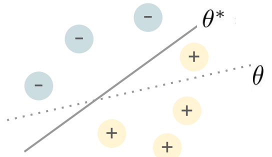
Characterizing Private Clipped Gradient Descent on Convex Generalized Linear Problems

Shuang Song, Om Thakkar, Abhradeep Thakurta

Empirical Risk Minimization (ERM)

Empirical risk: $\mathcal{L}(\theta; D) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\log\left(1+\exp\left(-y_i \cdot \langle \theta, \mathbf{x}_i \rangle\right)\right)}_{\ell(\theta; d_i)}$

$D = \{d_1, \cdots, d_n\}$ where $d_i = (\mathbf{x}_i, y_i)$



Minimizer: $\theta^* = \arg\min_{\theta \in \mathcal{C}} \mathcal{L}(\theta; D)$

ERM: (private) algorithm returns model θ

Excess empirical risk: $R(\theta) = \mathcal{L}(\theta; D) - \mathcal{L}(\theta^*; D)$

Excess Risk for Differentially Private ERM

Matching upper & lower bound [BST'14]

$\sqrt{p}/(\varepsilon \cdot n)$

model parameters
Privacy guarantee
Dataset size

- Constrained
- Gradient doesn't have structure, e.g. lying in low-rank space

Can we get better bound if:

- unconstrained
- gradients lie in (unknown) low-rank subspace

[JT'14]: population risk, objective / output perturbation

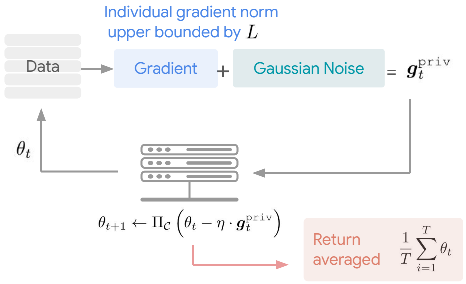
- $\tilde{O}(1/\varepsilon\sqrt{n})$
- Unclear if this is tight
- Bounds for non-convex objectives?

Our Setting: GLMs and Differentially Private Gradient Descent

Generalized Linear Models (GLMs):

- $\ell(\theta; d) = \ell(\langle \theta, \mathbf{x} \rangle; y)$ for $d = (\mathbf{x}, y)$
- Binary logistic regression, SVM, etc.
- Convex GLM: convex loss, convex space

Differentially private gradient descent (DP-GD)



Unconstrained Convex GLMs

Theorem: M being the projector to the eigenspace of matrix $\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^T$

$$\mathbb{E}\left[R(\theta^{priv})\right] \leq \tilde{O}\left(\frac{L\|\theta^*\|_2 \sqrt{rank(M)}}{\varepsilon n}\right)$$

→ vs. $\sqrt{p}$ from previous result
→ Dimension independent
→ rank(M) ≤ min(n, p)

Main technique: for Gaussian noise, L2 norm can be ≫ semi norm wrt. M (depend on rank)

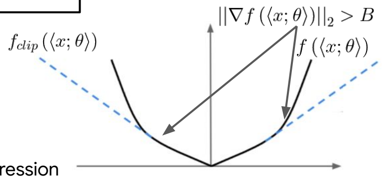
What about Non-convex GLMs?

- Common in robust regression, e.g., Savage loss [MV'09], Tangent loss [MMV'10], Tempered loss [AWAK'19]
- [New, informal]: For smooth losses, DP-GD converges to a first-order stationary point (FOSP).
  - **Dimension-independent** convergence; depends on rank of feature matrix
- Analysis via [Z'18] which shows first-order convergence of GD for non-convex losses
- Conjecture: Our result can be extended to second-order SPs

Clipped Differentially Private Gradient Descent

DP-GD requires knowledge of the Lipschitz constant
- Clipped DP-GD: "Clipping norm" *B* bounds norm of each gradient

- We show Clipping ≈ Huberization [HR'81] for convex GLMs
- [New, informal]: For convex GLMs, clipped DP-GD achieves dimension-independent convergence to minimum of a *well-defined* convex objective
- For functions not convex GLMs, objective may not be *well-defined* for Clipped DP-GD. E.g., multi-class logistic regression



Conclusions

- Dimension-independent excess risk upper bound for convex GLMs
  - Follow-up: Tight lower bounds (credit to Thomas Steinke)
- Dimension-independent convergence to FOSP for non-convex GLMs
- First convergence guarantee for Clipped DP-GD
- [In paper] Adverse effects of aggressive clipping