Differentially Private Nonparametric Regression Under a Growth Condition

COLT 2021

Noah Golowich

MIT

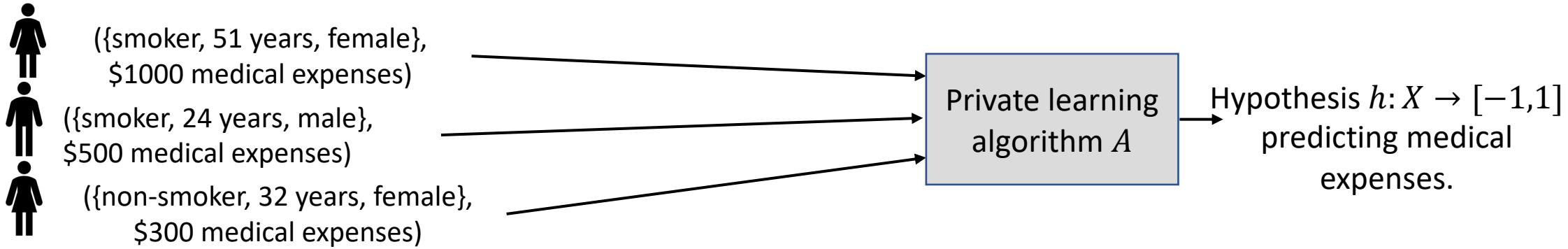
Overview: privacy-preserving regression

- Machine learning models often trained on sensitive data; important to protect privacy of users' data

- Recent development: connection between *private learnability* and *online learnability* *[Alon-Livni-Malliaris-Moran '19] [Bun-Livni-Moran '20] [Ghazi-G.-Kumar-Manurangsi '21] [Jung-Kim-Tewari '20] [G.-Livni, '21]*
  - Focus primarily on **binary classification**, with exception of *[Jung-Kim-Tewari '20]*
  - **This talk:** prove a sufficient condition for private **regression** in terms of online learnability parameters of a real-valued class.

Background: differential privacy

- Collection of individuals, each produces **example** $(x_i, y_i) \in X \times [-1,1]$
- **Dataset** $S_n = \{(x_1, y_1), \ldots, (x_n, y_n)\}$, (randomized) learner A:



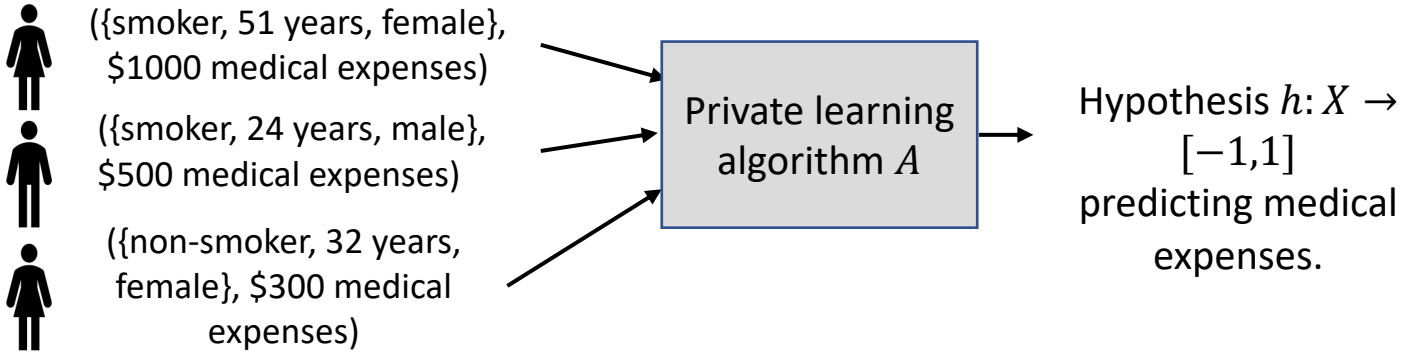
Definition: Algorithm A is **(ϵ, δ)-differentially private (DP)** if for all events E, for all *neighboring datasets* S_n, S_n',

$$\Pr_A[A(S_n) \in E] \leq e^{\epsilon} \cdot \Pr_A[A(S_n') \in E] + \delta$$

Neighboring datasets: those which differ in a single example (x_i, y_i)

