Accelerating Heterogeneous Federated Learning with Closed-form Classifiers

Eros Fanì, Raffaello Camoriano, Barbara Caputo, Marco Ciccone

Presenting: **Eros Fanì** - eros.fani@polito.it
Polytechnic University of Turin (PoliTO), Italy

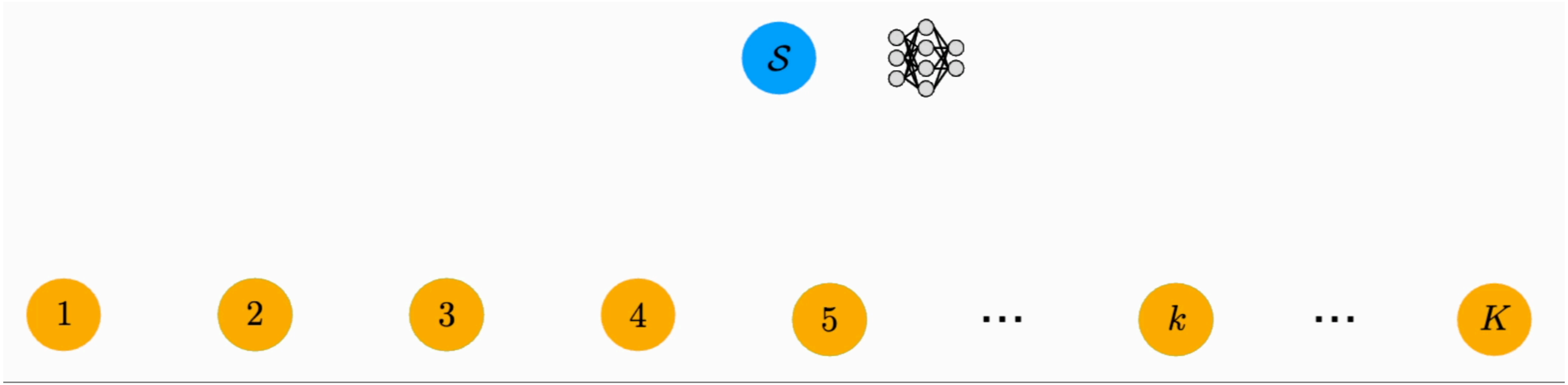
Forty-first International Conference on Machine Learning (ICML24), Wien, Austria, 2024

SCAN ME

Context and motivation

Federated Learning - Introduction

Novel distributed machine learning paradigm enabling model training using data from multiple devices without direct access, preserving users' privacy



Context and motivation

ICML
International Conference
On Machine Learning

Statistical Heterogeneity and Client Drift

Key issue:
Statistical Heterogeneity

Non-IID client distributions & partial participation:

→ **slow and unstable training**
→ loss in final performance and increased communication

