Federated Learning
in the Financial Industry

Jesse Cresswell, PhD
Senior Machine Learning Scientist
Layer 6 AI at TD

Oct. 6 2022 - Big Data & AI Conference Toronto

Agenda

- Motivation
- Federated Learning
- Applications

Data Privacy

Privacy is not security - privacy is about extracting general knowledge without revealing specific knowledge. It may entail collecting *less* data about our customers.

How can privacy be respected while extracting insights from data?

As the velocity of data increases, we need to rethink how we provide privacy to our customers. Consider anonymization:

Name	Age	Job	Salary
[redacted]	23	Clerk	50,000
[redacted]	45	Driver	60,000
[redacted]	61	Lawyer	100,000

Data Privacy

Privacy is not security - privacy is about extracting general knowledge without revealing specific knowledge. It may entail collecting *less* data about our customers.

How can privacy be respected while extracting insights from data?

As the velocity of data increases, we need to rethink how we provide privacy to our customers. Consider anonymization:

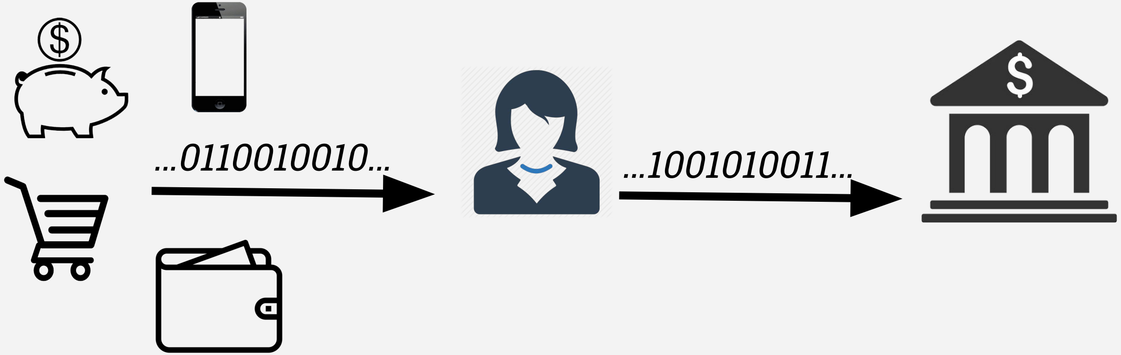
Name	Age	Job	Salary
[redacted]	23	Clerk	50,000
[redacted]	45	Driver	60,000
[redacted]	61	Lawyer	100,000

Name	Age	Job	Favorite Sport
Joe	[redacted]	Clerk	Soccer
Jun	[redacted]	Driver	Squash
Jaya	[redacted]	Lawyer	Hockey

Data Privacy

How we think about providing the best experience for our customers:

- A minimal amount of personal information is collected
- Customers are not beholden to credit scores, but can prove creditworthiness
- Open banking provides decentralized data ownership
- Improved fraud detection and risk management lowers costs



layer 6
AI at TD

Privacy Enhancing Technologies

The aim of PETs are to **minimize the risk** to individuals that their personally identifiable data will be exposed, while **maximizing the utility** of that data for analysis.

Open data is useful, but not private. Siloed data is safe, but not useful.

Four emerging PETs actively being researched in ML:

Federated Learning	Differential Privacy
Secure Multi-Party Computation	Homomorphic Encryption

Federated Learning

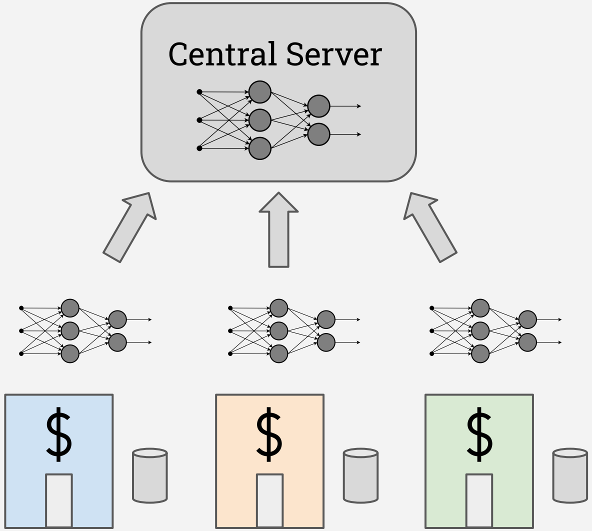
Federated Learning

The power of machine learning techniques scales with the amount and diversity of available data.

Federated Learning is a distributed ML approach where data is not pooled together on a centralized server.

