DIFFERENTIALLY PRIVATE MACHINE LEARNING

Aurélien Bellet (Inria)

Journées de Statistique (JDS)
June 13, 2022

- Massive collection of personal data by companies and public organizations, driven by the progress of data science and AI

- Data is increasingly sensitive and detailed: browsing history, purchase history, social network posts, speech, geolocation, health...

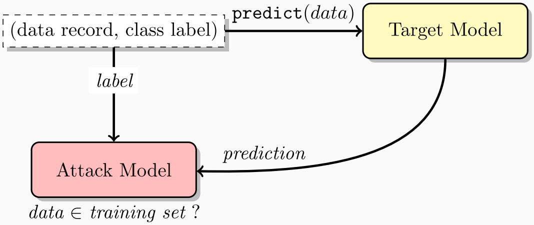
- **Quantifying privacy risks is challenging!**
  - Attacker may have prior knowledge
  - Same data used in multiple computations
  - Indirect leakage from aggregate quantities

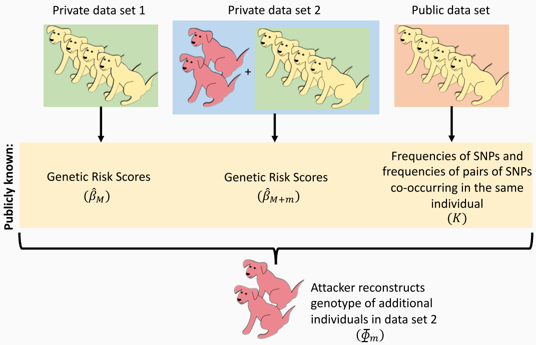
AGGREGATE STATISTICS ARE NOT SAFE

- Aggregate (potentially noisy) statistics about many individuals are vulnerable to various attacks on data privacy
- Membership inference attacks, i.e. inferring presence of known individual in a dataset from (high-dimensional) aggregate statistics
  - Example: statistics about genomic variants [Homer et al., 2008]
- Reconstruction attacks, i.e. inferring (part of) the dataset from the output of many aggregate statistics
  - After sufficiently many queries, one can reconstruct the dataset [Dinur and Nissim, 2003]

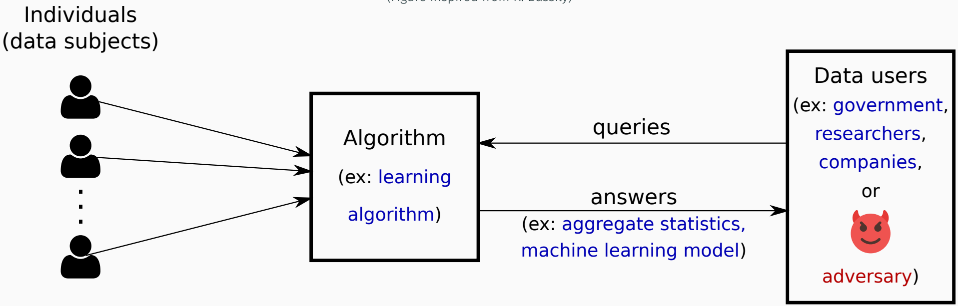
MACHINE LEARNING MODELS ARE NOT SAFE

- Machine learning models are elaborate kinds of aggregate statistics
- They are also susceptible to membership inference and reconstruction attacks, see e.g. [Shokri et al., 2017, Paige et al., 2020, Geiping et al., 2020]





(Figure inspired from R. Bassily)

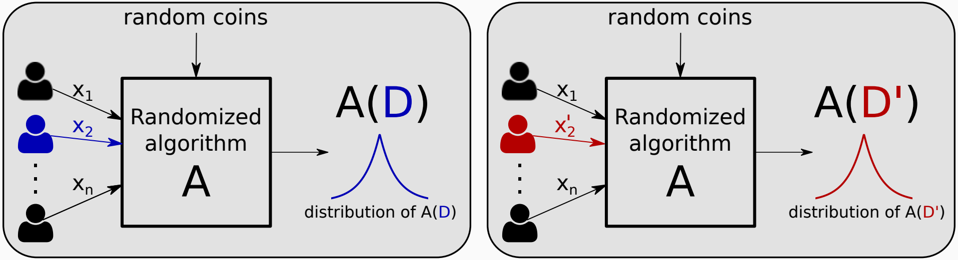


- **Goal:** achieve utility while preserving privacy (conflicting objectives!)
- Note: this is separate from security concerns (e.g., unauthorized access to the system)