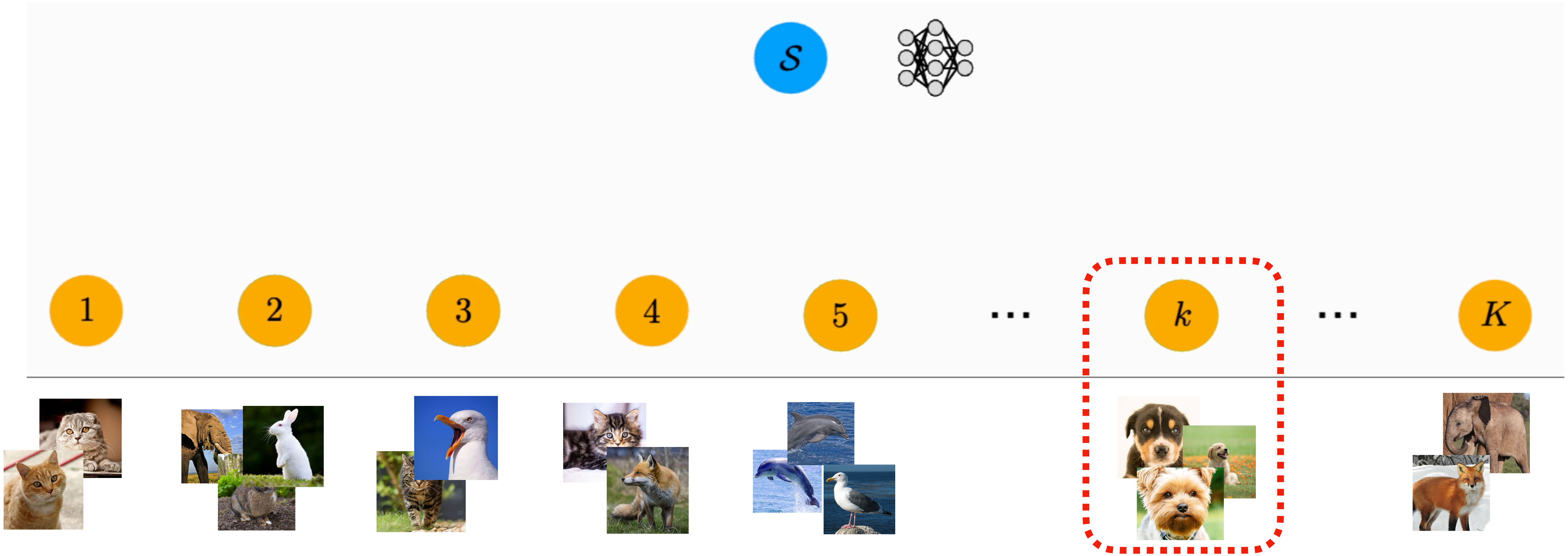
→

Client Drift
(Karimireddy, 2020)

Local updates **diverge** from global optimum

Karimireddy, S. P. et al. “Scaffold: Stochastic controlled averaging for federated learning” ICML 2020
Li, Z. et al. “No fear of classifier biases: Neural collapse inspired federated learning with synthetic and fixed classifier” ICCV 2023

Data Recency Bias in the classifier



Similarly to Continual Learning (Wang 2022), the classifier is prone to forgetting because of data recency bias (Luo 2021, Li 2023)

Wang, R. et al. “Schedule-robust online continual learning”
Luo, Mi et al. “No fear of heterogeneity: Classifier calibration for federated learning with non-iid data” NeurIPS 2021
Li, Z. et al. “No fear of classifier biases: Neural collapse inspired federated learning with synthetic and fixed classifier” ICCV 2023

Context and motivation

Our objective

Can we design an efficient FL method
***robust to client drift** in heterogeneous settings*
*and **unaffected by classifier bias?***

YES!

Contribution

- **Fed3R** - efficient and robust FL algorithm:
  - **Immune to statistical heterogeneity**
  - Faster convergence
  - Reduced computations and communication
- **Fed3R-RF:** non-linear variant based on random features
- **Fed3R+FT:** last layer initialization for faster fine-tuning

a) **Fed3R+FT**: fine-tune the whole model
b) **Fed3R+FTlp**: fine-tune only the classifier
c) **Fed3R+FTfeat**: fine-tune only the feature extractor

Method

Background - (centralized) Ridge Regression (RR)

$$W^* = \underset{W \in \mathbb{R}^{d \times C}}{\arg\min} \|Y - \varphi(X)W\|^2 + \lambda \|W\|^2$$

