Differentially Private Federated Learning on Heterogeneous Data

Utility & Privacy tradeoffs

Maxence Noble [1]

Joint work with Aymeric Dieuleveut [1] and Aurélien Bellet [2]

[1]CMAP, École Polytechnique, France; [2]MAGNET Team, INRIA, France

Table of contents

On Federated Learning and Privacy

Issues at stake

Centralized Federated Learning

Issues at stake

Centralized Federated Learning

- users collaboratively train one ML model via one server.

Issues at stake

Centralized Federated Learning

- users collaboratively train one ML model via one server.
- each user's dataset is kept private and decentralized.

Issues at stake

Centralized Federated Learning

- users collaboratively train one ML model via one server.
- each user's dataset is kept private and decentralized.
- the simplest `SGD` baseline: `FedAvg` ([5], 2017).

Centralized Federated Learning

- users collaboratively train one ML model via one server.
- each user's dataset is kept private and decentralized.
- the simplest `SGD` baseline: `FedAvg` ([5], 2017).

1. **Facing the challenge of heterogeneity between users**

Centralized Federated Learning

- users collaboratively train one ML model via one server.
- each user's dataset is kept private and decentralized.
- the simplest `SGD` baseline: `FedAvg` ([5], 2017).

1. Facing the challenge of heterogeneity between users

- non i.i.d data and/or divergence in local "true" models.

Centralized Federated Learning

- users collaboratively train one ML model via one server.
- each user's dataset is kept private and decentralized.
- the simplest `SGD` baseline: `FedAvg` ([5], 2017).

1. **Facing the challenge of heterogeneity between users**

- non i.i.d data and/or divergence in local "true" models.
- `SCAFFOLD` ([4], 2020): **use of control variates** to correct the direction of local gradients.

Centralized Federated Learning

- users collaboratively train one ML model via one server.
- each user's dataset is kept private and decentralized.
- the simplest `SGD` baseline: `FedAvg` ([5], 2017).

1. **Facing the challenge of heterogeneity between users**
   - non i.i.d data and/or divergence in local "true" models.
   - `SCAFFOLD` ([4], 2020): **use of control variates** to correct the direction of local gradients.
2. **Facing the challenge of privacy**

Centralized Federated Learning

- users collaboratively train one ML model via one server.
- each user's dataset is kept private and decentralized.
- the simplest SGD baseline: `FedAvg` ([5], 2017).

1. **Facing the challenge of heterogeneity between users**
   - non i.i.d data and/or divergence in local "true" models.
   - `SCAFFOLD` ([4], 2020): **use of control variates** to correct the direction of local gradients.
2. **Facing the challenge of privacy**
   - towards the server or a third party.

Centralized Federated Learning

- users collaboratively train one ML model via one server.
- each user's dataset is kept private and decentralized.
- the simplest `SGD` baseline: `FedAvg` ([5], 2017).

1. **Facing the challenge of heterogeneity between users**
   - non i.i.d data and/or divergence in local “true“ models.
   - `SCAFFOLD` ([4], 2020): **use of control variates** to correct the direction of local gradients.
2. **Facing the challenge of privacy**
   - towards the server or a third party.
   - Differential Privacy (DP) [2]: statistical approach to hide *individual* contributions to the dataset by adding noise to gradients.

Centralized Federated Learning

- users collaboratively train one ML model via one server.
- each user's dataset is kept private and decentralized.
- the simplest SGD baseline: `FedAvg` ([5], 2017).

1. **Facing the challenge of heterogeneity between users**

- non i.i.d data and/or divergence in local "true" models.
- `SCAFFOLD` ([4], 2020): **use of control variates** to correct the direction of local gradients.

2. **Facing the challenge of privacy**

- towards the server or a third party.
- Differential Privacy (DP) [2]: statistical approach to hide *individual* contributions to the dataset by adding noise to gradients.
- DP level is ensured by a budget $(\epsilon, \delta) \in {\mathbb{R}_+^*}^2$ (the lower, the better).

Related work

Contributions

Related work

- versions of DP-FedAvg [6, 3, 8], mostly without theoretical analysis.
- no approach designed to tackle data heterogeneity with DP.

Contributions

Related work

- versions of DP-FedAvg [6, 3, 8], mostly without theoretical analysis.
- no approach designed to tackle data heterogeneity with DP.

