && \text{(for } \varepsilon \leq 1/2) \quad (1) \\
&= \sum_{\ell=1}^{\frac{n}{2}} \frac{n!}{\ell!(n-\ell)!}\cdot\frac{(n-\ell)!}{(\lfloor\varepsilon\ell\rfloor)!(n-\ell-\lfloor\varepsilon\ell\rfloor)!}\cdot\left(\frac{\frac{(\ell+\lfloor\varepsilon\ell\rfloor)!}{\ell!\lfloor\varepsilon\ell\rfloor!}}{\frac{n!}{\ell!(n-\ell)!}}\right)^2 \\
&= \sum_{\ell=1}^{\frac{n}{2}} \frac{((n-\ell)!)^2((\ell+\lfloor\varepsilon\ell\rfloor)!)^2}{n!\ell!(\lfloor\varepsilon\ell\rfloor!)^3(n-\ell-\lfloor\varepsilon\ell\rfloor)!}. && (2)
\end{aligned}$$

Stirling's approximation says that for all $k \in \mathbb{N}$,

$$\sqrt{2\pi k}\left(\frac{k}{e}\right)^k e^{\frac{1}{12k+1}} < k! < \sqrt{2\pi k}\left(\frac{k}{e}\right)^k e^{\frac{1}{12k}}. \quad (3)$$

We apply (3) to each factorial in (2); specifically, we apply the upper bound to the numerator and the lower bound to the denominator. First, we count the contributions of e^k from the factorials in each term of (2):

$$\exp\left(n+\ell+3\lfloor\varepsilon\ell\rfloor+(n-\ell-\lfloor\varepsilon\ell\rfloor)-(2(n-\ell)+2(\ell+\lfloor\varepsilon\ell\rfloor))\right) = \exp(0) = 1.$$

Second, we count the contributions of $e^{\frac{1}{12k}}$ in the upper bound and $e^{\frac{1}{12k+1}}$ in the lower bound from the factorials in each term of (2):

$$\exp\left(\frac{1}{12}\left(\frac{2}{n-\ell}+\frac{1}{\ell+\lfloor\varepsilon\ell\rfloor}\right)-\left(\frac{1}{12n+1}+\frac{1}{12\ell+1}+\frac{3}{12\lfloor\varepsilon\ell\rfloor+1}+\frac{1}{12(n-\ell-\lfloor\varepsilon\ell\rfloor)+1}\right)\right)$$
$$\le \exp\left(\frac{1}{12}\cdot(2+1)\right)<1.29,$$

by noting that $n-\ell\ge 1$ and $\ell+\lfloor\varepsilon\ell\rfloor\ge 1$ for all $\ell\in[n/2]$. Take $\varepsilon=0.01$. Third, we count the contributions of k^k from the factorials in each term of (2):

$$\begin{aligned}
&\frac{(n-\ell)^{2(n-\ell)}(\ell+\lfloor\varepsilon\ell\rfloor)^{2(\ell+\lfloor\varepsilon\ell\rfloor)}}{n^n\ell^\ell\lfloor\varepsilon\ell\rfloor^{3\lfloor\varepsilon\ell\rfloor}(n-\ell-\lfloor\varepsilon\ell\rfloor)^{n-\ell-\lfloor\varepsilon\ell\rfloor}}\\
&\le\frac{(n-\ell)^{2(n-\ell)}(\ell+\varepsilon\ell)^{2(\ell+\varepsilon\ell)}}{n^n\ell^\ell(\varepsilon\ell)^{3\varepsilon\ell}(n-\ell-\varepsilon\ell)^{n-\ell-\varepsilon\ell}} &&(4)\\
&\le\frac{(1+\varepsilon)^{2(1+\varepsilon)}}{\varepsilon^{3\varepsilon}}\left(\frac{n-\ell}{n}\right)^{n-\ell+\lfloor\varepsilon\ell\rfloor}\left(\frac{\ell}{n}\right)^{\ell-\lfloor\varepsilon\ell\rfloor}\left(1+\frac{\lfloor\varepsilon\ell\rfloor}{n-\ell-\lfloor\varepsilon\ell\rfloor}\right)^{n-\ell-\lfloor\varepsilon\ell\rfloor}\\
&<1.18\cdot 1\cdot\left(\frac{1}{2}\right)^{0.99\ell}\cdot e^{0.01\ell} &&(\text{since } \ell\le n/2)\\
&<1.18\cdot 0.51^\ell.
\end{aligned}$$