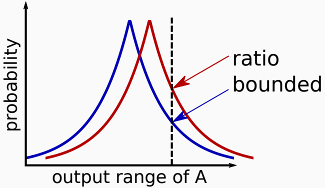
OUTLINE

1. Differential Privacy
2. Private learning in the centralized setting
3. Private learning without a trusted curator

Differential Privacy



- Neighboring datasets $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ and $\mathcal{D}' = \{x_1, x'_2, x_3, \ldots, x_n\}$
- **Requirement**: $\mathcal{A}(\mathcal{D})$ and $\mathcal{A}(\mathcal{D}')$ should have "close" distribution



Definition ([Dwork et al., 2006], informal)

A randomized algorithm $\mathcal{A}$ is (ε, δ)-differentially private (DP) if for all neighboring datasets $\mathcal{D} = \{x_1, x_2, \dots, x_n\}$ and $\mathcal{D}' = \{x_1, x_2', x_3, \dots, x_n\}$ and all sets S:

$$\Pr[\mathcal{A}(\mathcal{D}) \in S] \leq e^{\varepsilon} \Pr[\mathcal{A}(\mathcal{D}') \in S] + \delta.$$

- DP is a property of the analysis, not of a particular output
- Sufficient condition: for $o \sim \mathcal{A}(D)$, the privacy loss $\left| \ln \left(\frac{\Pr[\mathcal{A}(D)=o]}{\Pr[\mathcal{A}(D')=o]} \right) \right|$ is bounded by ϵ with probability $1 - \delta$ (note: ϵ can be seen as a function of δ)
- For meaningful privacy guarantees, think of $\varepsilon \leq 1$ and $\delta \ll 1/n$
- In 2017, Dwork, McSherry, Nissim & Smith won the Gödel prize for introducing DP
- In 2020, the US Census started to use DP for its data releases

PROPERTIES OF DP: ROBUSTNESS TO POSTPROCESSING AND AUXILIARY KNOWLEDGE

- Robustness to processing: informally, if $\mathcal{A}$ is (ϵ, δ)-DP, then so is $f \circ \mathcal{A}$ for any f
- Robustness to auxiliary knowledge: DP bounds the relative advantage that an adversary gets from observing the output of an algorithm
    - DP holds even if adversary knows all but one data record
    - Interpretation as hypothesis testing: adversary knows $\mathcal{A}$ and neighboring datasets $\mathcal{D}_0$ and $\mathcal{D}_1$, observes a realization of $\mathcal{A}(\mathcal{D}_b)$ for a secret bit $b \in \{0, 1\}$, and must guess whether it was drawn from $\mathcal{A}(\mathcal{D}_0)$ or $\mathcal{A}(\mathcal{D}_1)$
    - DP puts a bound on the trade-offs between the true positive rate and the false positive rate that can be achieved for this test

- Composition allows to control the *worst-case* cumulative privacy loss over multiple analyses run on the same dataset, including complex multi-step algorithms

Theorem (Simple composition)

Let $\mathcal{A}_1, \dots, \mathcal{A}_K$ be such that $\mathcal{A}_k$ satisfies (ϵ_k, δ_k)-DP. For any dataset $\mathcal{D}$, let $\mathcal{A}$ be such that $\mathcal{A}(\mathcal{D}) = (\mathcal{A}_1(\mathcal{D}), \dots, \mathcal{A}_k(\mathcal{D}))$. Then $\mathcal{A}$ is (ϵ, δ)-DP with $\epsilon = \sum_{k=1}^K \epsilon_k$ and $\delta = \sum_{k=1}^K \delta_k$.