Models are trained at the device/institution where data is collected.

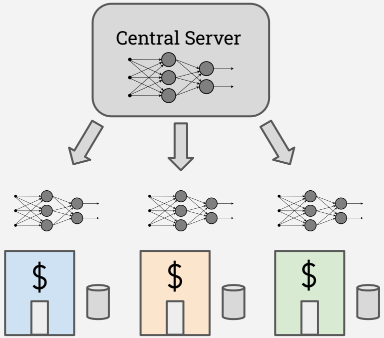
Intuitively this is more private, since the raw data never leaves the device/institution where it was generated.



Federated Learning - *FedAvg* [McMahan et al. 2017]

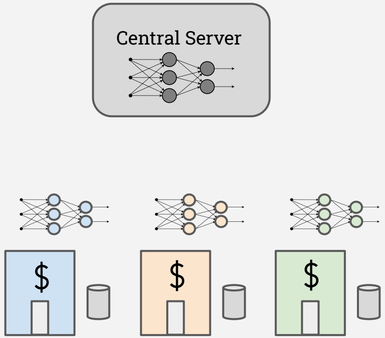
Repeat until convergence:

1.



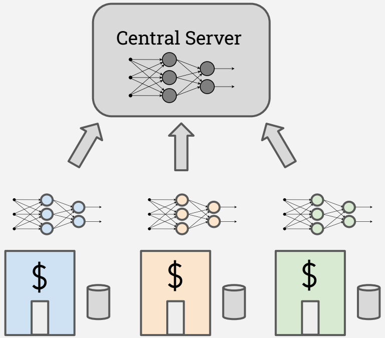
The central model is shared to each institution.

2.



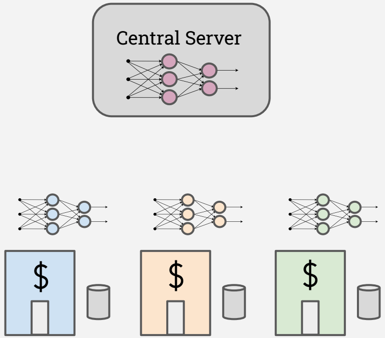
Institutions locally train the model on their data.

3.



Model updates are sent to the server.

4.



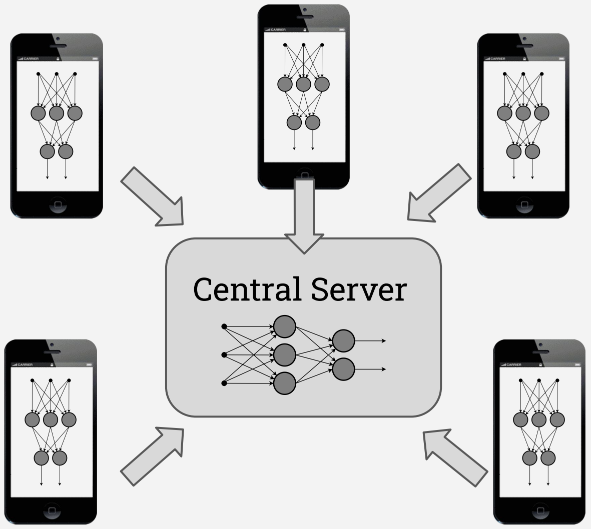
The server averages the updates, and applies them to the central model

Federated Learning - Settings

Federated learning was originally developed for computation across mobile devices.

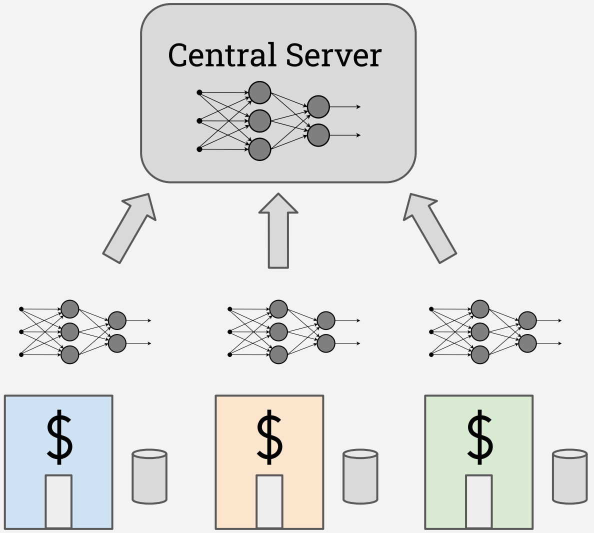
Cross-device FL:

Many clients, little data, low processing power,
costly communication, unreliable connections



