On the Privacy-Robustness-Utility Trilemma in Distributed Learning

Youssef Allouah Rachid Guerraoui Nirupam Gupta Rafael Pinot John Stephan

EPFL

Summary

1. Lower Bounds: we show that privacy and robustness induce a coupled cost
2. Upper Bounds: we show a matching upper bound using SMEA, our new high-dimensional robust aggregation rule

Distributed Machine Learning

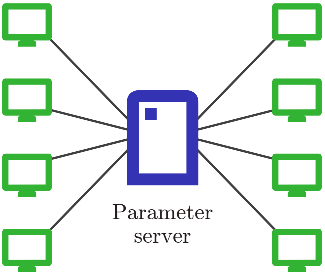
Distributed Machine Learning

Motivations:

- Performance: individual machines lack computing power and data, SOTA models are massive
- Privacy: local data should not be revealed

Distributed Machine Learning

Parameter server architecture: 1 central server, n workers holding m data points each



Desiderata

1. Protection against malicious workers sending corrupt gradients/models
2. Rigorous privacy guarantees for each worker

Goal 1: Byzantine Robustness

- Byzantine workers are omniscient, computationally unbounded and may collude

(f, ϱ)-Byzantine robustness

An algorithm is (f, ϱ)-robust if it can find a $\varrho-$approximate minimizer despite the presence of f Byzantine workers.

- Essentially requires a robust aggregation subroutine (median, trimmed mean, ...) and local variance reduction

Goal 2: Differential Privacy

(ε, δ)-distributed differential privacy

An algorithm satisfies (ε, δ)-distributed DP if the transcript of external communications Z of each worker satisfies (ε, δ)-DP with respect to their local data.

- Essentially requires local noising mechanism (Gaussian, Laplace, ...) with careful variance tuning

Results: Lower Bounds

Results: Lower Bounds

- Each problem is individually hard: BR and DP separately induce lower bounds:

$$\varrho_{\mathrm{BR}} = \Omega\left(\frac{f}{n} G^2\right), \qquad \varrho_{\mathrm{DP}} = \Omega\left(\frac{d}{\varepsilon^2 n m^2}\right),$$

where G measures data heterogeneity and d is the model dimension.

Results: Lower Bounds

- Each problem is individually hard: BR and DP separately induce lower bounds:

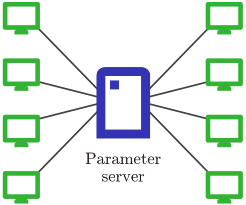
$$\varrho_{\mathrm{BR}} = \Omega\left(\frac{f}{n} G^2\right), \qquad \varrho_{\mathrm{DP}} = \Omega\left(\frac{d}{\varepsilon^2 n m^2}\right),$$

  where G measures data heterogeneity and d is the model dimension.
- Simultaneously achieving both induces a coupled lower bound:

$$\varrho_{\mathrm{DP+BR}} = \tilde{\Omega}\left(\frac{f}{n} \cdot \frac{1}{\varepsilon^2 m^2}\right).$$

Distributed Gradient Descent (D-GD)

- Goal: Exchange gradients to train a single global model θ minimizing loss



At each iteration t:

1. Server broadcasts θ_t
2. Worker computes noisy gradient: $\tilde{g}_t^{(i)} = g_t^{(i)} + \mathcal{N}(0, \sigma_{\mathrm{DP}}^2)$
3. Worker updates local momentum: $\tilde{m}_t^{(i)} = \beta_{t-1} m_t^{(i)} + (1 - \beta_{t-1})\tilde{g}_t^{(i)}$
4. Server aggregates momentums $R_t = F(\tilde{m}_t^{(1)}, \ldots, \tilde{m}_t^{(n)})$
5. Server updates $\theta_{t+1} = \theta_t - \gamma_t R_t$

Results: Matching Upper Bound

- A key ingredient is the following robustness property: for any $x_1, \ldots, x_n \in \mathbb{R}^d$, $\mathcal{H} \subseteq [n], |\mathcal{H}| = n - f$, the aggregation output $\hat{x}$ satisfies

$$\|\hat{x} - \overline{x}_{\mathcal{H}}\|^2 \leq \kappa \cdot \lambda_{\max}\left(\frac{1}{|\mathcal{H}|}\sum_{i\in\mathcal{H}}(x_i - \overline{x}_{\mathcal{H}})(x_i - \overline{x}_{\mathcal{H}})^\top\right),$$

where κ is of order $\frac{f}{n}$ for SMEA.

Results: Matching Upper Bound

- A key ingredient is the following robustness property: for any $x_1, \ldots, x_n \in \mathbb{R}^d$, $\mathcal{H} \subseteq [n], |\mathcal{H}| = n - f$, the aggregation output $\hat{x}$ satisfies

$$\|\hat{x} - \overline{x}_{\mathcal{H}}\|^2 \leq \kappa \cdot \lambda_{\max}\left(\frac{1}{|\mathcal{H}|}\sum_{i \in \mathcal{H}}(x_i - \overline{x}_{\mathcal{H}})(x_i - \overline{x}_{\mathcal{H}})^\top\right),$$

  where κ is of order $\frac{f}{n}$ for SMEA.

- Using the heavy ball method with local Gaussian mechanism and SMEA, we match the lower bound:

$$\varrho = \tilde{O}\left(\frac{d}{\varepsilon^2 n m^2} + \frac{f}{n} \cdot \frac{1}{\varepsilon^2 m^2} + \frac{f}{n} G^2\right).$$

Results: Matching Upper Bound

- A key ingredient is the following robustness property: for any $x_1, \ldots, x_n \in \mathbb{R}^d$, $\mathcal{H} \subseteq [n], |\mathcal{H}| = n - f$, the aggregation output $\hat{x}$ satisfies

$$\|\hat{x} - \overline{x}_{\mathcal{H}}\|^2 \leq \kappa \cdot \lambda_{\max}\left(\frac{1}{|\mathcal{H}|}\sum_{i \in \mathcal{H}}(x_i - \overline{x}_{\mathcal{H}})(x_i - \overline{x}_{\mathcal{H}})^\top\right),$$

  where κ is of order $\frac{f}{n}$ for SMEA.

- Using the heavy ball method with local Gaussian mechanism and SMEA, we match the lower bound:

$$\varrho = \tilde{O}\left(\frac{d}{\varepsilon^2 n m^2} + \frac{f}{n} \cdot \frac{1}{\varepsilon^2 m^2} + \frac{f}{n} G^2\right).$$

- Interestingly in high-dimension $d \geq f$, thanks to SMEA, the coupled cost is dominated by the DP cost