Theorem (Advanced composition)

Let $\epsilon, \delta, \delta' > 0$. If $\mathcal{A}_k$ satisfies (ϵ, δ)-DP, then $\mathcal{A}$ is $(\epsilon', K\delta + \delta')$-DP with

$$\epsilon' = \sqrt{2K \ln(1/\delta')}\epsilon + K\epsilon(e^\epsilon - 1)$$

- The sequence of algorithms can be chosen adaptively
- Numerically tighter composition can be obtained with through a variant of DP based on the Rényi divergence [Mironov, 2017]

ENFORCING DP WITH THE GAUSSIAN MECHANISM

- Consider f taking as input a dataset and returning a p-dimensional real vector

Gaussian mechanism $\mathcal{A}_{\text{Gauss}}(\mathcal{D}, f, \varepsilon, \delta)$

1. Compute sensitivity $\Delta = \max_{(\mathcal{D},\mathcal{D}') \text{ are neighboring}} \|f(\mathcal{D}) - f(\mathcal{D}')\|_2$
2. Output $f(\mathcal{D}) + \eta$, where $\eta \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_p)$ with $\sigma = \frac{\sqrt{2\ln(1.25/\delta)}\Delta}{\varepsilon}$

Theorem

Let $\varepsilon, \delta > 0$. The Gaussian mechanism $\mathcal{A}_{\text{Gauss}}(\cdot, f, \varepsilon, \delta)$ is (ε, δ)-DP.

- Noise calibrated using sensitivity of f and privacy budget (ε and δ)
- Sketch of proof: tail bound for the Gaussian distribution + simplifications
- DP induces a privacy-utility trade-off, here in terms of the variance of the estimate
- Note: the MSE achieved by the Gaussian mechanism is worst-case optimal

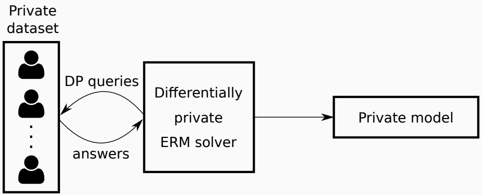
Private learning in the centralized setting

- A trusted curator wants to privately release a model trained on data $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^n$
- We focus here on approximately solving an Empirical Risk Minimization (ERM) problem under an (ϵ, δ)-DP constraint:

$$\min_{\theta \in \Theta} \left\{ F(\theta; \mathcal{D}) := \frac{1}{n} \sum_{i=1}^{n} L(\theta; x_i, y_i) \right\}$$

(Note: in some cases, DP can imply generalization [Bassily et al., 2016, Jung et al., 2021])

- We can achieve this by designing a differentially private ERM solver



Algorithm: Differentially Private SGD $\mathcal{A}_{\text{DP-SGD}}(\mathcal{D}, L, \epsilon, \delta)$

- Initialize parameters to $\theta^{(0)} \in \Theta$ (must be independent of $\mathcal{D}$)
- For $t = 0, \ldots, T-1$:
  - Pick random mini-batch $\mathcal{B}^{(t)} \subseteq \{1, \ldots, n\}$ of size m
  - $\eta^{(t)} \leftarrow (\eta_1^{(t)}, \ldots, \eta_p^{(t)}) \in \mathbb{R}^p$ where each $\eta_j^{(t)} \sim \mathcal{N}(0, \sigma^2)$ with $\sigma = \frac{16 l \sqrt{T \ln(2/\delta) \ln(1.25T/\delta n)}}{n\epsilon}$
  - $\theta^{(t+1)} \leftarrow \Pi_\Theta\left(\theta^{(t)} - \gamma_t(\nabla L(\theta^{(t)}; \mathcal{B}^{(t)}) + \eta^{(t)})\right)$ (Π_Θ projection operator)
- Return $\theta^{(T)}$

- More data (larger n) $\rightarrow$ less noise added to each gradient
- More iterations (larger T) $\rightarrow$ more noise added to each gradient

Theorem (DP guarantees for DP-SGD)

Let $\epsilon \leq 1, \delta > 0$. Let the loss function $L(\cdot; x, y)$ be l-Lipschitz w.r.t. the ℓ_2 norm for all $x, y \in \mathcal{X} \times \mathcal{Y}$. Then $\mathcal{A}_{\text{DP-SGD}}(\cdot, L, \epsilon, \delta)$ is (ϵ, δ)-DP.

Sketch of proof.

