Federated Learning

Idea, Applications, and

Daniel Schalk

November 5, 2018

Table of Contents

Terminology

Distributed Learning

Classical parallelization, advantages:

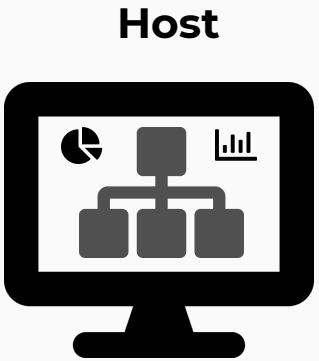
- Speed up fitting process
- Train model on much more data
- Idea behind Spark, Hadoop, ...
- Assumption that we already have a database which we want to distribute, hence data of the splits should follow the same distribution

Federated/Decentralized Learning

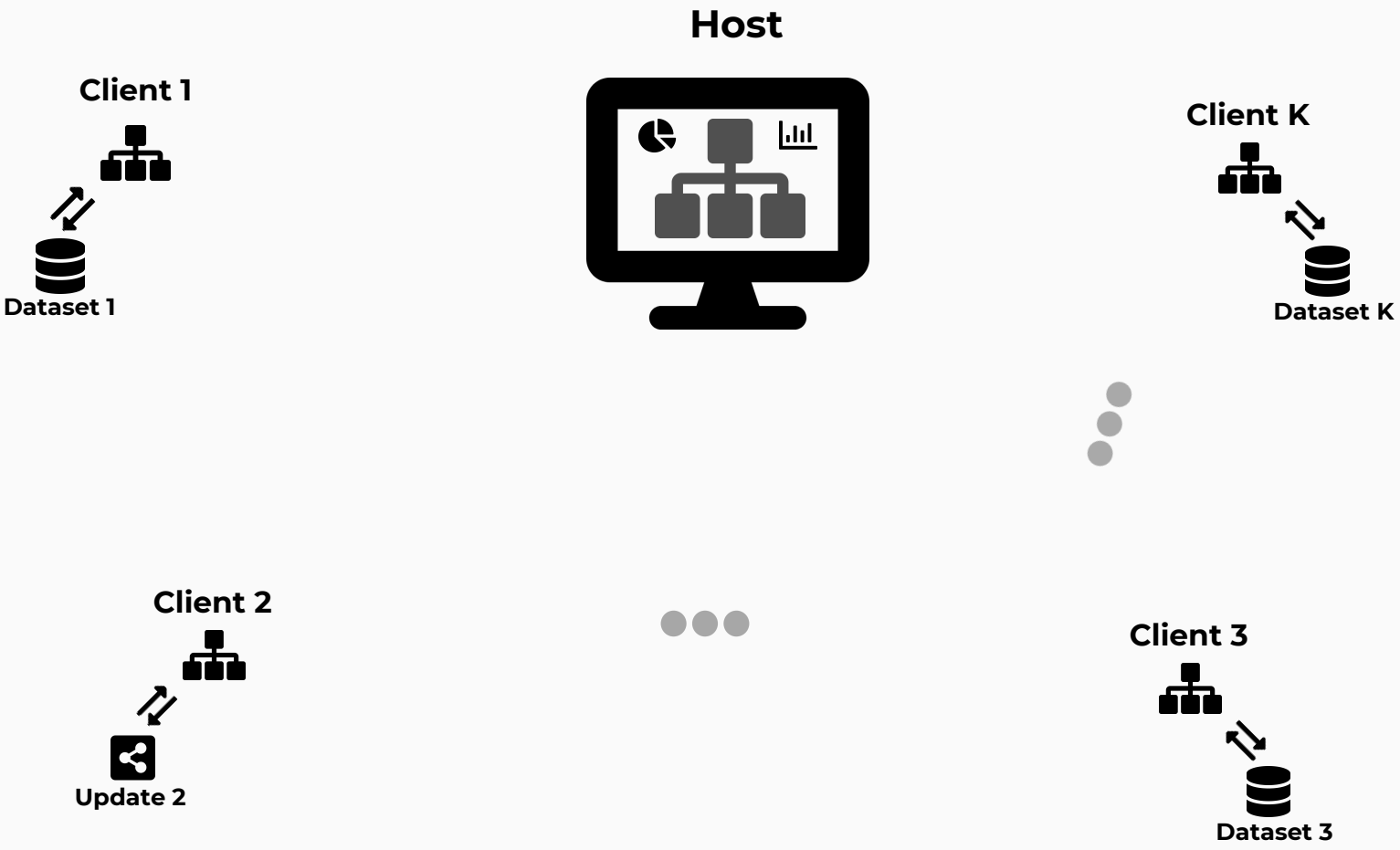
Federated Learning as learning on decentralized data with the following properties:

