# A Quick Introduction to Federated Learning

Methods, Challenges, and Applications

Ethan Young

UCLA

Stats 147: Data Technologies for Data Scientists
Winter 2023

# What is Federated Learning?

**Federated learning** (FL) is a machine learning (ML) technique that trains an algorithm across multiple servers (nodes) holding local data, without exchanging them

# What is Federated Learning?

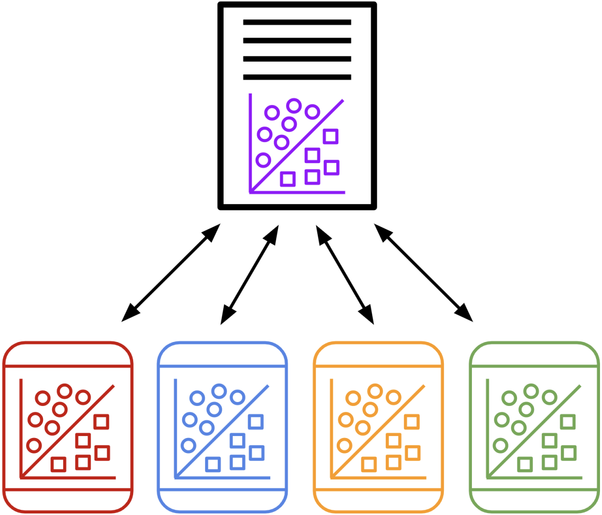
There are 3 main FL settings[1]:

- **Centralized**
  - Central server coordinates the participating nodes in the learning process
- **Decentralized**
  - Nodes coordinate themselves to train the global model
- **Heterogeneous**[2]
  - Involves a large set of heterogeneous clients such as mobile phones and internet of things (IoT) devices

[1] Kairouz et al., *Advances and Open Problems in Federated Learning*, 2019

[2] Diao, Ding, and Tarokh, *HeteroFL: Computation and Communication Efficient Federated Learning for Heterogeneous Clients*, 2020

# What is Federated Learning?



# Why Federated Learning?

The appeal of FL lies in building a robust ML model without sharing data, which helps to address major societal concerns such as data privacy, data security, and access to data[3]

[3] Li et al., "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection", 2021

# Learning Procedure

A summary[4] of the learning procedure in the centralized setting is as follows:

1. Initialization
2. Client selection
3. Configuration
4. Reporting
5. Termination

[4] Bonawitz et al., *Towards Federated Learning at Scale: System Design*, 2019

# Initialization

1. **Initialization**
   - Initialize a ML model to be trained on the nodes
2. Client selection
3. Configuration
4. Reporting
5. Termination

# Client Selection

1. Initialization
2. **Client selection**
   - A fraction of nodes are selected to start training on local data
3. Configuration
4. Reporting
5. Termination

# Configuration

1. Initialization
2. Client selection
3. **Configuration**
   - Central server orders selected nodes to train the model on their local data
4. Reporting
5. Termination

# Reporting

1. Initialization
2. Client selection
3. Configuration
4. **Reporting**
   - Selected nodes send their models to the central server for aggregation
   - Central server aggregates the models and sends updated model to nodes
   - Next round starts by returning to client selection
5. Termination

# Termination

1. Initialization
2. Client selection
3. Configuration
4. Reporting
5. **Termination**
   - When a pre-specified criterion is met the central server aggregates the updated models into the final global model

# Popular Variations

- There are an overwhelming number of FL variations that exist to address issues such as heterogeneous data and non-IID data
- Below are 2 that are widely-used as benchmarks and serve as the foundation of many other FL frameworks include:
  - **Federated stochastic gradient descent**[5] (FedSGD)
  - **Federated averaging**[6] (FedAvg)

[5] Shokri and Shmatikov, "Privacy-Preserving Deep Learning", 2015

[6] McMahan et al., *Communication-Efficient Learning of Deep Networks from Decentralized Data*, 2016

# Recent Advances

- It is difficult to know which methods are preferable to others due to constant developments in the field
- Below are several (of many) relatively recent advances that have captured much attention:
  - **Inverse Distance Aggregation**[7] (IDA)
  - **FL with Dynamic Regularization**[8] (FedDyn)
  - **Hybrid Federated Dual Coordinate Ascent**[9] (HyFDCA)

[7] Yeganeh et al., *Inverse Distance Aggregation for Federated Learning with Non-IID Data*, 2020

[8] Acar et al., *Federated Learning Based on Dynamic Regularization*, 2021

[9] Overman, Blum, and Klabjan, *A Primal-Dual Algorithm for Hybrid Federated Learning*, 2022

# Adversarial Attacks

- Understanding the impact of malicious actors (attackers) is a major challenge to the robustness of models learned by FL
- Chen et al.[10] details several types of attacks on different aspects of FL
  - Examples include **Byzantine attacks**, **reconstruction attacks**, and **poisoning attacks**

[10] Chen et al., *Federated Learning Attacks and Defenses: A Survey*, 2022

# Defense Mechanisms

- An active area of FL research is developing defense methods to prevent data breaches
- Chen et al.[11] describes 2 different levels of defenses: **security-based** and **privacy-based**
    - Examples include **data anonymization**, **differential privacy**, and **secure multi-party computation**