- Recall that for a query with ℓ_2 sensitivity Δ, achieving (ϵ', δ') with the Gaussian mechanism requires to add noise with standard deviation $\sigma' = \frac{\sqrt{2\ln(1.25/\delta')}\Delta}{\epsilon'}$
- The loss function L is l-Lipschitz, which implies that ℓ_2-norm of individual gradients is bounded by l and therefore $\Delta = 2l/m$
- Hence, with $\sigma = \frac{16l\sqrt{T\ln(2/\delta)\ln(1.25T/\delta n)}}{n\epsilon}$, each noisy gradient is $\left(\frac{n\epsilon}{4m\sqrt{2T\ln(2/\delta)}}, \frac{\delta n}{2mT}\right)$-DP
- Using privacy amplification by subsampling [Balle et al., 2018] allows to leverage the randomness in the choice of $\mathcal{B}$: each noisy gradient is in fact $\left(\frac{\epsilon}{2\sqrt{2T\ln(2/\delta)}}, \frac{\delta}{2T}\right)$-DP
- DP-SGD is an adaptive composition of T DP mechanisms, so by advanced composition we obtain that it is (ϵ, δ)-DP

□

Theorem (Utility guarantees for DP-SGD [Bassily et al., 2014])

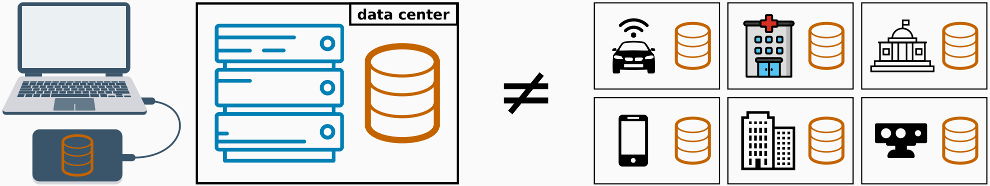
Let Θ be a convex domain of diameter bounded by R, and let the loss function L be convex and l-Lipschitz over Θ. For $T = n^2$ and $\gamma_t = O(R/\sqrt{t})$, DP-SGD guarantees:

$$\mathbb{E}[F(\theta^{(T)})] - \min_{\theta \in \Theta} F(\theta) \leq O\left(\frac{lR\sqrt{p \ln(1/\delta)} \ln^{3/2}(n/\delta)}{n\epsilon}\right).$$

- Proof: plug variance of stochastic gradients in analysis of SGD [Shamir and Zhang, 2013]
- Utility gap w.r.t. the non-private model is $\widetilde{O}(\sqrt{p}/\epsilon n)$, which is worst-case optimal
- In practice: drop Lipschitz assumption and use gradient clipping [Abadi et al., 2016], which introduces a bias-variance trade-off in gradient estimation

Private learning without a trusted curator

- In the real world data is often decentralized across different parties



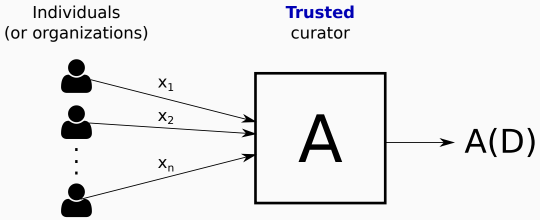
- Data may be considered too sensitive to be shared (e.g., due to legal restrictions, intellectual property rights, or because it provides a competitive advantage)
- Inferior performance and/or biased results if each party learns independently

Federated Learning (FL) aims to
collaboratively train ML models
while keeping the data decentralized

- FL is a booming and multidisciplinary topic: see collaborative survey [Kairouz et al., 2021] to know more about existing work and open problems
- FL does not itself provide any privacy guarantees: in fact, it offers an additional attack surface compared to the centralized setting as participants will observe some intermediate results [Nasr et al., 2019, Geiping et al., 2020]

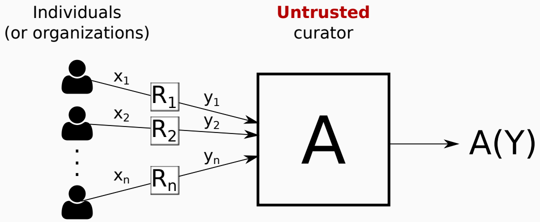
Central DP: a trusted curator collects raw data and runs a DP algorithm on it

→ the observed output is only the final result



Local DP: no trusted curator so each party must locally run a DP algorithm

→ the observed output consists of all messages shared by all parties



A KEY FUNCTIONALITY FOR FL: DP AGGREGATION

- Consider K parties, with each party k holding local dataset $\mathcal{D}_k$
- Many FL algorithms rely on a coordinating server and proceed as follows:

  for $t = 1$ to T **do**
  At each party k: compute $\theta_k \leftarrow \text{LocalUpdate}(\theta, \theta_k; \mathcal{D}_k)$, send θ_k to server
  At server: compute $\theta \leftarrow \frac{1}{K}\sum_k \theta_k$, send θ back to the parties