In this talk: $\epsilon \leq O(1)$ (e.g., $\epsilon = 0.01$), $\delta < 1/n^{\omega(1)}$ (e.g., $\delta = n^{-\log n}$)

Private Regression



- Given a *known* class H of **hypotheses**, i.e., functions $h: X \to [-1,1]$
- $S_n = \{(x_1, y_1), \ldots, (x_n, y_n)\}$ is drawn i.i.d. from *unknown* distribution P on $X \times [-1,1]$
- Goal: algorithm $A(S_n)$ outputs $\hat{h}: X \to [-1,1]$ minimizing

$$\mathrm{err}_P(\hat{h}) := \mathbb{E}_{(x,y)\sim P}[|\hat{h}(x) - y|]$$

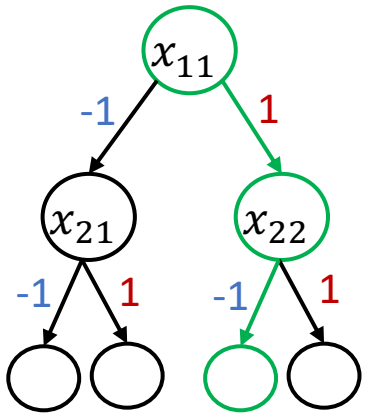
Definition: H is **privately learnable** if: for all η, ϵ, δ, there is n and a (ϵ, δ)-DP algorithm A mapping $S_n \to \hat{h}$ so that for all P, with high probability over S_n,

$$\mathrm{err}_P(\hat{h}) \leq \inf_{h \in H} \mathrm{err}_P(h) + \eta$$

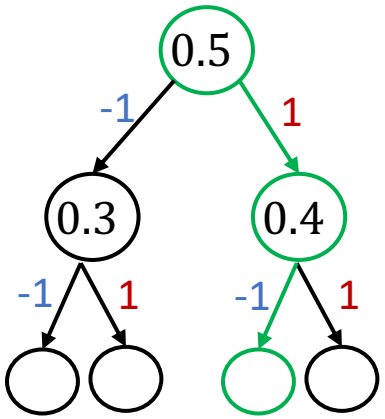
Sequential fat-shattering dimension

- Main (partially open) question: ***Which classes are privately learnable?***
- **Sequential fat-shattering dimension** provides (partial) characterization:

X-labeled tree T:



Assignments:



Defn: Fix $\alpha > 0$, and a binary tree T with all internal nodes labeled by elements of X, edges labeled by $\{-1,1\}$.

- T is **α-shattered** by H if there is an assignment of some element $s \in [-1,1]$ to each node in T so that:
- For each leaf ℓ there is some $h_\ell \in H$ so that for each node $(x,s) \in X \times [-1,1]$ on the root-to-leaf path of ℓ:

$$k \cdot (f(x) - s) \geq \alpha/2$$

- E.g., for the green leaf: need
  - $h_\ell(x_{11}) - 0.5 \geq \alpha/2$; and
  - $-1 \cdot (h_\ell(x_{22}) - 0.4) \geq \alpha/2$

Edge label (-1 or 1) for next edge on root-to-leaf path

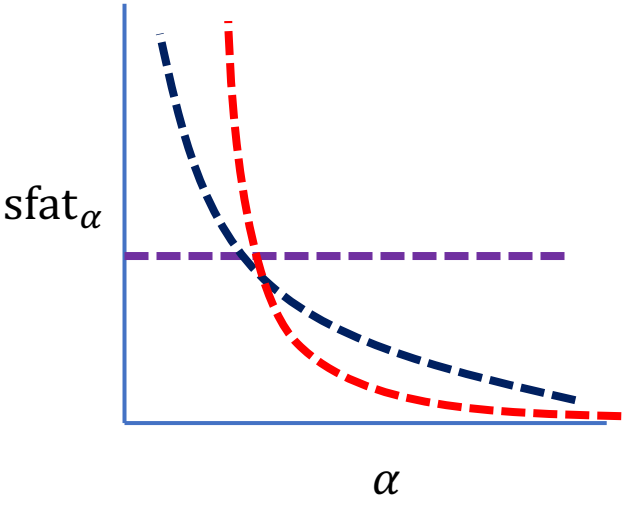
Defn: **α-sequential fat-shattering dimension** of hypothesis class H, denoted $\mathrm{sfat}_\alpha(H)$, is largest d so that there exists tree of depth d that is α-shattered by H.