Linear predictor: $f(x; W) = W^\top \varphi(x)$

Closed-form solution: $W^* = (A + \lambda I_d)^{-1} b$ ★

$$A := \varphi(X)^T \varphi(X), \; b := \varphi(X)^T Y$$

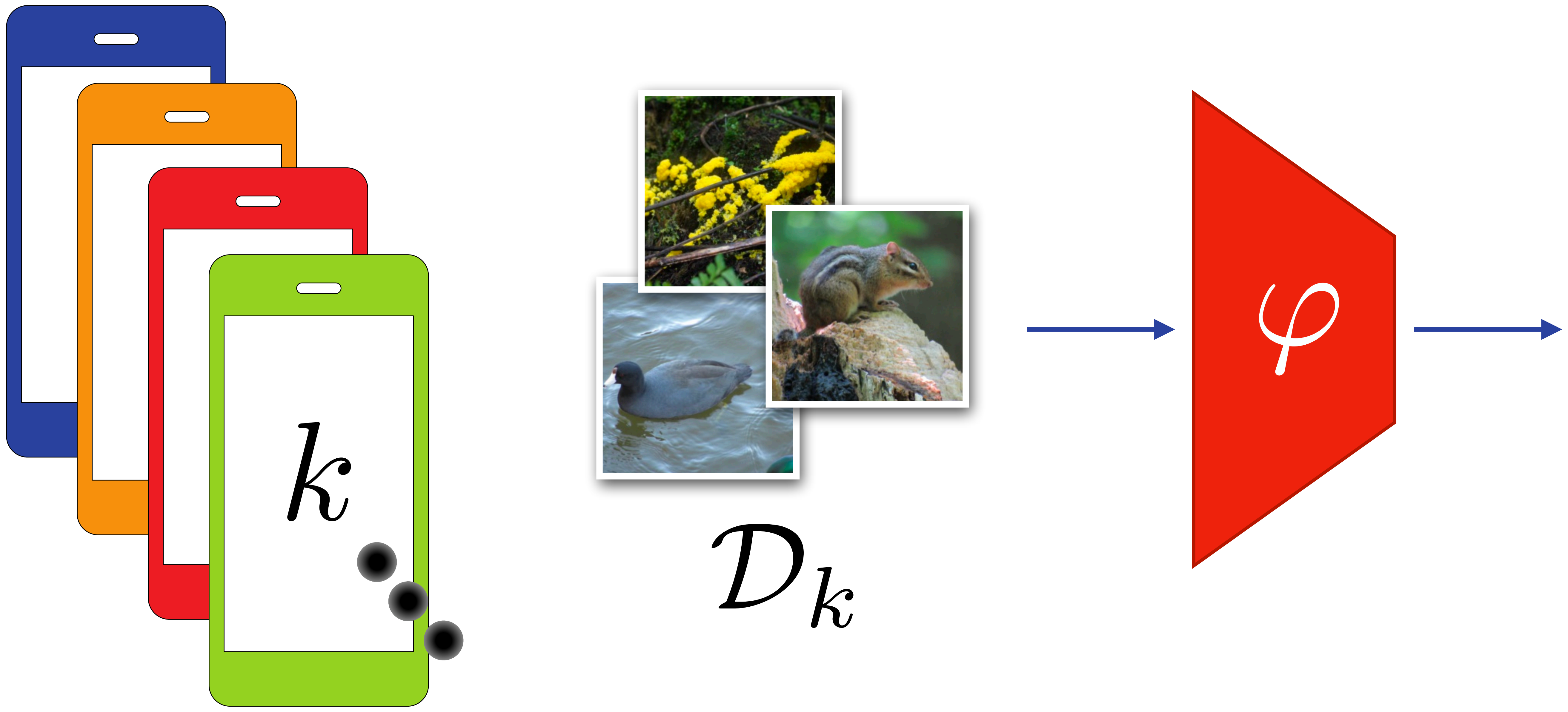
RR can also be used for the classification task (Rifkin, 2003)

Rifkin, R., Yeo, G., Poggio, T., et al. Regularized least squares classification. Nato Science Series Sub Series III Computer and Systems Sciences, 2003.

