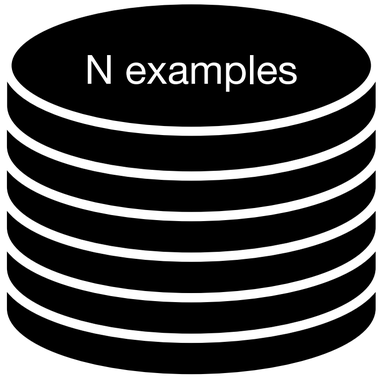
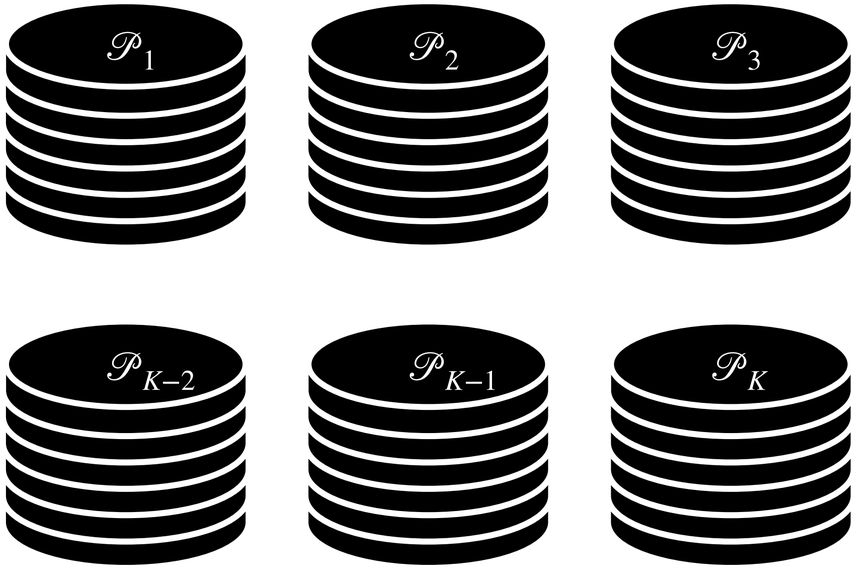


$$\min_{w \in \mathbb{R}^d} f(w) \quad \text{where} \quad f(w) = \frac{1}{N} \sum_{i=1}^{N} f_i(w), \quad f_i(w) = \ell(D_i;\, w)$$



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

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

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

FedSGD and FedAVG

FedSGD

$$w_{t+1} \leftarrow w_t - \eta \sum_{k=1}^{K} \frac{n_k}{n} g_k$$

FedAVG

$$w^k \leftarrow w^k - \eta g_k \qquad\qquad w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$$

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