

Federated Multi-Task Learning

MOCHA: A Framework for Federated Multi-Task Learning

Federated Multi-Task Learning, NIPS, 2017

$$\min_{W, \Omega} \left[\sum_{k=1}^K \sum_{i \in \mathcal{P}_k} f_k(w_k^\top x_k^i, y_k^i) + \mathcal{R}(W, \Omega) \right]$$

Not jointly convex in W and Ω

$$\mathcal{R}(W, \Omega) = \lambda_1 \operatorname{tr}(W \Omega W^T) + \lambda_2 \|W\|_F^2$$

When fixing Ω , updating W depends on the data X .

When fixing W , optimizing for Ω only depends on W , not on X .









Federated Multi-Task Learning

MOCHA : A Framework for Federated Multi-Task Learning

$$\min_{W, \Omega} \left[\sum_{k=1}^K \sum_{i \in \mathcal{P}_k} f_k(w_k^\top x_k^i, y_k^i) + \mathcal{R}(W, \Omega) \right]$$
$$\mathcal{R}(W, \Omega) = \lambda_1 \text{tr}(W\Omega W^\top) + \lambda_2 \|W\|_F^2$$

- ✓ Not jointly convex in W and Ω
- ✓ When fixing Ω , updating W depends on the data X .
- ✓ When fixing W , optimizing for Ω only depends on W , not on X .

Federated Multi-Task Learning

Federated Update of W

Dual Problem $\min_{\alpha \in \mathbb{R}^N} \left[D(\alpha) := \sum_{k=1}^K \sum_{i \in \mathcal{P}_k} f_k^*(-\alpha_k^i) + \mathcal{R}^*(X\alpha) \right] \quad X = \text{diag}(X_1, \dots, X_K)$

Data-local subproblem

$$\min_{\Delta \alpha_k \in \mathbb{R}^N} G_k^{\sigma'}(\Delta \alpha_k; v_k, \alpha_k) := \sum_{i \in \mathcal{P}_k} f_k^*(-\alpha_k^i - \Delta \alpha_k^i) + \langle w_k(\alpha), X_k \Delta \alpha_k \rangle + \frac{\alpha}{2} \|X_k \Delta \alpha_k\|_{M_k}^2 + c(\alpha)$$