Introduction

Is Client-level DP Suitable for Cross-Silo FL?
1. Client-level DP can be an “overkill” for the in-silo data subjects that require privacy protection, particularly when silos have large local datasets.
2. Small # of persistent clients ⇒ hard to achieve strong DP targets compared to cross-device FL.
3. In practice, clients in cross-silo FL may need to publicly disclose their participation (e.g. hospitals).

Silo-specific sample-level DP

• Each silo k sets \((\varepsilon_i, \delta_i)\) sample-level DP for its own dataset.
• Under FL, every silo simply runs DSGD when computing updates, w/ noise calibrated to spend \((\varepsilon_i, \delta_i)\) over training.
• All updates from silo k satisfy \((\varepsilon_i, \delta_i)\)-DP (w.r.t. silo k’s local examples) against all external adversaries (e.g. the server).
• Explored in previous work, this notion is applicable to, e.g.:
  - Voting records
  - Student records
  - Vaccination records
  - Voting centers
  - Schools
  - Exam
  - Election
  - Disease

On Privacy and Personalization in Cross-Silo Federated Learning

Paper Summary

1. We study silo-specific sample-level DP for cross-silo FL.
2. We find that model personalization can play a role in an emerging privacy & data heterogeneity cost tradeoff.
3. We show that mean-regularized multi-task learning (MR-MTL) is a very simple and strong baseline due to three key desiderata: noise reduction, smooth interpolation, and minimal privacy overhead.
4. We theoretically analyze how MR-MTL navigates the privacy-heterogeneity cost tradeoff under federated scalar mean estimation.

Emerging Characteristics

1. Silos incur privacy costs when querying their local data, but not when participating in federated training; in particular,
   - Local training & FedAvg has identical privacy costs.
   - Local fine-tuning may not work as expected (under a standard trust model where the learned models must be private w.r.t. silo’s datasets).

2. Less trust assumptions compared to client-level DP, which necessitates some trust on server for non-local DP (even with distributed DP).
3. Tradeoff emerges between costs from privacy & heterogeneity:
   - Silos’ independent DP noises manifest in model updates and can be mitigated via model averaging (FedAvg), but doing so implies cost from heterogeneity.

MR-MTL & The Privacy-Heterogeneity Cost Tradeoff

**MR-MTL**

- Every silo \(k\) participates in FL, but maintains its own updated model with mean-regularization towards the server mean model:
  \[
  w_k^{(t+1)} = w_k^{(t)} - \eta \left( g_t + \lambda \left( w_k^{(t)} - \bar{w}^{(t)} \right) \right)
  \]
- The regularization param \(\lambda\) gives a (rough) personalization spectrum between local training & FedAvg.
- \(\lambda\) allows MR-MTL to navigate the emerging tradeoff with no privacy overhead. At the optimal \(\lambda^*\), it can outperform both local & FedAvg.

Definition, Instantiation, Examples

- Each silo \(k\) sets \((\varepsilon_i, \delta_i)\) sample-level DP for its own dataset.
- Under FL, every silo runs DSGD when computing updates, w/ noise calibrated to spend \((\varepsilon_i, \delta_i)\) over training.
- All updates from silo \(k\) satisfy \((\varepsilon_i, \delta_i)\)-DP (w.r.t. silo \(k\)’s local examples) against all external adversaries (e.g. the server).
- Explored in previous work, this notion is applicable to, e.g.:
  - Voting records
  - Student records
  - Vaccination records
  - Voting centers
  - Schools
  - Exam
  - Election
  - Disease

Baselines

**Error of MR-MTL under Mean Estimation:**

\[
\hat{\varepsilon}(\lambda) = \left( \frac{1}{1 - \frac{1}{K}} \right) \sigma^2 + \frac{\lambda^2 \sigma^2}{(\lambda + 1)^2} + \frac{\sigma^2_\text{loc}}{K}
\]

The above informs: (1) the existence and value of optimal \(\lambda^*\), (2) the utility “bump” observed on the left, (3) how MR-MTL compares against local & FedAvg, and (4) how \(\lambda\) interfaces with DP noises and data heterogeneity.

Broader open question: the privacy cost of tuning \(\lambda\) may already outweigh the benefits of MR-MTL.

Theoretical Characterization

MR-MTL is a strong baseline against many SoA methods (which may incur privacy overhead from extra iterations, private selection, etc.) under silo-specific sample-level DP.