Robust Aggregation for Federated Learning

Krishna Pillutla, Sham Kakade, Zaid Harchaoui

### Robust federated learning

#### Our Setting
- Some clients are adversarial:
  - Data poisoning
  - Model poisoning
- Poor predictions for all clients

#### Objective
$$\min_w \sum_{i \in \theta} F_i(w)$$
- Set of "inlier" clients
- Loss on client $i$

### Goals of the work
A robust aggregation approach that is
1. Communication-efficient
2. Implementable via secure summation

- Server only sees $\xi_i, \xi'_i \sim \text{Unif}(\Theta)$ but calculates the correct sum $x_i + x'_i = x_i + x'_i$

### RFA: Geometric Median Aggregation

#### Weiszfeld’s Algorithm:
$$\beta_{ij} = \frac{1}{\max \{1, \|z_i - w_i\|_2, \nu\}}$$
$$z_{i+1} = \frac{\sum \beta_{ij} w_i}{\sum \beta_{ij}}$$

#### Theory (Least Squares)
Suppose $Y_i = X_i^T w^* + \xi_i$ where $\xi_i \sim \mathcal{N}(0, \sigma^2)$

**Theorem:** Assume that $F(w)$ is strongly convex, $\|X_i\|_1 \leq 1$ and # local steps $\propto 2^t$. Let $\Theta$ denote the event that $\geq 1/2 + c/2$ non-corrupted devices are chosen (out of the $1/2 + \epsilon$ available) in each round.

Then, RFA with $\epsilon$-approximate GM satisfies
$$E[\|w_t - w^*\|^2] \leq \frac{\|w_0 - w^*\|^2}{2^t} + \epsilon^2 \left( \frac{d a^2}{2^t} + \frac{\epsilon^2}{m^2} + \frac{\Omega^2}{2^t} \right)$$

### Handling Heterogeneity
- Personalize parts of the model

#### Improving the Communication Cost

**Single Weiszfeld iteration is also robust!**

#### Software for the geometric median
Install: `pip install geom-median`

Documentation: [github.com/krishnap25/geom_median](https://github.com/krishnap25/geom_median)

### Experiments

- **EMNIST Linear**
- **Test Accuracy** vs **Corruption Level**
- **RFA** vs **FedAvg**
- **RFA more robust!**

Unavoidable due to lower bounds of robust mean estimation

**Code**

- [krishnap25](https://twitter.com/krishnap25)
- [Krishna Pillutla](https://github.com/krishnap25)

---

**References**

- [Bonawitz et al. CCS (2017), Bell et al. CCS (2020)]
- [Robust federated learning](https://doi.org/10.1145/3372297.3379199)
- [Optimization error](https://doi.org/10.1145/3372297.3379199)
- [Some clients are adversarial:](https://doi.org/10.1145/3372297.3379199)
- [Our Setting](https://doi.org/10.1145/3372297.3379199)