- Therefore: DP aggregation $+$ Composition property of DP $\implies$ DP-FL
- **DP aggregation:** given a private value $\theta_k \in [0, 1]$ for each party k, we want to accurately estimate $\theta^{avg} = \frac{1}{K}\sum_k \theta_k$ under a DP constraint
- Central DP: trusted server computes θ^{avg} and adds Gaussian noise
- Local DP: each party k adds Gaussian noise to θ_k before sharing it

Error is $\sqrt{K}$ larger in local DP $\rightarrow$ study intermediate trust models

- Assume that pairs of parties can communicate through secure channels (the server may serve as relay), e.g. using a public key infrastructure

Algorithm GOPA protocol [Sabater et al., 2020]

Each party k generates independent Gaussian noise η_k
Each party k selects a random set of m other parties
for all selected pairs of parties $k \sim l$ **do**
  Parties k and l securely exchange pairwise-canceling Gaussian noise $\Delta_{k,l} = -\Delta_{l,k}$
Each party k sends $\hat{\theta}_k = \theta_k + \sum_{k \sim l} \Delta_{k,l} + \eta_k$ to the server

- **Estimate of the average:** $\hat{\theta}^{avg} = \frac{1}{K} \sum_k \hat{\theta}_k = \theta^{avg} + \frac{1}{K} \sum_k \eta_k$
- Intuition: pairwise noise does not affect utility but helps protecting individual values

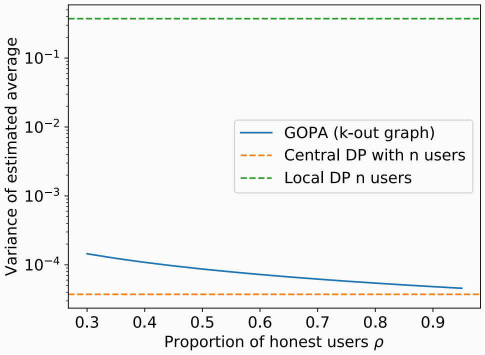
PRIVACY GUARANTEES FOR GOPA

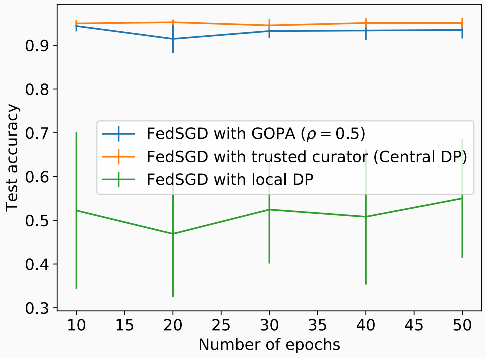
- **Adversary**: coalition of the server with a proportion $1 - \tau$ of the parties

Theorem (Privacy of GOPA [Sabater et al., 2020], informal)

- *Let each party select $m = O(\log(\tau K)/\tau)$ other parties*
- *Set the independent noise variance so as to satisfy (ϵ, δ')-DP in the central model*
- *For large enough pairwise noise variance, GOPA is (ϵ, δ)-DP with $\delta = O(\delta')$.*

- Same utility as central DP with only logarithmic number of messages per party
- Our theoretical results give practical values for the quantities above
- More generally, we precisely quantify the effect of the graph of communications between honest parties on the privacy guarantees

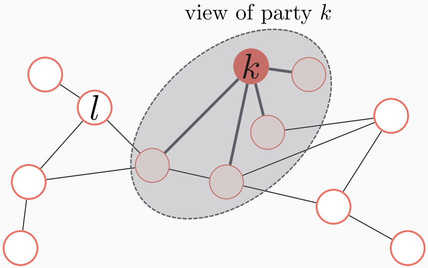




- For reasonable proportions ρ of honest parties, the variance of the estimated average produced by GOPA is similar to the trusted curator setting
- As expected, the resulting FL model also has similar accuracy

PRIVACY & FULL DECENTRALIZATION

- In fully decentralized FL, global aggregations are replaced by local aggregations among neighbors in a graph (thus, the previous approach cannot be applied)



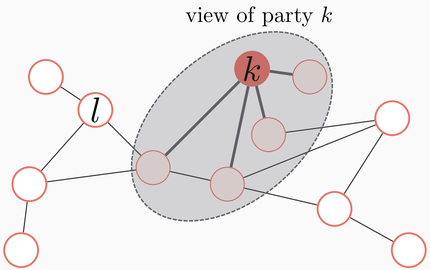
- But there is no server observing all messages, and each party k has a limited view
- Can this be used to prove stronger differential privacy guarantees?

