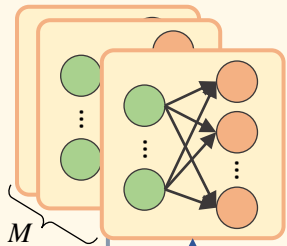


Stacking as a p -way tensor

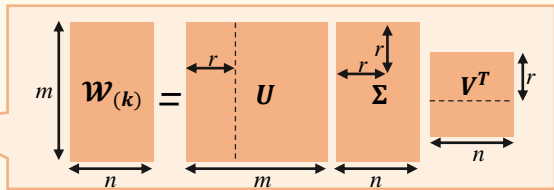
$$\mathcal{W} \in \mathbb{R}^{d_1 \times \dots \times d_p}$$

$d_p = M$



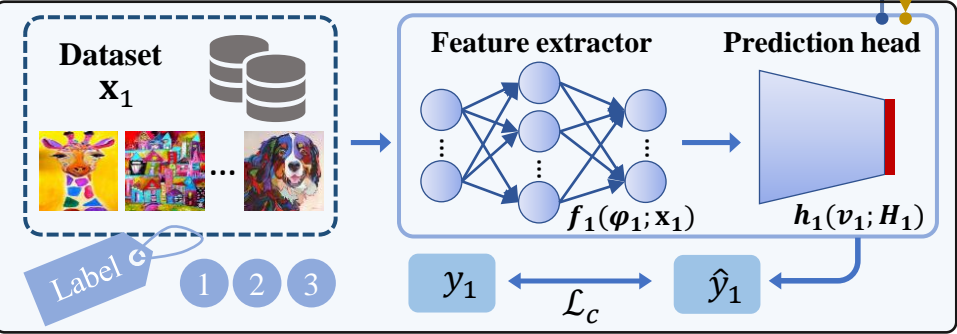
Central Server

$$\mathcal{L}_r = \sum_{l=1}^L \|\mathcal{W}^l\|_*, \quad \|\mathcal{W}\|_* = \sum_{k=1}^p \|\mathcal{W}_{(k)}\|_*$$

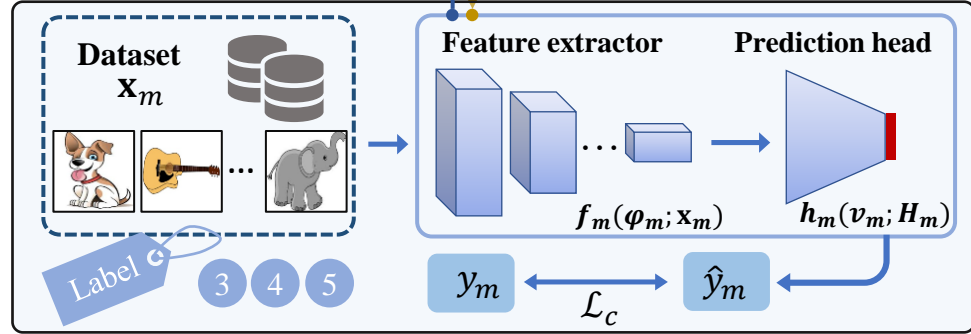


Upload: DH: $w = \varphi \circ v$, MH: $w = v$, TH: $w = \varphi$

Gradient Descent



Local Client 1



Local Client m