Note that (4) is due to the fact that for each term in (1),

$$\binom{n}{\ell}\binom{n-\ell}{\lfloor\varepsilon\ell\rfloor}\left(\frac{\binom{\ell+\lfloor\varepsilon\ell\rfloor}{\ell}}{\binom{n}{\ell}}\right)^2\le\binom{n}{\ell}\binom{n-\ell}{\varepsilon\ell}\left(\frac{\binom{\ell+\varepsilon\ell}{\ell}}{\binom{n}{\ell}}\right)^2,$$

by the monotonicity of the generalized binomial coefficient (with factorials replaced by the gamma function $\Gamma(r)=(r-1)!$). Fourth, we count the contributions of $\sqrt{2\pi k}$ from the factorials in each term of (2). Note that $k^k>\sqrt{2\pi k}$ for all $k\ge 1.99$. Since $n>3$, then $n-\ell\ge n-(1+\varepsilon)\ell\ge 1.99$ for all $\ell\in[n/2]$. Note also that $(1+\varepsilon)\ell\ge 2$ for all $\ell\in\{2,\dots,n/2\}$. Hence, the contribution of $\sqrt{2\pi k}$ is not upper bounded by k^k from the factorials in each term only if $\ell=1$, in which case, since $n\ge 3$,

$$\binom{n}{\ell}\binom{n-\ell}{\lfloor\varepsilon\ell\rfloor}\left(\frac{\binom{\ell+\lfloor\varepsilon\ell\rfloor}{\ell}}{\binom{n}{\ell}}\right)^2=\binom{n}{1}\binom{n-1}{\lfloor 0.01\cdot 1\rfloor}\left(\frac{\binom{1+\lfloor 0.01\cdot 1\rfloor}{1}}{\binom{n}{1}}\right)^2=\frac{1}{n}\le\frac{1}{3}.$$

Moreover, the contribution of $\sqrt{2\pi k}$ from each term with $\ell\in\{2,\dots,n/2\}$ is at most $1.18\cdot 0.51^\ell$. Therefore,

$$\begin{aligned}
&\mathbb{P}\left[\exists L\subset L_0, |L|\le\frac{n}{2}, |N(L)|<(1+\varepsilon)|L|\right]\\
&\le\sum_{\ell=1}^{\frac{n}{2}}\binom{n}{\ell}\binom{n-\ell}{\lfloor\varepsilon\ell\rfloor}\left(\frac{\binom{\ell+\lfloor\varepsilon\ell\rfloor}{\ell}}{\binom{n}{\ell}}\right)^2\\
&=\binom{n}{1}\binom{n-1}{\lfloor 0.01\cdot 1\rfloor}\left(\frac{\binom{1+\lfloor 0.01\cdot 1\rfloor}{1}}{\binom{n}{1}}\right)^2+\sum_{\ell=2}^{\frac{n}{2}}\binom{n}{\ell}\binom{n-\ell}{\lfloor\varepsilon\ell\rfloor}\left(\frac{\binom{\ell+\lfloor\varepsilon\ell\rfloor}{\ell}}{\binom{n}{\ell}}\right)^2
\end{aligned}$$

$$
\begin{aligned}
&\le \frac{1}{3} + \sum_{\ell=2}^{\frac{n}{2}} 1 \cdot 1.29 \cdot \left(1.18 \cdot 0.51^{\ell}\right)^2 \\
&< 0.34 + 1.8 \sum_{\ell=2}^{\frac{n}{2}} 0.27^{\ell} \\
&< 0.34 + 1.8 \sum_{\ell=2}^{\infty} 0.27^{\ell} \\
&= 0.34 + 1.8 \cdot 0.27^2 \cdot \frac{1}{1-0.27} \\
&< 0.52.
\end{aligned}
$$

It follows that

$$
\begin{aligned}
\mathbb{P}\left[|N(L)| \ge (1+\varepsilon)|L| \ \forall L \subset L_0, |L| \le \frac{n}{2}\right] &\ge 1 - \mathbb{P}\left[\exists L \subset L_0, |L| \le \frac{n}{2}, |N(L)| < (1+\varepsilon)|L|\right] \\
&> 1 - 0.52 = 0.48.
\end{aligned}
$$

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,

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

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. $\square$

(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),

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

Rearranging the above inequality and applying Parseval's identity yield

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

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

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

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 $j \in [n]$,

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

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. $\square$

4. (a) *Collaborators and sources:* Guanghao Ye.

First, we give an algorithm in Algorithm 2 which, given a set S of samples, finds a consistent decision list h such that $h(x) = f(x)$ for all $x \in S$.

$h \leftarrow$ empty decision list
repeat
    foreach *literal ℓ (i.e., a variable or its negation)* **do**
        $S' \leftarrow \{x \in S : \ell(x) = f(x)\}$
        if *there exists $b \in \{0,1\}$ such that $f(x) = b$ for all $x \in S'$* **then**
            append to h a decision "if $\ell(x)$ then output b"
            $S \leftarrow S \setminus S'$
            break
until $S \neq \emptyset$
return h

Algorithm 2: An algorithm which, given a set S of samples, finds a consistent decision list h such that $h(x) = f(x)$ for all $x \in S$.