- **Non-IID** The training data on a given client is typically based on the usage of the mobile device by a particular user, and hence any particular user's local dataset will not be representative of the population distribution.
- **Unbalanced** Similarly, some users will make much heavier use of the service or app than others, leading to varying amounts of local training data.
- **Massively distributed** We expect the number of clients participating in an optimization to be much larger than the average number of examples per client.
- **Limited communication** Mobile devices are frequently offline or on slow or expensive connections.

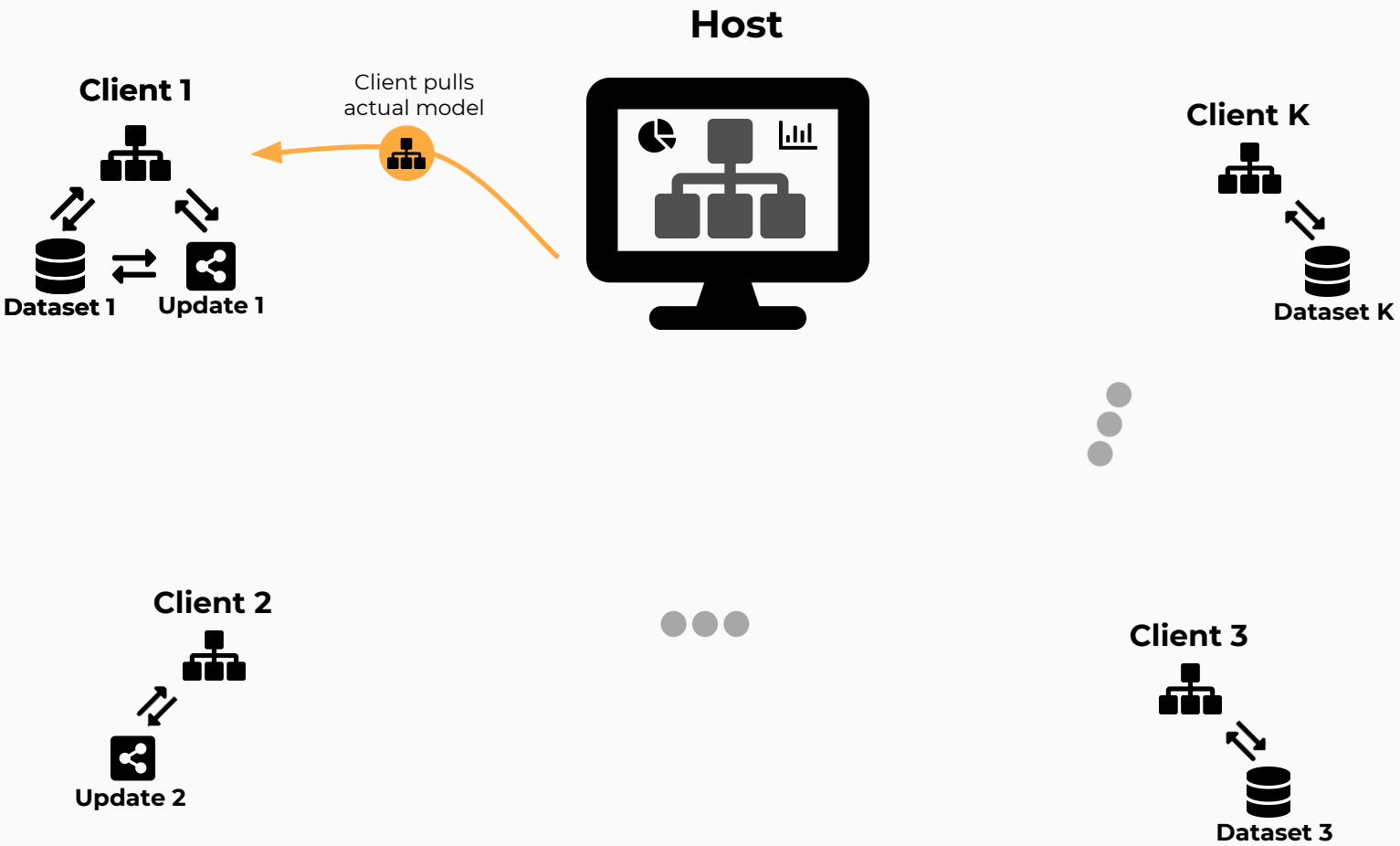
What is it About?



