FEDERATED LEARNING: ADVANCES AND OPEN CHALLENGES

Aurélien Bellet (MAGNET)

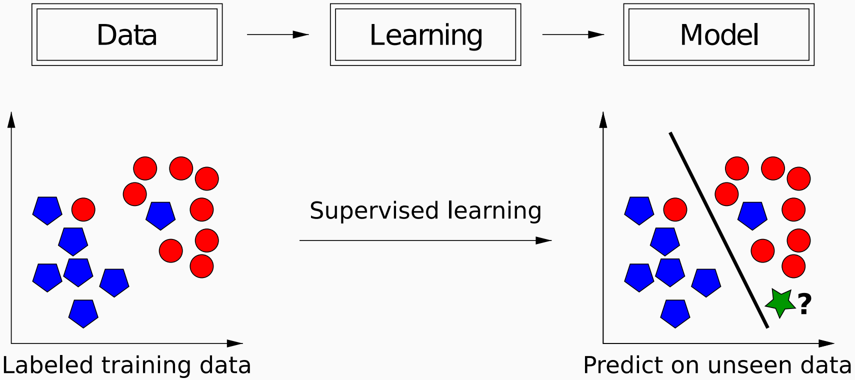
Journées Scientifiques Inria 2021
Session IA

OUTLINE

1. What is Federated Learning?
2. A concrete Federated Learning algorithm
3. Some challenges in Federated Learning
4. Wrapping up

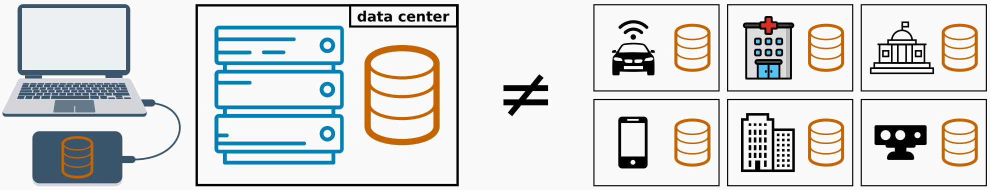
What is Federated Learning?

(SUPERVISED) MACHINE LEARNING



A SHIFT OF PARADIGM: FROM CENTRALIZED TO DECENTRALIZED DATA

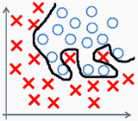
- The standard setting in Machine Learning (ML) considers a centralized dataset
- But in the real world data is often decentralized across different parties



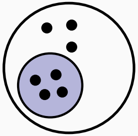
WHY DON'T WE ALWAYS CENTRALIZE DATA?

1. Sending the data may be too costly
    - Self-driving cars are expected to generate several TBs of data a day
    - Some wireless devices have limited bandwidth/power
2. Data may be considered too sensitive
    - Growing public awareness and regulations on data privacy
    - Keeping control of data can give a competitive advantage in business and research

1. The local dataset may be too small
   - Sub-par predictive performance (e.g., due to overfitting)
   - Non-statistically significant results (e.g., medical studies)



2. The local dataset may be biased
   - Not representative of the target distribution



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

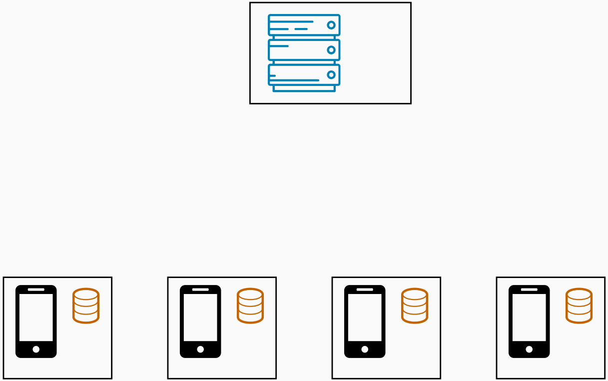
- FL is a booming topic
  - Term first coined in 2016; more than 1,000 papers in first half of 2020 alone[1]
  - First real-world deployments by companies and researchers
- FL is multidisciplinary: ML, optimization, privacy & security, networks, systems…
- FL could eventually enable remote data science, make AI accessible to citizens for collaborative tasks on personal data, …

[1] `https://www.forbes.com/sites/robtoews/2020/10/12/the-next-generation-of-artificial-intelligence/`

A concrete Federated Learning algorithm

- We consider a set of K parties
- Each party k holds a dataset $\mathcal{D}_k$ of n_k points, so there is $n = \sum_k n_k$ points in total
- We denote by θ the model parameters (e.g., weights of a neural network)
- We want to find the parameters that minimize the overall prediction error:

$$\min_{\theta} \sum_{k=1}^{K} \frac{n_k}{n} \text{Loss}(\theta; \mathcal{D}_k)$$