Method

Fed3R: Federated Recursive Ridge Regression

Step 1 (client side): Local computations



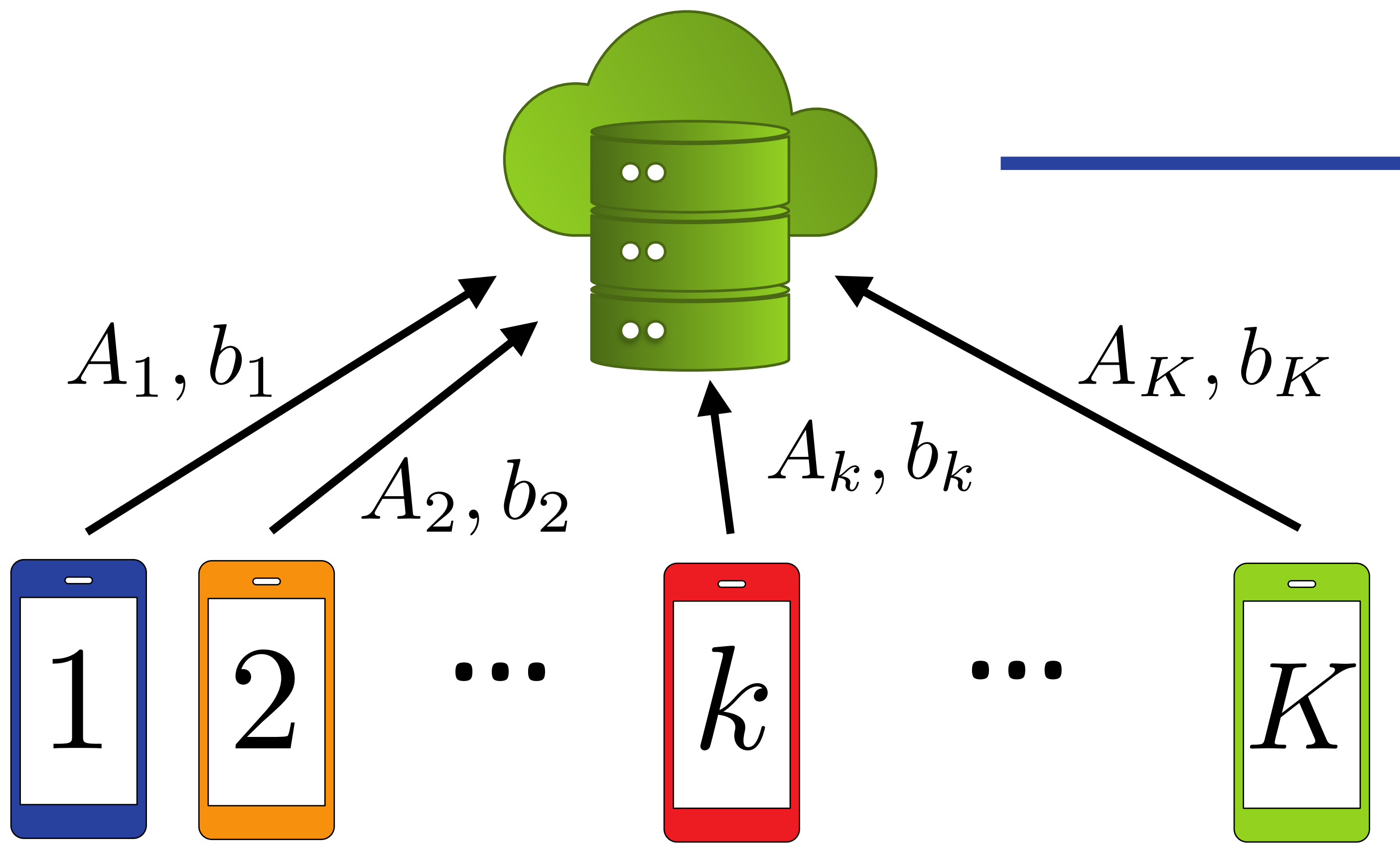
$$\sum_{(x,y)\in\mathcal{D}_k} \varphi(x)\varphi(x)^\top = A_k$$

$$\sum_{(x,y)\in\mathcal{D}_k} \varphi(x)e_y^\top = b_k$$

Method

Fed3R: Federated Recursive Ridge Regression

Step 2 (server side): Exact aggregation



Compute the aggregate statistics

$$A = \sum_{(x,y)\in\mathcal{D}} \varphi(x)\varphi(x)^\top = \sum_{k\in\mathcal{K}} \sum_{(x,y)\in\mathcal{D}_k} \varphi(x)\varphi(x)^\top = \sum_{k\in\mathcal{K}} A_k$$

$$b = \sum_{(x,y)\in\mathcal{D}} \varphi(x)e_y^\top = \sum_{k\in\mathcal{K}} \sum_{(x,y)\in\mathcal{D}_k} \varphi(x)e_y^\top = \sum_{k\in\mathcal{K}} b_k$$