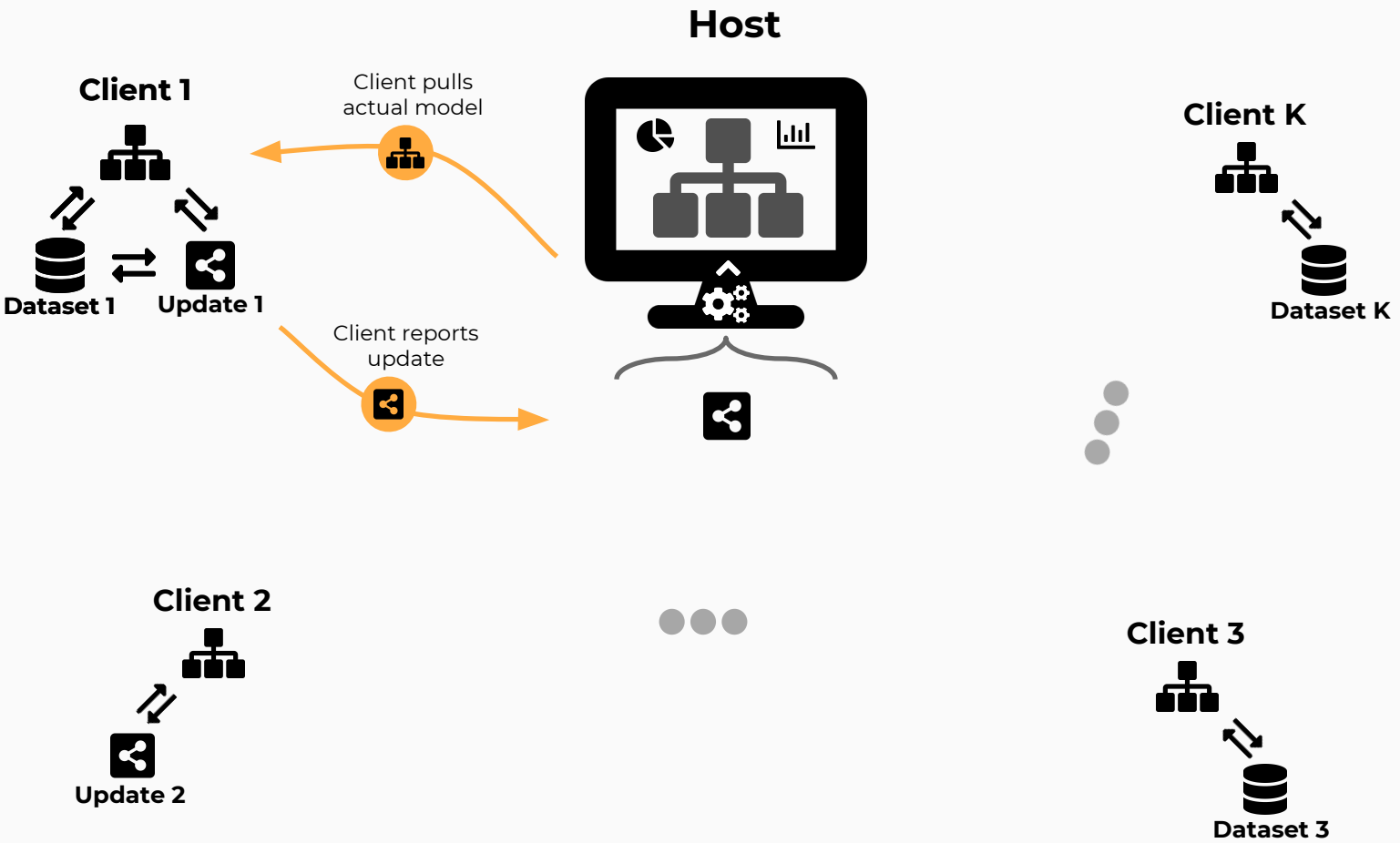
What is it About?