Observation: $\mathrm{sfat}_\alpha(H)$ is non-increasing function of α

Examples for sequential fat-shattering dimension

Defn: **α-sequential fat-shattering dimension** of hypothesis class H, denoted $\mathrm{sfat}_\alpha(H)$, is largest d so that there exists tree of depth d that is α-shattered by H.

- If H is actually **binary**, i.e., $h(x) \in \{-1,1\}$ for all $h \in H$:
  - $\mathrm{sfat}_\alpha(H)$ equal to *Littlestone dimension* of H for all $\alpha \leq 1$
- **One-dimensional linear regression:**
  - Suppose $H = \{x \to ax + b : |x| \leq 1, |a| \leq 1, |b| \leq 1\}$;
  - Then $\mathrm{sfat}_\alpha(H) \asymp \log(1/\alpha)$
- **Infinite-dimensional linear regression:**
  - Suppose $H = \{x \to \langle w, x\rangle : x, w \in \ell_2^\infty, |x|_2 \leq 1, |w|_2 \leq 1\}$;
  - Then $\mathrm{sfat}_\alpha(H) \asymp 1/\alpha^2$



[Rakhlin-Sridharan-Tewari, '13]: H is *online learnable* iff $\mathrm{sfat}_\alpha(H)$ is finite for all $\alpha > 0$

Prior work: private regression

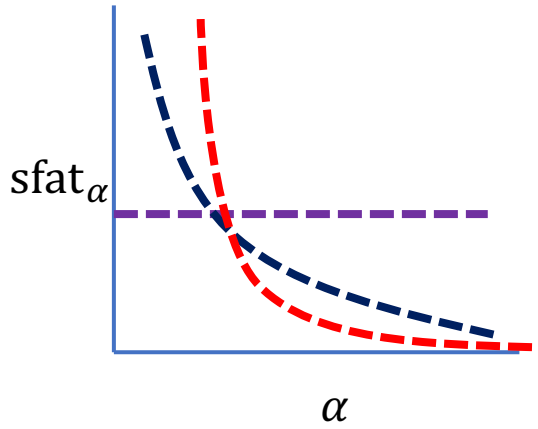
Definition: H is **privately learnable** if: for all η, ϵ, δ, there is (ϵ, δ)-DP algorithm A mapping $S_n \to \hat{h}$ so that for all P, with high probability over S_n,

$$\mathrm{err}_P(\hat{h}) \leq \inf_{h \in H} \mathrm{err}_P(h) + \eta$$

Which real-valued classes H are privately learnable?

- *[Jung-Kim-Tewari, '20]:* **Online learnability is necessary for private learnability:**
  - If H is privately learnable, then $\mathrm{sfat}_\alpha(H)$ is finite for all $\alpha > 0$
- **Open: is online learnability sufficient for private learnability?**
  - *[Jung-Kim-Tewari, '20]:* if $\lim_{\alpha \downarrow 0} \mathrm{sfat}_\alpha(H)$ is finite, then H is privately learnable

Generalizes *[Bun-Livni-Moran, '20],* that finiteness of *Littlestone dimension* is sufficient for private learnability in *binary case*



Our question: can we show private learnability for some classes with $\mathrm{sfat}_\alpha(H)$ **diverging** as $\alpha \downarrow 0$?

Private regression under a growth condition

- Recall: H is a class of hypotheses $h : X \to [-1,1]$.
- P is a distribution on $X \times [-1,1]$.