Algorithm FedAvg (server-side)

initialize θ

for each round $t = 0, 1, \ldots$ **do**

  for each party k in parallel **do**

    $\theta_k \leftarrow \text{ClientUpdate}(k, \theta)$

  $\theta \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} \theta_k$

Algorithm ClientUpdate(k, θ)

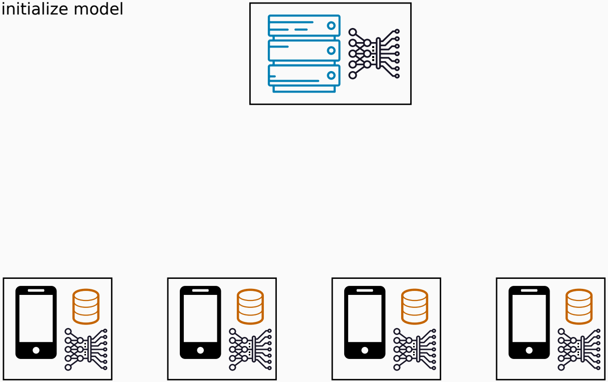
Parameters: # steps L, step size η

for $1, \ldots, L$ **do**

  $\theta \leftarrow \theta - \eta \nabla \text{Loss}(\theta; \mathcal{D}_k)$

send θ to server

A BASELINE FL ALGORITHM: FEDAVG [McMahan et al., 2017]



Algorithm FedAvg (server-side)

initialize θ

for each round $t = 0, 1, \ldots$ **do**

  for each party k in parallel **do**

    $\theta_k \leftarrow \text{ClientUpdate}(k, \theta)$

  $\theta \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} \theta_k$

Algorithm ClientUpdate(k, θ)

Parameters: # steps L, step size η

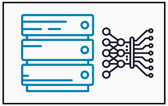
for $1, \ldots, L$ **do**

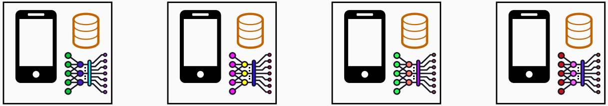
  $\theta \leftarrow \theta - \eta \nabla \text{Loss}(\theta; \mathcal{D}_k)$

send θ to server

A BASELINE FL ALGORITHM: FEDAVG [McMahan et al., 2017]

each party makes an update
using its local dataset





Algorithm FedAvg (server-side)

initialize θ

for each round $t = 0, 1, \ldots$ **do**

  for each party k in parallel **do**

    $\theta_k \leftarrow \text{ClientUpdate}(k, \theta)$

  $\theta \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} \theta_k$

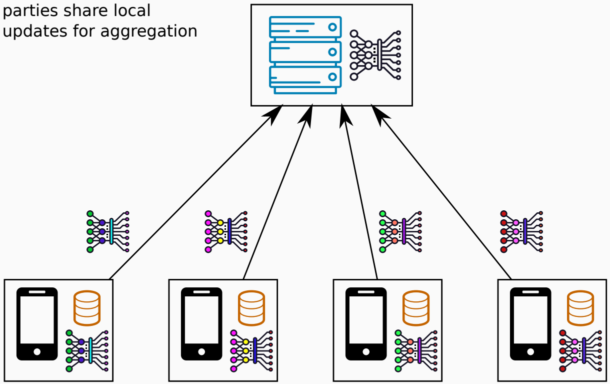
Algorithm ClientUpdate(k, θ)

Parameters: # steps L, step size η

for $1, \ldots, L$ **do**

  $\theta \leftarrow \theta - \eta \nabla \text{Loss}(\theta; \mathcal{D}_k)$

send θ to server



Algorithm FedAvg (server-side)

initialize θ

for each round $t = 0, 1, \ldots$ **do**

  for each party k in parallel **do**

    $\theta_k \leftarrow \text{ClientUpdate}(k, \theta)$

  $\theta \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} \theta_k$

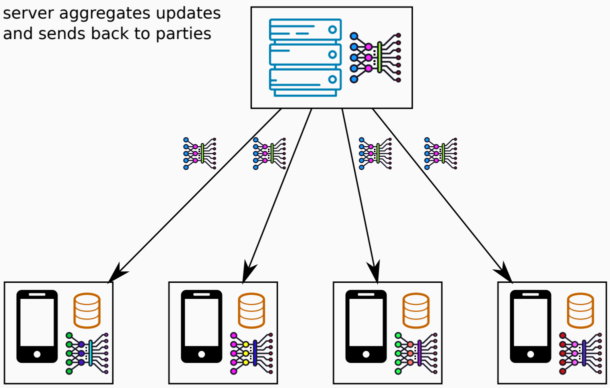
Algorithm ClientUpdate(k, θ)