Closed-form RR solution: $W^* = (A + \lambda I_d)^{-1} b$ ★

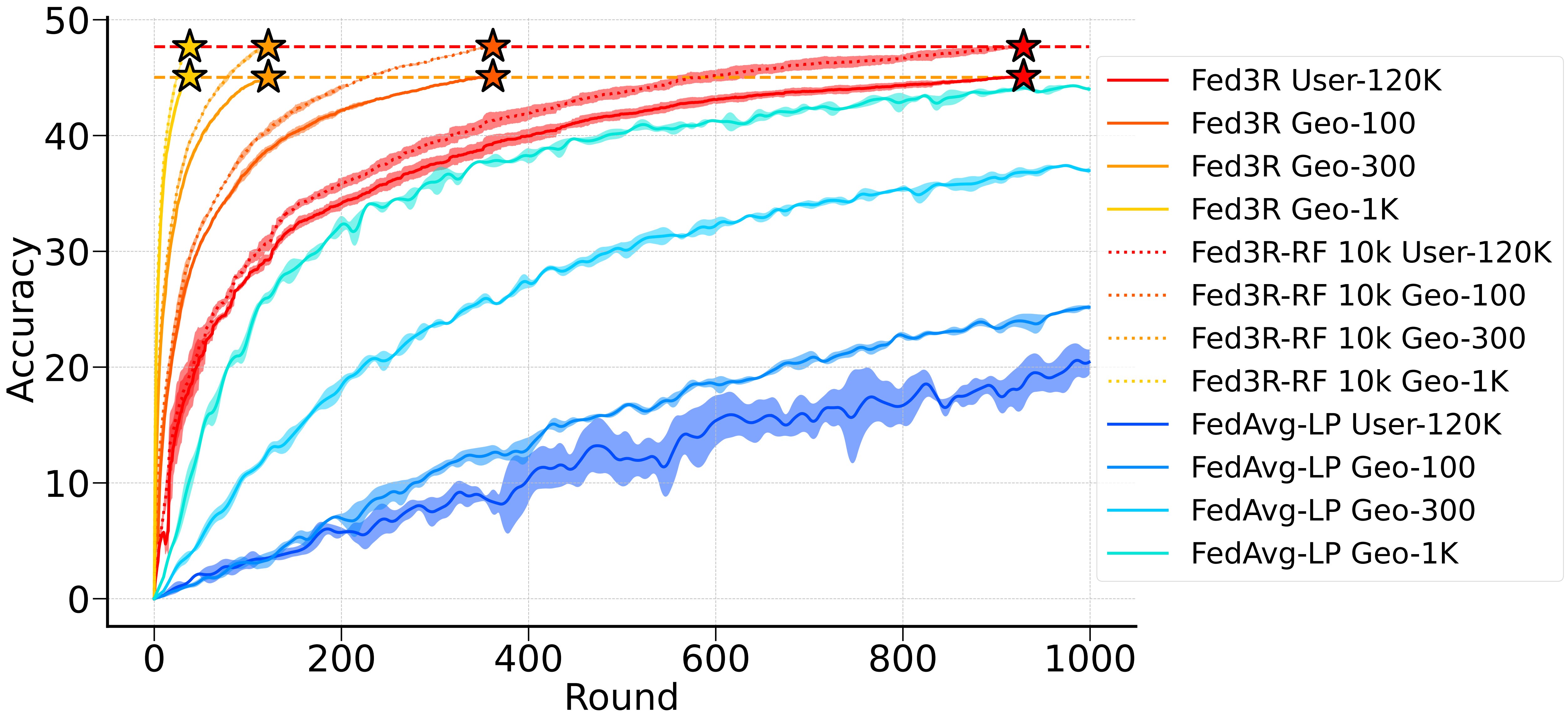
Use Eq. ★ to obtain the exact aggregate classifier W^*