Our results: expressing utility and privacy guarantees for federated learning with fine results of DP theory [7], considering

Contributions

Related work

- versions of DP-FedAvg [6, 3, 8], mostly without theoretical analysis.
- no approach designed to tackle data heterogeneity with DP.

Our results: expressing utility and privacy guarantees for federated learning with fine results of DP theory [7], considering

- heterogeneity issues between users,

Contributions

Related work

- versions of DP-FedAvg [6, 3, 8], mostly without theoretical analysis.
- no approach designed to tackle data heterogeneity with DP.

Our results: expressing utility and privacy guarantees for federated learning with fine results of DP theory [7], considering

- heterogeneity issues between users,
- convex and non-convex objective functions.

Contributions

Related work

- versions of DP-FedAvg [6, 3, 8], mostly without theoretical analysis.
- no approach designed to tackle data heterogeneity with DP.

Our results: expressing utility and privacy guarantees for federated learning with fine results of DP theory [7], considering

- heterogeneity issues between users,
- convex and non-convex objective functions.

Our algorithm: DP-SCAFFOLD(-warm)

Related work

- versions of `DP-FedAvg` [6, 3, 8], mostly without theoretical analysis.
- no approach designed to tackle data heterogeneity with DP.

Our results: expressing utility and privacy guarantees for federated learning with fine results of DP theory [7], considering

- heterogeneity issues between users,
- convex and non-convex objective functions.

Our algorithm: `DP-SCAFFOLD(-warm)`

- using gradient perturbation via Gaussian noise with scale σ_g,

Contributions

Related work

- versions of DP-FedAvg [6, 3, 8], mostly without theoretical analysis.
- no approach designed to tackle data heterogeneity with DP.

Our results: expressing utility and privacy guarantees for federated learning with fine results of DP theory [7], considering

- heterogeneity issues between users,
- convex and non-convex objective functions.

Our algorithm: DP-SCAFFOLD(-warm)

- using gradient perturbation via Gaussian noise with scale σ_g,
- taking advantage of control variates,

Contributions

Related work

- versions of DP-FedAvg [6, 3, 8], mostly without theoretical analysis.
- no approach designed to tackle data heterogeneity with DP.

Our results: expressing utility and privacy guarantees for federated learning with fine results of DP theory [7], considering

- heterogeneity issues between users,
- convex and non-convex objective functions.

Our algorithm: `DP-SCAFFOLD(-warm)`

- using gradient perturbation via Gaussian noise with scale σ_g,
- taking advantage of control variates,
- fairly comparing our results to DP-FedAvg(σ_g) performance.

Theoretical results

Theoretical results

We provide two main results in our article.

Theoretical results

We provide two main results in our article.

Privacy analysis

Theoretical results

We provide two main results in our article.

Privacy analysis

- with the same scale of noise σ_g^*, DP-SCAFFOLD and DP-FedAvg are both $(O(\epsilon),\delta)$-DP w.r.t. the whole dataset towards any third party.

Theoretical results

We provide two main results in our article.

Privacy analysis

- with the same scale of noise σ_g^*, DP-SCAFFOLD and DP-FedAvg are both $(O(\epsilon),\delta)$-DP w.r.t. the whole dataset towards any third party.

Utility analysis

Theoretical results

We provide two main results in our article.

Privacy analysis

- with the same scale of noise σ_g^*, DP-SCAFFOLD and DP-FedAvg are both $(O(\epsilon), \delta)$-DP w.r.t. the whole dataset towards any third party.

Utility analysis

- if $\sigma_g = \sigma_g^*$, DP-SCAFFOLD-warm converges faster than DP-FedAvg,

Theoretical results

We provide two main results in our article.

Privacy analysis

- with the same scale of noise σ_g^*, DP-SCAFFOLD and DP-FedAvg are both $(O(\epsilon), \delta)$-DP w.r.t. the whole dataset towards any third party.

Utility analysis

- if $\sigma_g = \sigma_g^*$, DP-SCAFFOLD-warm converges faster than DP-FedAvg,
- the proof for DP-FedAvg relies on an extra assumption on gradients.

Theoretical results

We provide two main results in our article.

Privacy analysis

- with the same scale of noise σ_g^*, `DP-SCAFFOLD` and `DP-FedAvg` are both $(O(\epsilon), \delta)$-DP w.r.t. the whole dataset towards any third party.

