Federated Optimization

Key properties different from a typical distributed optimization problem

- Non i.i.d
  - Local dataset will not be representative of the population distributio
- Unbalanced
  - Varying amounts of local training data
- Massively distributed
  - The number of clients is much larger than the number of examples per client
- Limited communication
  - Devices are frequently offline or on slow

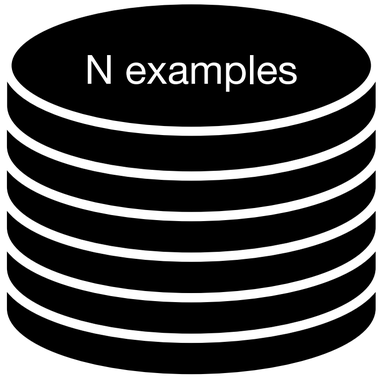
Communication-Efficient Learning of Deep Networks from Decentralized Data, AISTATS, 2017

Federated Optimization

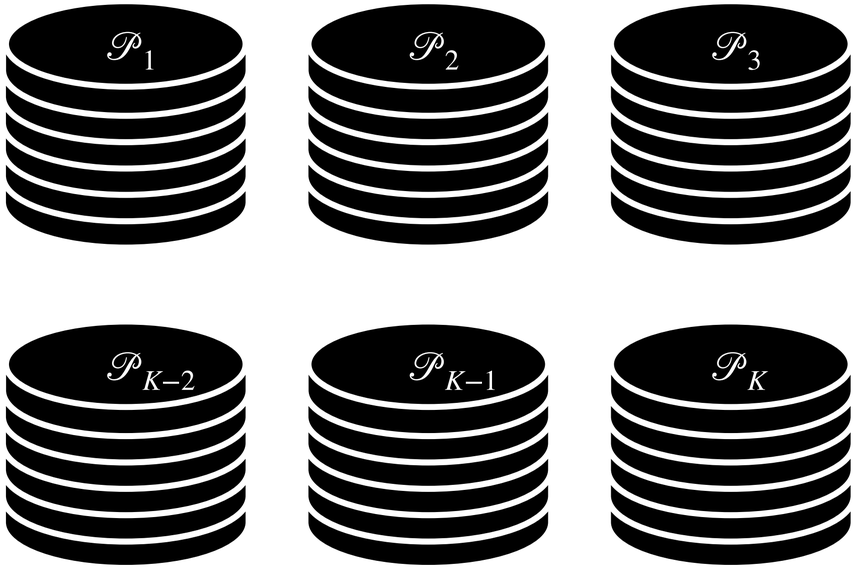
Key properties different from a typical distributed optimization problem

- Non i.i.d
  - Local dataset will not be representative of the population distributio
- Unbalanced
  - Varying amounts of local training data
- Massively distributed
  - The number of clients is much larger than the number of examples per client
- Limited communication
  - Devices are frequently offline or on slow

Communication-Efficient Learning of Deep Networks from Decentralized Data, AISTATS, 2017



$$\min_{w \in \mathbb{R}^d} f(w) \quad \text{where} \quad f(w) = \frac{1}{N} \sum_{i=1}^{N} f_i(w), \quad f_i(w) = \ell(D_i;\, w)$$



$$f(w) = \sum_{k=1}^{K} \frac{n_k}{n} F_k(w), \quad \text{where} \quad F_k(w) = \frac{1}{n_k} \sum_{i \in \mathcal{P}_k} f_i(w)$$

$$\nabla f(w) = \sum_{k=1}^{K} \frac{n_k}{n} \underbrace{\nabla F_k(w)}_{g_k}, \quad \nabla F_k(w) = \frac{1}{n_k} \sum_{i \in \mathcal{P}_k} \nabla f_i(w)$$

Communication-Efficient Learning of Deep Networks from Decentralized Data, AISTATS, 2017