[11] Chen et al., *Federated Learning Attacks and Defenses: A Survey*, 2022

# Healthcare

- The capacity for FL to address challenges of data privacy makes it a powerful tool for ML applications in healthcare
- **Partial meta-federated learning**[12] (PMFL) shows great potential in its fast training speed and high accuracy when applied to heterogeneous medical records

[12] Zhang et al., *PMFL: Partial Meta-Federated Learning for heterogeneous tasks and its applications on real-world medical records*, 2021

# Satellite Constellations

- Low Earth Orbit (LEO) constellations that contain many satellites have become a large data source, although that data is expensive and slow to transfer
- **Asynchronous federated learning for LEO satellite constellations**[13] (AsyncFLEO) outperforms existing methods by increasing convergence time and model accuracy

[13] Elmahallawy and Luo, *AsyncFLEO: Asynchronous Federated Learning for LEO Satellite Constellations with High-Altitude Platforms*, 2022

# Internet of Things

- ML models are increasingly popular in industrial settings due to sensor data from production machinery becoming more widely available
- **Autoencoder-based federated learning**[14] applied to sensor data reduced the network usage and demonstrates the success of FL in the industrial IoT

[14] Becker et al., *Federated Learning for Autoencoder-based Condition Monitoring in the Industrial Internet of Things*, 2022

# Additional Resources

- ***Federated Learning***[15] is an in-depth exploration of relevant challenges and methods in FL
- **FedML**[16] is an open-source and collaborative research library that supports FL algorithm development
  - Other libraries mentioned by the authors include **TensorFlow Federated** (TFF) and **PySyft**
- **OpenFL**[17] is another open-source framework with TensorFlow and PyTorch training pipelines

[15] Ludwig and Baracaldo, *Federated Learning*, 2022

[16] He et al., *FedML: A Research Library and Benchmark for Federated Machine Learning*, 2020

[17] Foley et al., *OpenFL: the open federated learning library*, 2022

# Current Landscape

- While FL circumvents a number of issues faced by traditional, centralized ML approaches, many open problems remain[18]
- These include node trustworthiness, robustness to adversarial attacks, improvements to communication efficiency, and development of privacy-preserving techniques

[18] Kairouz et al., *Advances and Open Problems in Federated Learning*, 2019

# References I


- Acar, Durmus Alp Emre et al. *Federated Learning Based on Dynamic Regularization*. 2021. DOI: 10.48550/ARXIV.2111.04263.
- Becker, Soeren et al. *Federated Learning for Autoencoder-based Condition Monitoring in the Industrial Internet of Things*. 2022. DOI: 10.48550/ARXIV.2211.07619.
- Bonawitz, Keith et al. *Towards Federated Learning at Scale: System Design*. 2019. DOI: 10.48550/ARXIV.1902.01046.
- Chen, Yao et al. *Federated Learning Attacks and Defenses: A Survey*. 2022. DOI: 10.48550/ARXIV.2211.14952.
- Diao, Enmao, Jie Ding, and Vahid Tarokh. *HeteroFL: Computation and Communication Efficient Federated Learning for Heterogeneous Clients*. 2020. DOI: 10.48550/ARXIV.2010.01264.

# References II


- Elmahallawy, Mohamed and Tie Luo. *AsyncFLEO: Asynchronous Federated Learning for LEO Satellite Constellations with High-Altitude Platforms*. 2022. DOI: 10.48550/ARXIV.2212.11522.
- Foley, Patrick et al. *OpenFL: the open federated learning library*. 2022. DOI: 10.1088/1361-6560/ac97d9.
- He, Chaoyang et al. *FedML: A Research Library and Benchmark for Federated Machine Learning*. 2020. DOI: 10.48550/ARXIV.2007.13518.
- Kairouz, Peter et al. *Advances and Open Problems in Federated Learning*. 2019. DOI: 10.48550/ARXIV.1912.04977.
- Li, Qinbin et al. "A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection". In: *IEEE Transactions on Knowledge and Data Engineering* (2021). DOI: 10.1109/tkde.2021.3124599.

# References III


- Ludwig, Heiko and Nathalie Baracaldo, eds. *Federated Learning. A Comprehensive Overview of Methods and Applications*. Springer Cham, 2022. DOI: https://doi.org/10.1007/978-3-030-96896-0.
- McMahan, H. Brendan et al. *Communication-Efficient Learning of Deep Networks from Decentralized Data*. 2016. DOI: 10.48550/ARXIV.1602.05629.
- Overman, Tom, Garrett Blum, and Diego Klabjan. *A Primal-Dual Algorithm for Hybrid Federated Learning*. 2022. DOI: 10.48550/ARXIV.2210.08106.
- Shokri, Reza and Vitaly Shmatikov. "Privacy-Preserving Deep Learning". In: *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*. 2015. DOI: 10.1145/2810103.2813687.

# References IV


- Yeganeh, Yousef et al. *Inverse Distance Aggregation for Federated Learning with Non-IID Data*. 2020. DOI: 10.48550/ARXIV.2008.07665.
- Zhang, Tianyi et al. *PMFL: Partial Meta-Federated Learning for heterogeneous tasks and its applications on real-world medical records*. 2021. DOI: 10.48550/ARXIV.2112.05321.