Parameters: # steps L, step size η

for $1, \ldots, L$ **do**

  $\theta \leftarrow \theta - \eta \nabla \text{Loss}(\theta; \mathcal{D}_k)$

send θ to server



Algorithm FedAvg (server-side)

initialize θ

for each round $t = 0, 1, \ldots$ **do**

  for each party k in parallel **do**

    $\theta_k \leftarrow \text{ClientUpdate}(k, \theta)$

  $\theta \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} \theta_k$

Algorithm ClientUpdate(k, θ)

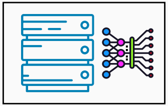
Parameters: # steps L, step size η

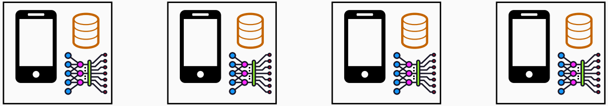
for $1, \ldots, L$ **do**

  $\theta \leftarrow \theta - \eta \nabla \text{Loss}(\theta; \mathcal{D}_k)$

send θ to server

parties update their copy
of the model and iterate





Algorithm FedAvg (server-side)

initialize θ
for each round $t = 0, 1, \ldots$ **do**
  for each party k in parallel **do**
    $\theta_k \leftarrow \text{ClientUpdate}(k, \theta)$
  $\theta \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} \theta_k$

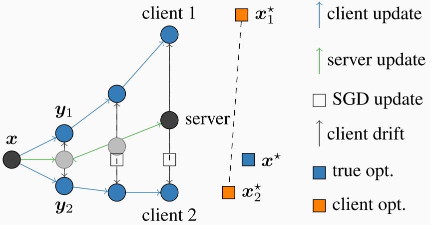
Algorithm ClientUpdate(k, θ)

Parameters: # steps L, step size η
for $1, \ldots, L$ **do**
  $\theta \leftarrow \theta - \eta \nabla \text{Loss}(\theta; \mathcal{D}_k)$
send θ to server

Some challenges in Federated Learning

1. Dealing with heterogeneous data

- Unlike distributed ML on a cluster, local data distributions may be arbitrarily different
- When data is heterogeneous across parties, FedAvg suffers from local drift



(Figure taken from [Karimireddy et al., 2020])

- Challenges: design algorithms which minimize communication costs, ensure that model is fair to all parties, automatically adapt the network topology...

- Instead of training a single global model, learn personalized models collaboratively!
- Inspired by multi-task learning, we proposed to learn personalized models along with relationships between tasks in fully decentralized networks:[2]

$$F(\theta_1, \dots, \theta_K, W; \mathcal{D}) = \frac{1}{K}\sum_{k=1}^{K} \mathrm{Loss}(\theta_k; \mathcal{D}_k) + \sum_{k<l} W_{k,l} \|\theta_k - \theta_l\|^2$$

- Ongoing collaboration with NEO team (G. Neglia) on formulations based on clear statistical assumptions that can offer generalization guarantees

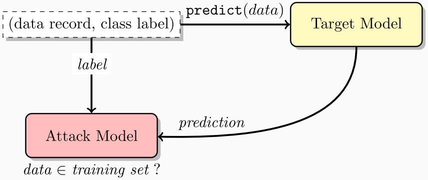
[2] [Vanhaesebrouck et al., 2017, Bellet et al., 2018, Zantedeschi et al., 2020]

Some challenges in Federated Learning

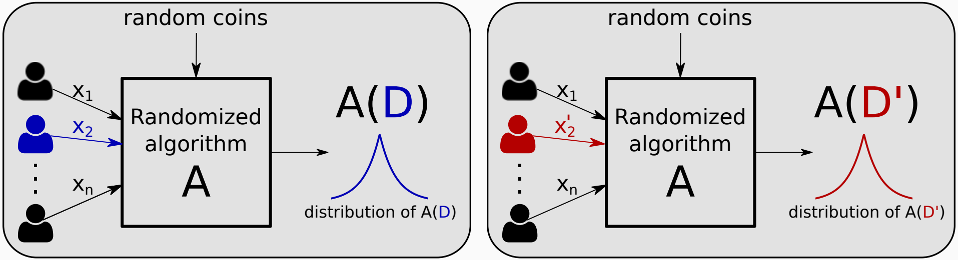
2. Preserving Privacy

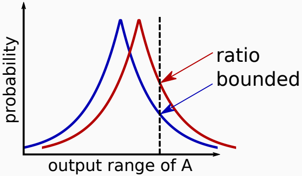
PRIVACY ISSUES IN (FEDERATED) ML

- ML models are susceptible to various attacks on data privacy
- Membership inference attacks try to infer the presence of a known individual in the training set [Shokri et al., 2017]



- Reconstruction attacks try to infer some of the points used to train the model [Paige et al., 2020]
- Federated Learning offers an additional attack surface because the server and/or other parties observe model updates (not only the final model) [Nasr et al., 2019]