We show that Algorithm 2 is correct. Consider an iteration of the **repeat** loop, with set S_0 of remaining samples. Then $S_0 \neq \emptyset$. It suffices to show that there exist a literal ℓ and $b \in \{0,1\}$ such that $S' \neq \emptyset$ and $f(x) = b$ for all $x \in S'$, where $S' = \{x \in S_0 : \ell(x) = f(x)\}$. Since f is a decision list, let "if $\ell^*(x)$ then output b^*" be the first decision in f that has not been added to h so far. Let S_0^* be the set of samples that are not output by decisions prior to ℓ^* in f. Let S^* be the set of samples output by decision "if $\ell^*(x)$ then output b^*" in f. Then $S^* = \{x \in S_0^* : \ell^*(x) = f(x)\}$ and $f(x) = b^*$ for all $x \in S^*$. Let $S' = \{x \in S_0 : \ell^*(x) = f(x)\}$. Since all decisions prior to ℓ^* in f has been added to h, then $S_0 \subset S_0^*$, so $S' \subset S^*$. Therefore, $f(x) = b^*$ for all $x \in S'$. If $S' \neq \emptyset$, then we are done. Otherwise, we run the same argument for the next decision in f, until we find a decision in f for which $S' \neq \emptyset$. We claim that this is possible. To see this, since $\emptyset \neq S_0 \subset S_0^*$, then there exists a decision after literal ℓ^* in f for which $S' \neq \emptyset$. This completes the proof.

Second, we show that $|S| = \ln(n!4^n)$ is necessary to ensure (ε, δ) PAC learning. To see this, we claim that each variable appears at most once in a decision list. Suppose that variable x_i first appears in a decision list f as literal ℓ. Let S_0 be the set of points $\{0,1\}^n$ that remain after this decision. Then $\ell(x) \neq f(x)$ for all $x \in S_0$. This shows that literal ℓ need not appear in h again. Moreover, $\neg\ell(x) = f(x)$ for all $x \in S_0$. At any point after literal ℓ in f, if we were to add literal $\neg\ell$, then there would exist $b \in \{0,1\}$ such that $f(x) = b$ for all remaining x such that $\neg f(x) = f(x)$; since $\neg\ell(x) = f(x)$ for all remaining x, then we can simply end the decision list. Therefore, if $|S| < \ln(n!4^n)$, then we can possibly output wrong answers for 2^{n-1} points.

5. *Collaborators and sources:* Guanghao Ye.

Proof. We apply Occam's Razor; i.e., we give Algorithm 3.

```
draw M = (1/ε)(ln|C| + ln(1/δ)) samples from {0,1}^n, where C is the set of all decision lists
run Algorithm 2 to find a consistent decision list h such that h(x) = f(x) for all x ∈ S
return h
```

Algorithm 3: An algorithm which finds a decision list h such that $\mathbb{P}_{x \sim \mathcal{D}}[f(x) \neq h(x)] < \varepsilon$ with probability at least $1 - \delta$.

First, we claim that each variable appears at most once in a decision list. Suppose that variable x_i first appears in a decision list f as literal ℓ. Let S_0 be the set of points $\{0,1\}^n$ that remain after this decision. Then $\ell(x) \neq f(x)$ for all $x \in S_0$. This shows that literal ℓ need not appear in h again. Moreover, $\neg\ell(x) = f(x)$ for all $x \in S_0$. At any point after literal ℓ in f, if we were to add literal $\neg\ell$, then there would exist $b \in \{0,1\}$ such that $f(x) = b$ for all remaining x such that $\neg f(x) = f(x)$; since $\neg\ell(x) = f(x)$ for all remaining x, then we can simply end the decision list.

Therefore,

$$|\mathcal{C}| = \sum_{k=1}^{n} k!2^k 2^k \leq n \cdot n!4^n \leq n \cdot n^n 4^n = n^{n+1}4^n.$$

It follows that

$$M = \frac{1}{\varepsilon}\left(\ln|\mathcal{C}| + \ln\frac{1}{\delta}\right) \leq \frac{1}{\varepsilon}\left(\ln\left(n^{n+1}4^n\right) + \ln\frac{1}{\delta}\right) = O\left(\frac{1}{\varepsilon}\left(n\log n + \log\frac{1}{\delta}\right)\right).$$

It is easy to see that Algorithm 3 runs in $O(|S| \cdot 2n \cdot n) = O(Mn^2) = O((n^2/\varepsilon)(n\log n + \log(1/\delta)))$ time, which is polynomial in n, $1/\varepsilon$ and $\log(1/\delta)$.

By Occam's Razor, the probability that any decision list h such that $\mathbb{P}_{x \in \mathcal{D}}[f(x) \neq h(x)] \geq \varepsilon$ is not consistent with the samples with probability at most δ. This completes the proof. □