Cross-silo FL:

Few clients, large datasets,
high processing power, reliable connections



Federated Learning - Weaknesses

Memorization and reconstruction attacks

There are many examples of neural networks memorizing individual training datapoints, which can be recovered by analyzing the model. [Fredrikson, Jha, Ristenpart 2015]

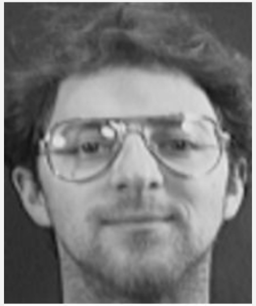
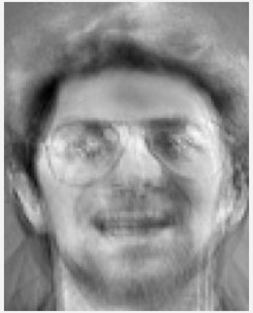


Image seen by model during training

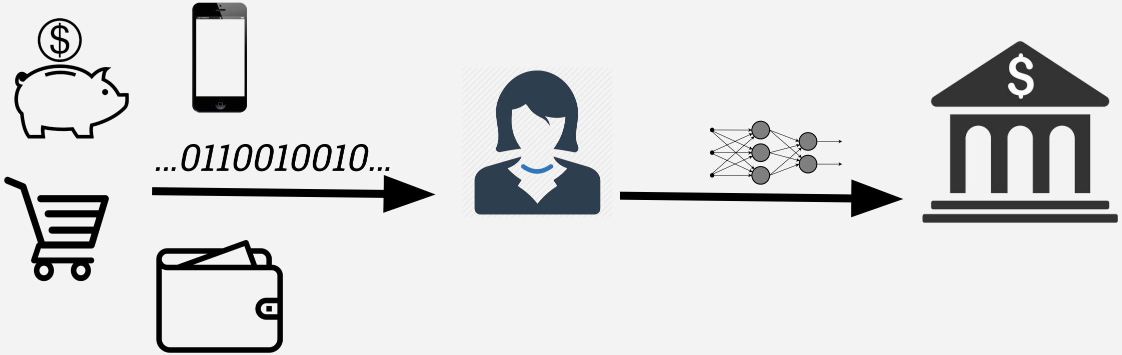


Reconstructed image inferred from trained model

FL should be used in conjunction with other PETs to strengthen privacy

Applications

Applications – Open Banking

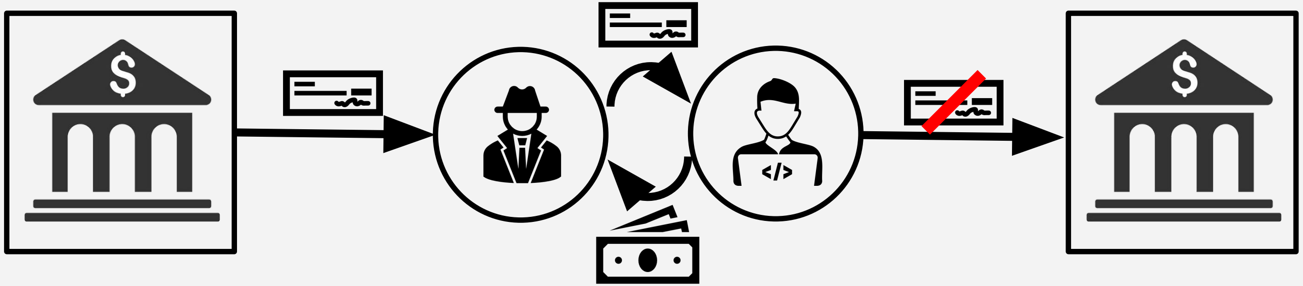


Customers retain control over who can access their data, while institutions can learn behaviour from richer, more diverse datasets.

Through FL, even when customers opt-in, their raw data is never exposed. Only high-level information in the form of a model update is shared.

Applications – Fraud Detection

Fraudulent activities and money laundering often include multiple parties across multiple institutions.
Malicious transactions may appear benign to every institution involved, but can be detected in a network.



FL provides a direct link between institutions to privately share information about fraudulent activities and learn detectable patterns.

Conclusions – Future State

Federated learning may enable institutions to collaborate with one another while better protecting people's privacy.

It will be used in the future financial system where data is selectively provided by individuals.

Areas for further development:

Personalization: Central models may not work equally well for all participants.

Decentralization: Eliminating central models puts control in the hands of participants.

Privacy: Other PETs can supplement FL to provide rigorous privacy guarantees.

Thank You!

jesse.cresswell@td.com