- Let $\mathcal{O}_k$ be the set of messages sent and received by party k

Definition (Network DP [Cyffers and Bellet, 2022])

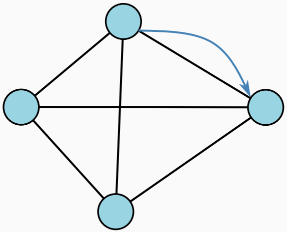
An algorithm $\mathcal{A}$ satisfies (ϵ, δ)-network DP if for all pairs of distinct parties $k, l \in \{1, \ldots, K\}$ and all pairs of datasets $\mathcal{D}, \mathcal{D}'$ that differ only in the local dataset of party l, we have:

$$\Pr[\mathcal{O}_k(\mathcal{A}(\mathcal{D}))] \leq e^{\epsilon} \Pr[\mathcal{O}_k(\mathcal{A}(\mathcal{D}'))] + \delta.$$



- This is a relaxation of local DP: if $\mathcal{O}_k$ contains the full transcript of messages, then network DP boils down to local DP

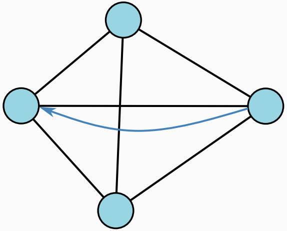
- Consider the standard objective $F(\theta; \mathcal{D}) = \frac{1}{K}\sum_{k=1}^{K} F_k(\theta; \mathcal{D}_k)$ and a complete graph
- We consider a fully decentralized algorithm where the model is updated sequentially by following a random walk



Algorithm Private decentralized SGD on a complete graph

Initialize model θ
for $t = 1$ to T **do**
  Current party updates θ by a gradient update with Gaussian noise
  Current party sends θ to a random party
return θ

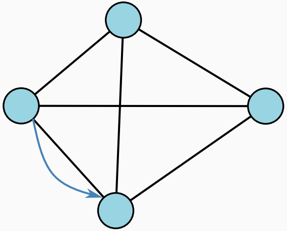
- Consider the standard objective $F(\theta; \mathcal{D}) = \frac{1}{K}\sum_{k=1}^{K} F_k(\theta; \mathcal{D}_k)$ and a complete graph
- We consider a fully decentralized algorithm where the model is updated sequentially by following a random walk



Algorithm Private decentralized SGD on a complete graph

Initialize model θ
for $t = 1$ to T **do**
  Current party updates θ by a gradient update with Gaussian noise
  Current party sends θ to a random party
return θ

- Consider the standard objective $F(\theta; \mathcal{D}) = \frac{1}{K} \sum_{k=1}^{K} F_k(\theta; \mathcal{D}_k)$ and a complete graph
- We consider a fully decentralized algorithm where the model is updated sequentially by following a random walk



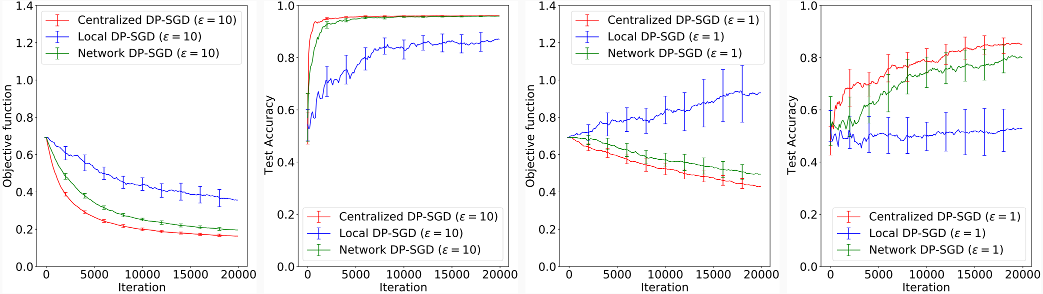
Algorithm Private decentralized SGD on a complete graph

Initialize model θ
for $t = 1$ to T **do**
  Current party updates θ by a gradient update with Gaussian noise
  Current party sends θ to a random party
return θ

Theorem ([Cyffers and Bellet, 2022], informal)

To achieve a fixed (ϵ, δ)-DP guarantee with the previous algorithm, the standard deviation of the noise is $O(\sqrt{K}/\ln K)$ smaller under network DP than under local DP.

