Byzantine Resilient and Fast Federated Few-Shot Learning

Ankit Pratap Singh [1] Namrata Vaswani [1]

[1]Iowa State University

Problem Setting

Multi-task learning addresses statistical challenges in the federated setting by learning separate models for each device. We let the representation function class be Low-Dimensional Linear Representations i.e., $\{\boldsymbol{x} \mapsto \boldsymbol{U}^T\boldsymbol{x} | \boldsymbol{U} \in \Re^{n\times r}\}$ [Du et al. 2020]. The goal is to find the optimal representation $\boldsymbol{U}^*$ in a federated setting, resilient to **Byzantine Attacks**.

$$\boldsymbol{Y}_{m\times q} = [(\boldsymbol{X}_1)_{m\times n}(\boldsymbol{\theta}_1^*)_{n\times 1}, ..., (\boldsymbol{X}_q)_{m\times n}(\boldsymbol{\theta}_q^*)_{n\times 1}] = [(\boldsymbol{X}_1)_{m\times n}\boldsymbol{U}^*_{n\times r}(\boldsymbol{b}_1^*)_{r\times 1}, ..., (\boldsymbol{X}_q)_{m\times n}\boldsymbol{U}^*_{n\times r}(\boldsymbol{b}_q^*)_{r\times 1}]$$

Solving this problem requires solving (AltGDmin [Nayer and Vaswani 2022] and FedRep [Collins et al. 2021]),

$$\min_{\substack{\tilde{\boldsymbol{U}}\in\Re^{n\times r}\\ \tilde{\boldsymbol{B}}\in\Re^{r\times q}}} f(\tilde{\boldsymbol{U}}, \tilde{\boldsymbol{B}}) = \min_{\substack{\tilde{\boldsymbol{U}}\in\Re^{n\times r}\\ \tilde{\boldsymbol{B}}\in\Re^{r\times q}}} \sum_{k=1}^{q} \|\boldsymbol{y}_k - \boldsymbol{X}_k\tilde{\boldsymbol{U}}\tilde{\boldsymbol{b}}_k\|^2$$

Federated Setting

- In the federated setting, we assume that there are a total of L nodes. Each observes a different disjoint subset ($\widetilde{m} = m/L$) of rows of $\boldsymbol{Y}$. At most τL nodes can be Byzantine with $\tau < 0.4$. The nodes can only communicate with the center.
- **Byzantine attack** is a "model update poisoning" attack where it can design the worst possible attacks at each algorithm iteration.

Theorem 1: Byz-Fed-AltGDmin-Learn

Assume Right Singular Vectors' Incoherence for Θ^* If

$$\frac{m}{L}q \geq C\kappa^4\mu^2(n+q)r^2\log(1/\epsilon)$$

then, w.p. at least $1 - TLn^{-10}$,

$$\boldsymbol{SD}_F(\boldsymbol{U}^*, \boldsymbol{U}_T) \leq \epsilon$$

and $\|(\boldsymbol{\theta}_k)_\ell - \boldsymbol{\theta}_k^*\| \leq \epsilon\|\boldsymbol{\theta}_k^*\|$ for all $k \in [q]$, $\ell \in [L]$. The communication cost per node is order $nr\log(\frac{n}{\epsilon})$. The computational cost at any node is order $nqr\log(\frac{n}{\epsilon})$ while that at the center it is $n^2L\log^3(Lr/\epsilon)$.

Subspace-Median

In solving this problem, we also introduce **Subspace Median**, a novel, secure solution to the federated subspace learning meta-problem that occurs in many different applications e.g., Federated PCA.

Estimate principal subspace $span(\boldsymbol{U}^*)$ of an unknown $n\times n$ symmetric matrix $\boldsymbol{\Phi}^*$ in a federated setting while being resilient to **Byzantine Attacks**.

$$\boldsymbol{D}_{n\times q} = [(\boldsymbol{D}_1)_{n\times q_1}, ..., (\boldsymbol{D}_\ell)_{n\times q_\ell}, ..., (\boldsymbol{D}_L)_{n\times q_L}]$$

- $\boldsymbol{U}^*$ is an $n\times r$ matrix denoting the top r eigenvectors of $\boldsymbol{\Phi}^*$
- **Federated Setting:** Each node $\ell \in [L]$ observes a data matrix $\boldsymbol{D}_\ell$, that allows it to estimate $\boldsymbol{\Phi}^*$ and subsequently $\boldsymbol{U}^*$.

