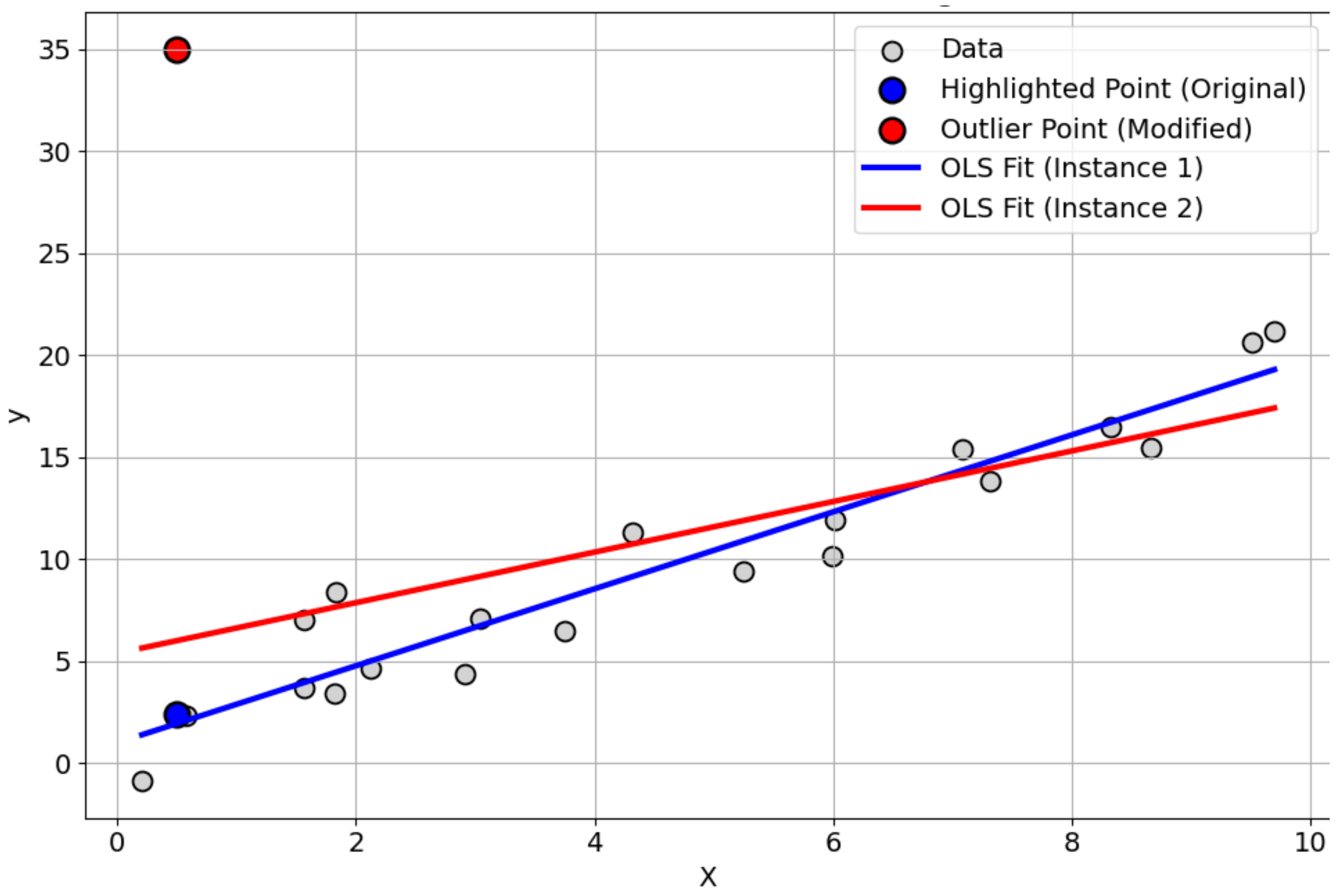
Sample-Optimal Private Regression in Polynomial Time

Prashanti Anderson, Ainesh Bakshi, Mahbod Majid, Stefan Tiegel

ETH zürich · MIT CSAIL

Motivation



Problem Setup

Input: $\{(x_i, y_i)\}_{i\in[n]}$ i.i.d. samples where

$y_i = \langle \theta, x_i\rangle + \zeta_i$ and $x_i \sim \mathcal{N}(0,\Sigma)$, $\zeta_i \sim \mathcal{N}(0,1)$

Output: Privately estimate $\hat{\theta}$ with small *generalization error* (equivalent to parameter recovery as follows: $\|\Sigma^{1/2}(\hat{\theta}-\theta)\| \le \alpha$, which is closeness in some unknown geometry)

- ε-DP: one input point changes → probability of any subset of outputs changes by a multiplicative factor $\le e^{\varepsilon}$ (on worst-case input)

What is the optimal sample complexity for an efficient ε-DP estimator?

Results

Sample-Optimal Pure DP Estimators:

Exists an efficient ε-DP estimator for regression s.t.

- Conditions: $\|\theta\| \le R$ and $\Sigma \preceq L \cdot I_d$
- Error: $\|\Sigma^{1/2}(\hat{\theta}-\theta)\| \le \alpha$
- Sample complexity:

$$\tilde{\Omega}\left(\frac{d^2+\log^2(1/\beta)}{\alpha^2} + \frac{d+\log(1/\beta)}{\alpha\varepsilon} + \frac{d\log(R\sqrt{L})}{\varepsilon}\right)$$

Lower Bounds:

- SQ lower bounds for d^2/α^2 (computational)
- Info-theoretic lower bounds for

$$\frac{d+\log(1/\beta)}{\alpha\varepsilon} + \frac{d\log(\sqrt{L}R)}{\varepsilon}$$

Extensions: Approx DP estimators & DP estimators for mean estimation w/ unknown cov

High Level Approach

- Robustness ⇝ Privacy reduction [HKMN23]
- Output $\hat{\theta} \propto \exp(-\varepsilon \cdot \mathrm{score}(\hat{\theta}))$
- score uses robust Sum-of-Squares estimators

$\mathrm{score}(\hat{\theta}) \approx$ "How many points do I have to change to make $\hat{\theta}$ close to the output of a robust estimator on the new input?"

Challenges

- Existing SoS estimators (with correct rate) need quasi poly sample complexity + runtime
- Explicitly learning the closeness geometry privately is too expensive – $\Omega\left(d^2/(\alpha\varepsilon)\right)$ samples

Technical Innovations

- "One shot" SoS algorithm for robust regression
- Internal representation of the covariance in robust algorithm used as a proxy for the geometry of the space in score function