- Accounting for the limited view in fully decentralized algorithms amplifies privacy guarantees by a factor of $O(\ln K/\sqrt{K})$, nearly recovering the utility of central DP
- The proof leverages recent results on privacy amplification by iteration [Feldman et al., 2018] and exploits the randomness of the path taken by the model
- We show some robustness to collusion (albeit with smaller privacy amplification)



- Results are consistent with our theory: network DP-SGD significantly amplifies privacy guarantees compared to local DP-SGD

Wrapping up

SUMMARY

- Differential privacy provides a robust mathematical definition of privacy and a strong algorithmic framework allowing to design complex private algorithms
- DP induces a privacy-utility trade-off which depends on the trust model: the two extreme cases are the central (trusted curator) model and the local model (trust no one and nothing except oneself)
- In the context of federated learning, we can leverage appropriate relaxations of local DP to nearly match the privacy-utility trade-off of the central model

SOME OPEN PROBLEMS IN PRIVACY & ML

- **Going beyond worst-case privacy-utility trade-offs:** leverage the structure of some machine learning problems to design better DP algorithms
- **Better privacy accounting:** tight, automatic and personalized
- **Correctness guarantees under malicious parties:** make computation verifiable while preserving privacy guarantees
- **Combining DP with secure multi-party computation:** identify tractable secure primitives under which one can achieve trusted curator utility for many problems
- **Concrete DP/FL deployments:** match DP bounds to protection against specific attacks, articulate with the law (GDPR), make FL transparent to end-users

THANK YOU FOR YOUR ATTENTION!

QUESTIONS?

REFERENCES I

[Abadi et al., 2016] Abadi, M., Chu, A., Goodfellow, I. J., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. (2016).
Deep learning with differential privacy.
In *CCS*.

[Balle et al., 2018] Balle, B., Barthe, G., and Gaboardi, M. (2018).
Privacy amplification by subsampling: tight analyses via couplings and divergences.
In *NeurIPS*.

[Bassily et al., 2016] Bassily, R., Nissim, K., Smith, A., Steinke, T., Stemmer, U., and Ullman, J. (2016).
Algorithmic stability for adaptive data analysis.
In *STOC*.

[Bassily et al., 2014] Bassily, R., Smith, A. D., and Thakurta, A. (2014).
Private Empirical Risk Minimization: Efficient Algorithms and Tight Error Bounds.
In *FOCS*.

[Cyffers and Bellet, 2022] Cyffers, E. and Bellet, A. (2022).
Privacy Amplification by Decentralization.
In *AISTATS*.

[Dinur and Nissim, 2003] Dinur, I. and Nissim, K. (2003).
Revealing information while preserving privacy.
In *PODS*.

[Dwork et al., 2006] Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006).
Calibrating noise to sensitivity in private data analysis.
In *Theory of Cryptography (TCC)*.

[Feldman et al., 2018] Feldman, V., Mironov, I., Talwar, K., and Thakurta, A. (2018).
Privacy Amplification by Iteration.
In *FOCS*.

[Geiping et al., 2020] Geiping, J., Bauermeister, H., Dröge, H., and Moeller, M. (2020).
Inverting gradients - how easy is it to break privacy in federated learning?
In *NeurIPS*.

[Homer et al., 2008] Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J. V., Stephan, D. A., Nelson, S. F., and Craig, D. W. (2008).
Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays.
PLOS Genetics, 4(8):1–9.

[Jung et al., 2021] Jung, C., Ligett, K., Neel, S., Roth, A., Sharifi-Malvajerdi, S., and Shenfeld, M. (2021).
A New Analysis of Differential Privacy's Generalization Guarantees (Invited Paper).