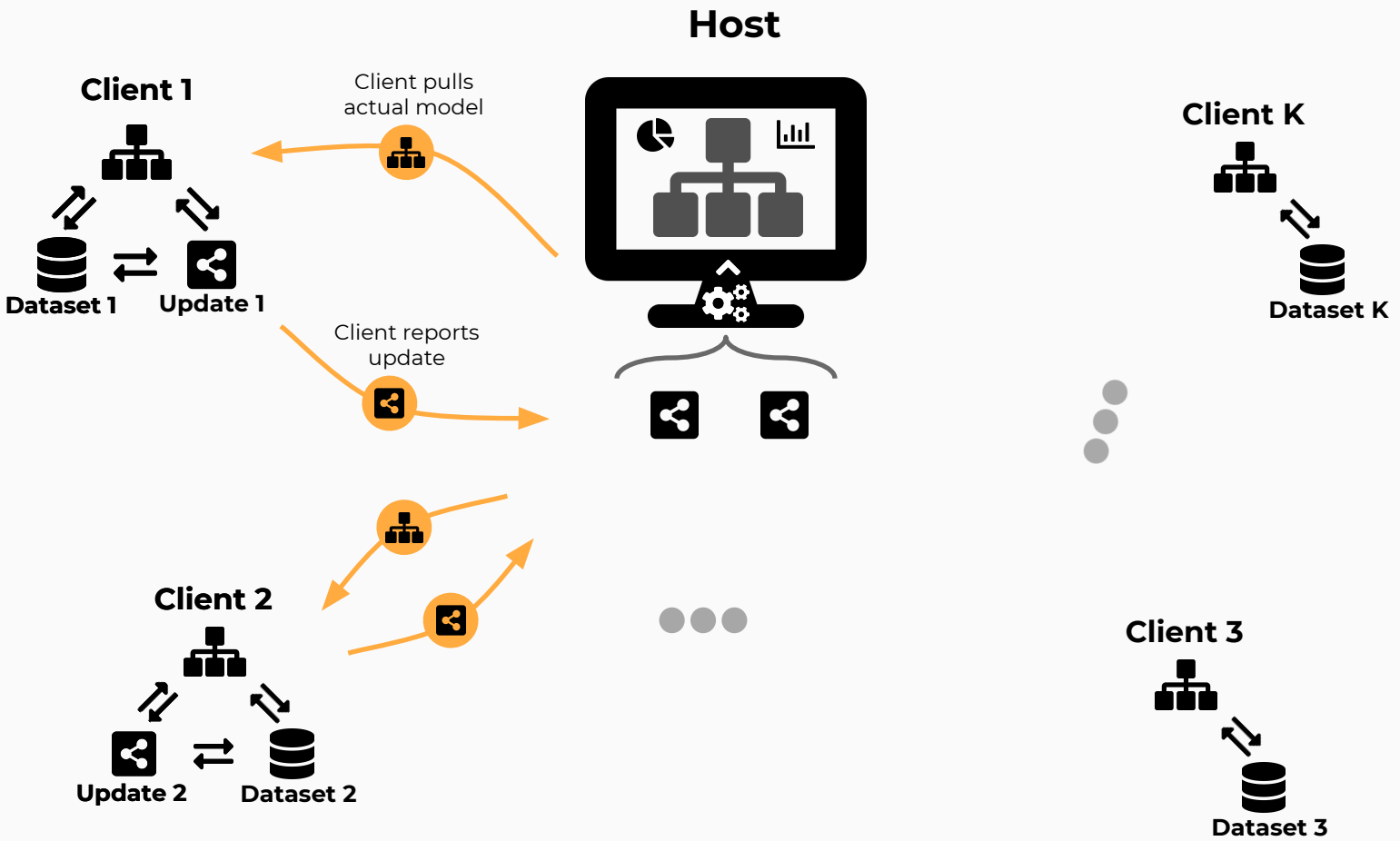
What is it About?