Byz-Fed-AltGDmin-Learn: Complete algorithm

Nodes $\ell = 1, ..., L$
Compute $(\boldsymbol{U}_0)_\ell$ which is the matrix of top r left singular vectors of $(\boldsymbol{\Theta}_0)_\ell := \Sigma_{k=1}^q (\boldsymbol{X}_k)_\ell{}^\top((\boldsymbol{y}_k)_\ell)_{\text{trunc}}\boldsymbol{e}_k^\top$
Key Idea 1: Subspace Median on $(\boldsymbol{U}_0)_\ell$'s
Central Server: Subspace Median
Orthonormalize: $\boldsymbol{U}_\ell \leftarrow QR((U_\ell)_0)$, $\ell \in [L]$
Compute $\mathcal{P}_{\boldsymbol{U}_\ell} \leftarrow \boldsymbol{U}_\ell\boldsymbol{U}_\ell^\top$, $\ell \in [L]$
Compute GM: $\mathcal{P}_{gm} \leftarrow \text{GeometricMedian}\{\mathcal{P}_{\boldsymbol{U}_\ell}, \ell \in [L]\}$
Find $\ell_{best} = \arg\min_\ell \|\mathcal{P}_{\boldsymbol{U}_\ell} - \mathcal{P}_{gm}\|_F$
Output $\boldsymbol{U}_0 = \boldsymbol{U}_{out} = \boldsymbol{U}_{\ell_{best}}$
for $t = 1$ to T **do**
  Nodes $\ell = 1, ..., L$
  Set $\boldsymbol{U} \leftarrow \boldsymbol{U}_{t-1}$
  With $\boldsymbol{U}$ fixed, Least-Squares step over $(\boldsymbol{b}_k)_\ell$ for all k
  With $\boldsymbol{B}$ fixed, Gradient of $f(\boldsymbol{U}, \boldsymbol{B})$ w.r.t. $\boldsymbol{U}$: ∇f_ℓ
  Central Server
  Key Idea 2: Calculate GM of $\nabla f_\ell'$s
  $\nabla f^{GM} \leftarrow \text{GeometricMedian}(\nabla f_\ell, \ell = 1, 2, \ldots L)$.
  Compute $\boldsymbol{U}^+ \leftarrow QR(\boldsymbol{U}_{t-1} - \frac{\eta}{\rho\widetilde{m}}\nabla f^{GM})$
  return Set $\boldsymbol{U}_t \leftarrow \boldsymbol{U}^+$. Push $\boldsymbol{U}_t$ to nodes.
end for

Resilient Federated PCA Experiment

Attacks	SubsMed (Proposed)	ResPowMeth	PowMeth (No Attack)
Alternating	0.091	0.898	0.050
Ones	0.091	0.952	0.050
Orthogonal	0.091	0.208	0.050

Table 1. $n = 1000$, $L = 3$, $L_{byz} = 1$, $r = 60$, $\tilde{q} = 600$, rank-$(r+1)$

- Resilient Power Method (ResPowMeth): GM based modification of the power method.
- Baseline Power Method for a no-attack setting (PowMeth).

References

- Nayer, S. and N. Vaswani (2022). "Fast and sample-efficient federated low rank matrix recovery from column-wise linear and quadratic projections". In: *IEEE TIT*.
- Collins, L., H. Hassani, A. Mokhtari, and S. Shakkottai (2021). "Exploiting shared representations for personalized federated learning". In: *ICML*.
- Du, S. S., W. Hu, S. M. Kakade, J. D. Lee, and Q. Lei (2020). "Few-shot learning via learning the representation, provably". In: *arXiv preprint arXiv:2002.09434*.