[Kairouz et al., 2021] Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawitz, K., Charles, Z., Cormode, G., Cummings, R., D'Oliveira, R. G. L., Eichner, H., Rouayheb, S. E., Evans, D., Gardner, J., Garrett, Z., Gascón, A., Ghazi, B., Gibbons, P. B., Gruteser, M., Harchaoui, Z., He, C., He, L., Huo, Z., Hutchinson, B., Hsu, J., Jaggi, M., Javidi, T., Joshi, G., Khodak, M., Konecný, J., Korolova, A., Koushanfar, F., Koyejo, S., Lepoint, T., Liu, Y., Mittal, P., Mohri, M., Nock, R., Özgür, A., Pagh, R., Qi, H., Ramage, D., Raskar, R., Raykova, M., Song, D., Song, W., Stich, S. U., Sun, Z., Suresh, A. T., Tramèr, F., Vepakomma, P., Wang, J., Xiong, L., Xu, Z., Yang, Q., Yu, F. X., Yu, H., and Zhao, S. (2021).
Advances and Open Problems in Federated Learning.
Foundations and Trends® in Machine Learning, 14(1–2):1–210.

[Mironov, 2017] Mironov, I. (2017).
Rényi Differential Privacy.
In *CSF*.

[Nasr et al., 2019] Nasr, M., Shokri, R., and Houmansadr, A. (2019).
Comprehensive Privacy Analysis of Deep Learning: Passive and Active White-box Inference Attacks against Centralized and Federated Learning.
In *IEEE Symposium on Security and Privacy*.

[Paige et al., 2020] Paige, B., Bell, J., Bellet, A., Gascón, A., and Ezer, D. (2020).
Reconstructing Genotypes in Private Genomic Databases from Genetic Risk Scores.
In *International Conference on Research in Computational Molecular Biology RECOMB*.

[Sabater et al., 2020] Sabater, C., Bellet, A., and Ramon, J. (2020).
Distributed Differentially Private Averaging with Improved Utility and Robustness to Malicious Parties.
Technical report, arXiv:2006.07218.

[Shamir and Zhang, 2013] Shamir, O. and Zhang, T. (2013).
Stochastic Gradient Descent for Non-smooth Optimization: Convergence Results and Optimal Averaging Schemes.
In *ICML*.

[Shokri et al., 2017] Shokri, R., Stronati, M., Song, C., and Shmatikov, V. (2017).
Membership Inference Attacks Against Machine Learning Models.
In *IEEE Symposium on Security and Privacy (S&P)*.

Definition (Rényi Differential Privacy)

Let $\alpha > 1$, $\epsilon > 0$. A randomized algorithm $\mathcal{A}$ is (α, ϵ)-RDP if for every adjacent datasets $\mathcal{D} \sim \mathcal{D}'$, we have:

$$D_\alpha\left(\mathcal{A}(\mathcal{D}) \| \mathcal{A}(\mathcal{D}')\right) \leq \epsilon,$$

where $D_\alpha(P\|Q)$ is the Rényi divergence of order α between probability distributions P and Q defined as:

$$D_\alpha(P\|Q) = \frac{1}{\alpha - 1} \log \mathbb{E}_{x \sim Q} \left[\frac{P(x)}{Q(x)}\right]^\alpha .$$

Proposition (From RDP to (ϵ, δ)-DP)

If $\mathcal{A}$ is an (α, ϵ)-RDP algorithm, then it also satisfies $(\epsilon + \frac{\log(1/\delta)}{\alpha - 1}, \delta)$-DP for any $\delta \in (0, 1)$.

Proposition (Gaussian mechanism in RDP)

Let f be a function taking as input a dataset, and has L2 sensitivity bounded by Δ. Then $\mathcal{A}(\mathcal{D}) = f(\mathcal{D}) + \eta$ with $\eta \sim \mathcal{N}(0, \sigma^2 \mathbb{I})$ satisfies (α, ϵ)-RDP for any $\alpha > 1$ and $\epsilon = \frac{\alpha\Delta}{2\sigma^2}$.

Proposition (Composition under RDP)

If $\mathcal{A}_1$ satisfies (α, ϵ_1)-RDP and $\mathcal{A}_2$ satisfies (α, ϵ_2)-RDP, then $\mathcal{A} = (\mathcal{A}_1, \mathcal{A}_2)$ satisfies $(\alpha, \epsilon_1 + \epsilon_2)$-RDP.

- RDP keeps tracks of the distribution of the privacy loss random variable
- Privacy accounting is done in RDP; then given the desired δ for the final guarantee, α is optimized (analytically or numerically) to get the best ϵ
- In practice this is much better than resorting to advanced composition