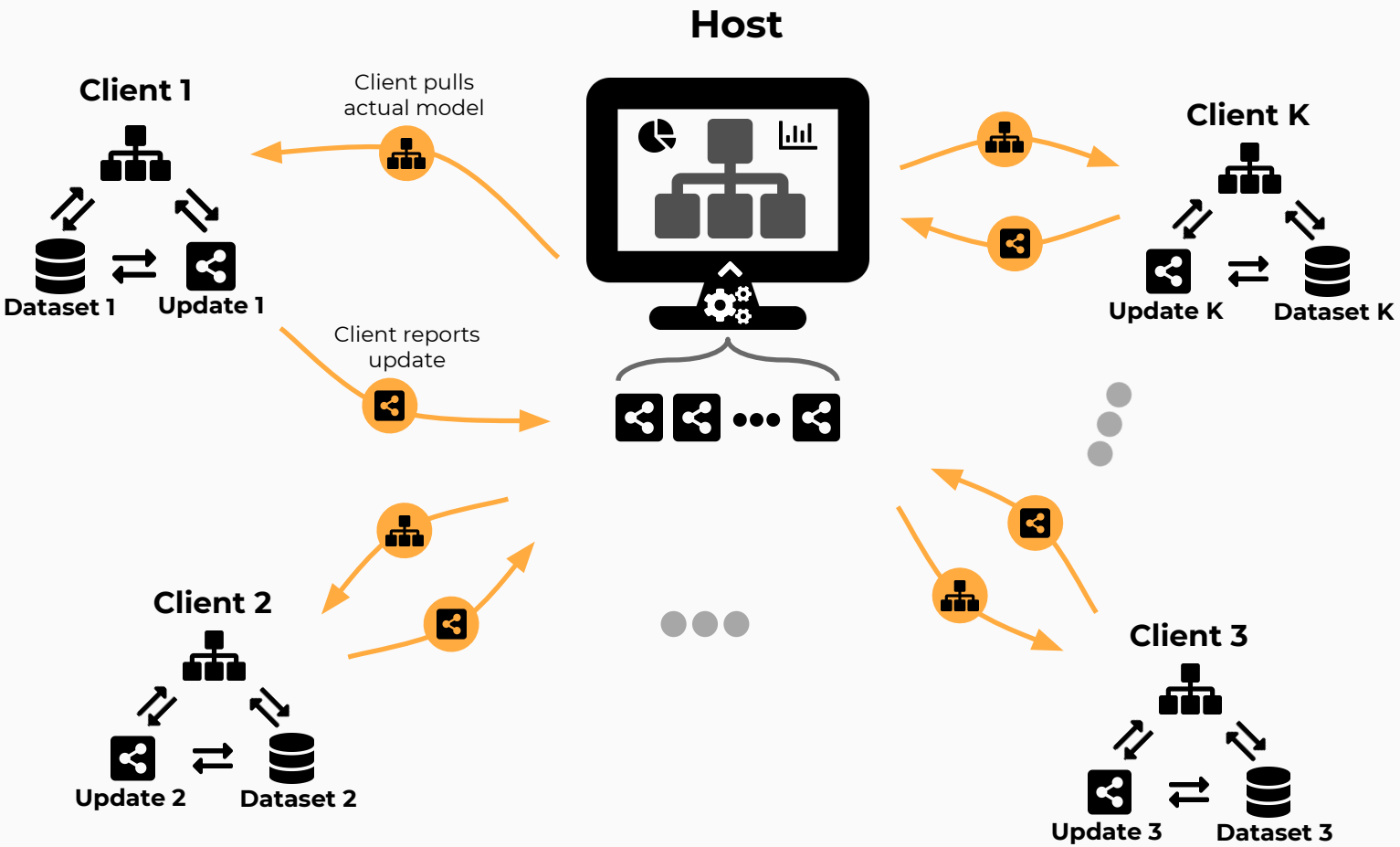
What is it About?



What is it About?



What is it About?



Federated Learning/Decentralized Learning

- Model comes to the data, not data to the model
- Privacy concerning method
- ...

Federated Learning of Gradient-Based Methods

Google Paper

- Problems

Federated Averaging

- Mention problems and how we can tackle them (more steps in one iteration . . .)

Challenges

Communication Costs vs. Training Costs

Evaluation of Federated Learning Systems

Example with Logistic Regression

Setup

- **Loss Function/Negative Log-Likelihood**

$$L\left(y, f(x,\theta)\right) = -y \log\left(f(x,\theta)\right) - (1-y) \log\left(1 - f(x,\theta)\right)$$

- **Response Function**

$$f(x) = \left(1 + \exp\left(-x^T\theta\right)\right)^{-1}$$

- **Score Function**

$$\frac{\delta}{\delta\theta} L(y, f\left(x,\theta\right)) = x\left(y - f(x,\theta)\right)$$

Boosting and Federated Learning