Theorem (main result): For any $\epsilon, \delta, \alpha \in (0,1)$, for some $n = \frac{2^{\tilde{O}(\mathrm{sfat}_\alpha(H))}}{\epsilon\, \alpha^4}$, there is an (ϵ, δ)-DP algorithm which, given n i.i.d. samples from any distribution P, outputs a hypothesis $\hat{h}$ s.t.:

$$\mathrm{err}_P(\hat{h}) \leq \inf_{h \in H} \mathrm{err}_P(h) + O(\alpha \cdot \mathrm{sfat}_\alpha(H)).$$

Corollary: If $\liminf_{\alpha \downarrow 0} \alpha \cdot \mathrm{sfat}_\alpha(H) = 0$, then H is privately learnable.

Proof overview: weak stability

- Common thread in private learnability is to use algorithms satisfying a notion of ***stability*** *[Bun-Livni-Moran, '20], [Jung-Kim-Tewari, '20], [G.-Ghazi-Kumar-Manurangsi, '21]*
- Fix a class H and $\alpha > 0$, set $d := \mathrm{sfat}_\alpha(H)$.

Lemma (informal; "weak stability"): For any distribution P, there is a hypothesis $\sigma: X \to [-1,1]$, so that for sufficiently large n, there is an algorithm A that takes $(x_1, y_1), \dots, (x_n, y_n) \sim P$ and outputs a set of hypotheses $\hat{h}_1, \dots \hat{h}_M : X \to [-1,1]$ so that:

1. **With probability $\approx \frac{1}{d}$ over the dataset, there is some $\hat{h}_i$ so that $\left|\hat{h}_i - \sigma\right|_\infty \le O(\alpha)$**
2. With high probability, all $\hat{h}_i$ satisfy $\mathrm{err}_P(\hat{h}_i) \le \inf_h \mathrm{err}_P(h) + O(\alpha \cdot d)$
3. $M \le \left(\frac{1}{\alpha}\right)^{2^{\tilde{O}(d)}}$

*This is the **"weak"** part; to get private learner using existing techniques, need $\hat{h}_i = \sigma$*

Above is proved closely following techniques of *[Golowich-Ghazi-Kumar-Manurangsi, '21]*

Strong stability

Weak stability lemma: A outputs $\hat{h}_1, \dots \hat{h}_M : X \to [-1,1]$ so that:

1. With probability $\approx \frac{1}{d}$, there is some $\hat{h}_i$ so that $\left|\hat{h}_i - \sigma\right|_\infty \leq O(\alpha)$
2. With high probability, all $\hat{h}_i$ satisfy $\mathrm{err}_P(\hat{h}_i) \leq O(\alpha \cdot d)$

- Recall H, α fixed, $d = \mathrm{sfat}_\alpha(H)$.
- Main technical contribution is procedure for upgrading *weak stability* (previous slide) to *strong stability*:

Lemma (informal; "weak stability → strong stability"): For any $\hat{g}, \sigma: X \to [-1,1]$ so that $|\hat{g} - \sigma|_\infty \leq O(\alpha)$ (+ some technical conditions), there is an algorithm B that takes only $\hat{g}$ as input and outputs a sequence $\hat{g}_1, \dots, \hat{g}_K$, so that

1. Each $\hat{g}_i$ satisfies $|\hat{g} - \hat{g}_i|_\infty \leq O(\alpha \cdot d)$
2. **There is some $\sigma^\star$, depending only on σ, so that $\sigma^\star = \widehat{g_i}$ for some i**
3. $K \leq \left(\frac{1}{\alpha}\right)^{d^{\tilde{O}(d)}}$

Use algorithm B a total of M times, for $\hat{g} = \hat{h}_1, \dots, \hat{h}_M$.

Strong *stability*

Proof of above lemma: uses notion of **irreducibility** originally developed in *[Ghazi-G.-Kumar-Manurangsi, '21]* for *sample-efficient* private learning in *binary* case

Open Questions

- Main question: can we relax the sufficient condition of

$$\liminf_{\alpha \downarrow 0} \alpha \cdot \mathrm{sfat}_\alpha(H) = 0$$

  for private learnability?
  - Interesting class to consider: infinite-dimensional ℓ_2 linear regression, $\mathrm{sfat}_\alpha(H) \asymp 1/\alpha^2$.
- Can we improve the sample complexity to polynomial in $\mathrm{sfat}_\alpha(H)$?
  - Current sample complexity is exponential in $\mathrm{sfat}_\alpha(H)$

Thank you for listening!