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

$\mathcal{A}$ is ε-differentially private (DP) if for all neighboring datasets $\mathcal{D} = \{x_1, x_2, \ldots, x_n\}$ and $\mathcal{D}' = \{x_1, x'_2, x_3, \ldots, x_n\}$ and all sets S:

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

DIFFERENTIALLY PRIVATE AGGREGATION

- In most FL algorithms, parties interact through an aggregation step $\theta \leftarrow \frac{1}{K}\sum_k \theta_k$
- DP in centralized setting: trusted curator adds (Gaussian) noise to the average θ
- DP in FL setting: each party k adds noise to its local update θ_k before sharing it
- The error due to privacy is $O(\sqrt{K})$ larger in the FL case
- Challenges: improve the privacy-utility trade-off while maintaining efficiency, model rich privacy constraints in complex systems...

A SIMPLE PROTOCOL FOR DP AGGREGATION WITH BETTER UTILITY [Sabater et al., 2020]

Algorithm GOPA protocol

Parameters: graph G, variances $\sigma_\Delta^2, \sigma_\eta^2 \in \mathbb{R}^+$

for all neighboring parties $\{k, l\}$ in G **do**

  k and l draw $y \sim \mathcal{N}(0, \sigma_\Delta^2)$

  set $\Delta_{k,l} \leftarrow y$, $\Delta_{l,k} \leftarrow -y$

for each party k **do**

  k draws $\eta_k \sim \mathcal{N}(0, \sigma_\eta^2)$

  k reveals $\hat{\theta}_k \leftarrow \theta_k + \sum_{l \sim k} \Delta_{k,l} + \eta_k$

1. Pairs of parties securely exchange pairwise-canceling Gaussian noise
2. Each party generates personal Gaussian noise
3. Each party reveals sum of local update, pairwise and personal noise terms

- Private & accurate: result $\hat{\theta} = \frac{1}{K} \sum_k \hat{\theta}_k$ can match the accuracy of centralized setting
- Scalable: it is sufficient for each party to communicate with $O(\log K)$ others
- Robust: it can handle some collusions, dropouts and even malicious behavior

Some challenges in Federated Learning

3. Putting FL to Practice

PUTTING FL TO PRACTICE

- Technological challenges: develop general-purpose software libraries which can be easily deployed in production systems
- Regulatory/legal challenges: when should model updates be considered as personal data? how to ensure compliance with current regulations (e.g., GDPR)?
- Convincing stakeholders: what are the key merits of FL for a given application? how to make FL as transparent as possible to the end-users?

COLLABORATION WITH LILLE UNIVERSITY HOSPITAL

- We are currently exploring these questions with Lille University Hospital (INCLUDE team) in the context of AEx FLAMED
- We have started developing our own FL library and will soon deploy a proof-of-concept across 4 hospitals of the GCS G4
- We will have some official support from CNIL on legal aspects in the context of its Bac à Sable 2021[3]
- Discussions with EPIONE team, who are also working on FL applied to health data

[3] https://www.cnil.fr/fr/bac-sable-donnees-personnelles-la-cnil-accompagne-12-projets-dans-le-domaine-de-la-sante-numerique

Wrapping up

Survey paper: Advances and Open Problems in FL [Kairouz et al., 2021]

- A large collaborative effort (50+ authors!)
- Updated in December 2020, to appear in FnTML 2021

Online seminar: Federated Learning One World (FLOW)

`https://sites.google.com/view/one-world-seminar-series-flow/`

- Weekly talks (usually on Wednesdays, 1pm UTC) covering all aspects of FL
- The videos and slides of all previous talks are available online

THANK YOU FOR YOUR ATTENTION!

QUESTIONS?

[Bellet et al., 2018] Bellet, A., Guerraoui, R., Taziki, M., and Tommasi, M. (2018).
Personalized and Private Peer-to-Peer Machine Learning.
In *AISTATS*.

[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)*.

[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.

[Karimireddy et al., 2020] Karimireddy, S. P., Kale, S., Mohri, M., Reddi, S. J., Stich, S. U., and Suresh, A. T. (2020).
SCAFFOLD: Stochastic Controlled Averaging for On-Device Federated Learning.
In *ICML*.

[McMahan et al., 2017] McMahan, H. B., Moore, E., Ramage, D., Hampson, S., and Agüera y Arcas, B. (2017).
Communication-efficient learning of deep networks from decentralized data.
In *AISTATS*.

[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.

[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)*.

[Vanhaesebrouck et al., 2017] Vanhaesebrouck, P., Bellet, A., and Tommasi, M. (2017).
Decentralized collaborative learning of personalized models over networks.
In *AISTATS*.

[Zantedeschi et al., 2020] Zantedeschi, V., Bellet, A., and Tommasi, M. (2020).
Fully Decentralized Joint Learning of Personalized Models and Collaboration Graphs.
In *AISTATS*.