Utility analysis

- if $\sigma_g = \sigma_g^*$, `DP-SCAFFOLD-warm` converges faster than `DP-FedAvg`,
- the proof for `DP-FedAvg` relies on an extra assumption on gradients.

We highlight theoretical trade-offs in our convergence bounds involving

- terms of **heterogeneity**,
- terms of **privacy** (ϵ, δ),
- terms from the **federated** framework (number of users, number of communication rounds, number of local `SGD` updates, sampling...).

Numerical experiments

Experiments (heterogeneity increasing from left to right)

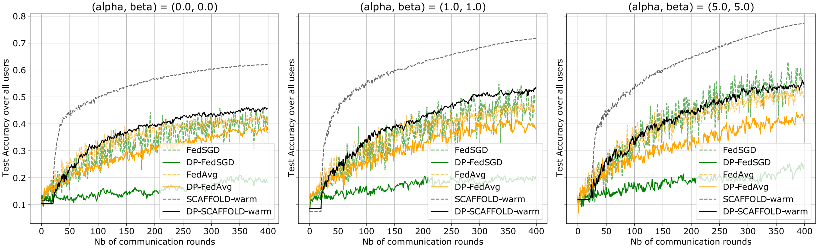

Figure 1: Test Accuracy on synthetic data (Logistic regression)

Experiments (heterogeneity increasing from left to right)

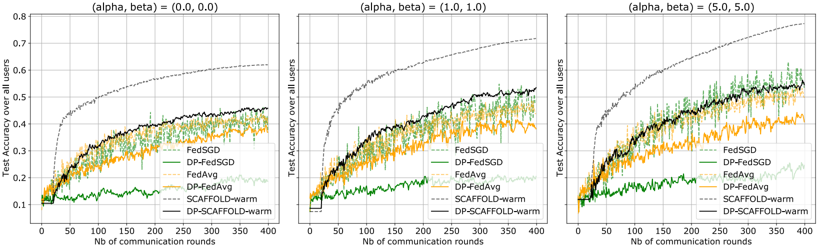

Figure 1: Test Accuracy on synthetic data (Logistic regression)

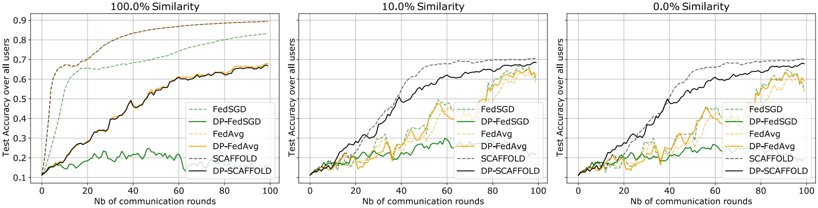

Figure 2: Test Accuracy on MNIST [1] data (Neural network, one hidden layer)

Differentially Private Federated Learning on Heterogeneous Data

Utility & Privacy tradeoffs

Maxence Noble [1]

Joint work with Aymeric Dieuleveut [1] and Aurélien Bellet [2]

[1]CMAP, École Polytechnique, France; [2]MAGNET Team, INRIA, France

References

[1] Gregory Cohen, Saeed Afshar, Jonathan Tapson, and Andre Van Schaik. Emnist: Extending mnist to handwritten letters. In *2017 international joint conference on neural networks (IJCNN)*, pages 2921–2926. IEEE, 2017.

[2] Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3-4):211–407, 2014.

[3] Robin C. Geyer, Tassilo Klein, and Moin Nabi. Differentially private federated learning: A client level perspective. *arXiv preprint arXiv:1712.07557*, 2017.

[4] Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank Reddi, Sebastian Stich, and Ananda Theertha Suresh. Scaffold: Stochastic controlled averaging for federated learning. In *International Conference on Machine Learning*, pages 5132–5143. PMLR, 2020.

[5] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Artificial intelligence and statistics*, pages 1273–1282. PMLR, 2017.

[6] H. Brendan McMahan, Daniel Ramage, Kunal Talwar, and Li Zhang. Learning differentially private recurrent language models. In *International Conference on Learning Representations*, 2018.

[7] Ilya Mironov. Rényi differential privacy. In *2017 IEEE 30th computer security foundations symposium (CSF)*, pages 263–275. IEEE, 2017.

[8] Aleksei Triastcyn and Boi Faltings. Federated learning with bayesian differential privacy. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 2587–2596. IEEE, 2019.