Tackling Data Heterogeneity in Federated Learning via Loss Decomposition (Supplementary Material)

Algorithm 1 FedLD

Input: number of clients selected in each round m, number of communication rounds T, number of local epochs E, local batch size B, learning rate η, regularization hyperparameter λ, number of eigenvectors selected L, number of local samples for each client n_i and $n = \sum_{i=1}^{m} n_i$

Output: global model w_T

Server executes: Initialize w_0

for *each round* $t = 1, \ldots, T$ **do**

  $S_t \leftarrow$ Random selected m local clients

  Receive local gradients $G = [\boldsymbol{g}_1, \cdots, \boldsymbol{g}_m]$ from S_t

  $\hat{\boldsymbol{g}} = \frac{1}{m} \sum_{i=1}^{m} \boldsymbol{g}_i$

  for $z = 1, \cdots, m$ **do**

    $\lambda_z, \boldsymbol{e}_z = \mathrm{SVD}_z\left(\frac{1}{m}\boldsymbol{G}^\top \boldsymbol{G}\right)$ \\ refer to Eq.(4)

    $\boldsymbol{v}_z = \boldsymbol{G} e_z$ \\ refer to Eq.(5)

    $\bar{\boldsymbol{v}}_z = \boldsymbol{v}_z, \text{ if } \langle \boldsymbol{v}_z, \hat{\boldsymbol{g}} \rangle \geq 0; -\boldsymbol{v}_z, \text{ otherwise}$ \\ refer to Eq.(6)

  end

  Select the principal gradients with the top L largest eigenvalues, $\{\bar{\boldsymbol{v}}_1, \cdots, \bar{\boldsymbol{v}}_L\}$

  for $i = 1, \cdots, m$ **do**

    for $l = 1, \cdots, L$ **do**

      $\boldsymbol{g}'_{i,l} = \frac{\boldsymbol{g}_i \bar{\boldsymbol{v}}_l}{\|\bar{\boldsymbol{v}}_l\| \|\bar{\boldsymbol{v}}_l\|} \bar{\boldsymbol{v}}_l.$

    end

    $\boldsymbol{g}_i^{revise} = \sum_{l=1}^{L} \frac{\|\boldsymbol{g}_i\|}{\|\boldsymbol{g}'_{i,l}\|} \frac{\lambda_l}{\|\lambda_l\|} \boldsymbol{g}'_{i,l}$ \\ refer to Eq.(7)

  end

  $\bar{\boldsymbol{g}} = \sum_{i=1}^{m} \frac{n_i}{n} \boldsymbol{g}_i^{revise}$

  Update the global model $\boldsymbol{w}_t \leftarrow \boldsymbol{w}_{t-1}$ with $\bar{\boldsymbol{g}}$

end

Clients update($i, \boldsymbol{w}_t$):

$\boldsymbol{w}_t^i \leftarrow \boldsymbol{w}_t$

$\mathcal{B} \leftarrow$ (split local dataset of client i into batches of size B)

for *local epoch* $e = 1, \cdots, E$ **do**

  for *batch* $b = \{x, y\} \in \mathcal{B}$ **do**

    $l = l_{CE}\left(y, f_{\boldsymbol{w}_t^i}(x)\right) + \lambda \log\left(1 + \left\| f_{\boldsymbol{w}_t^i}(x) \right\|_2^2\right)$

    $w_t^i \leftarrow w_t^i - \eta \nabla \ell$

  end

end

return w_t^i to server