Method

Fed3R & Fed3R-RF properties

- Immune to heterogeneity: equivalence to exact centralized solution
- Convergence guaranteed in a single pass on clients
- Inherits generalization properties of RR (Caponetto, 2007)
- Memory and computation efficient

	User-120K	Geo-100	Geo-300	Geo-1K
# clients	9275	3606	1208	368
# classes	1203	1203	1203	1203



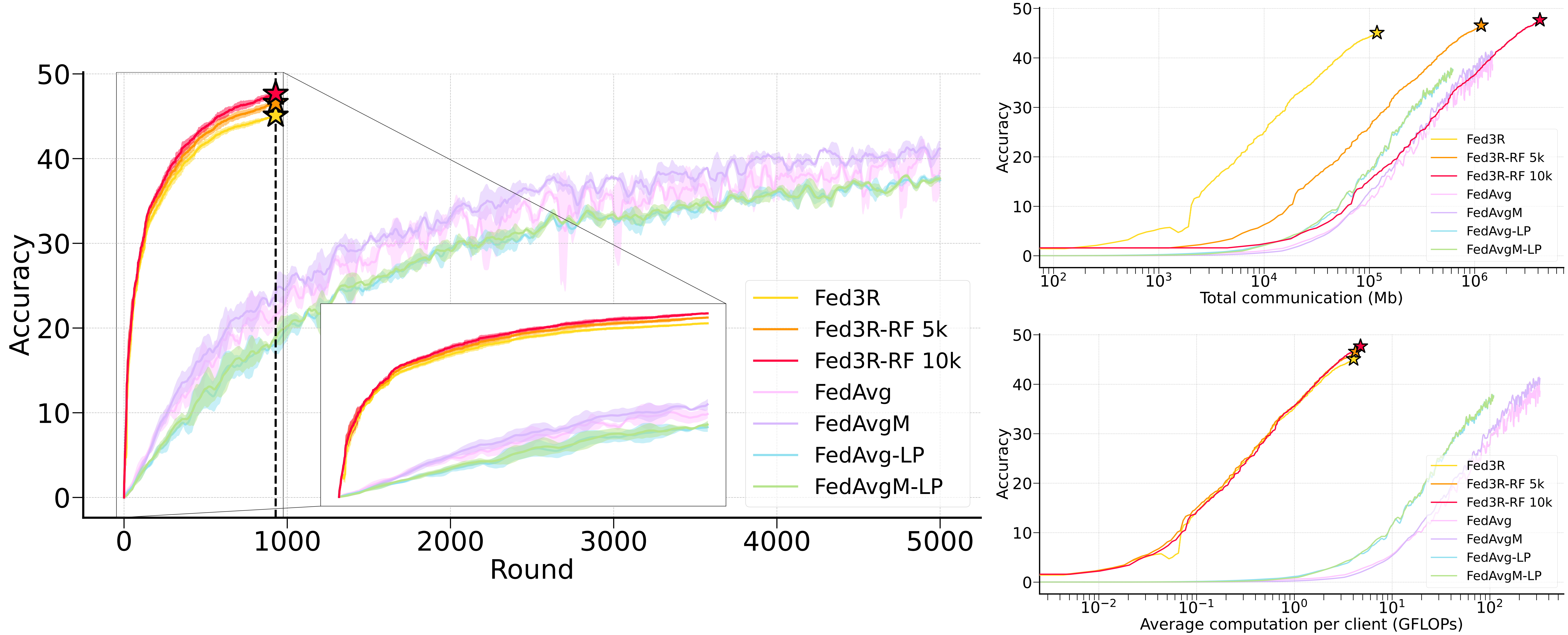
Fed3R performance is invariant to different federated splits of **iNaturalist** (Hsu, 2020).

Caponnetto, A. et al. "Optimal rates for the regularized least-squares algorithm" Foundations of Computational Mathematics 2007

Results

Fed3R & Fed3R-RF vs. baselines

on iNaturalist-Users120K - 9275 clients; 1203 classes

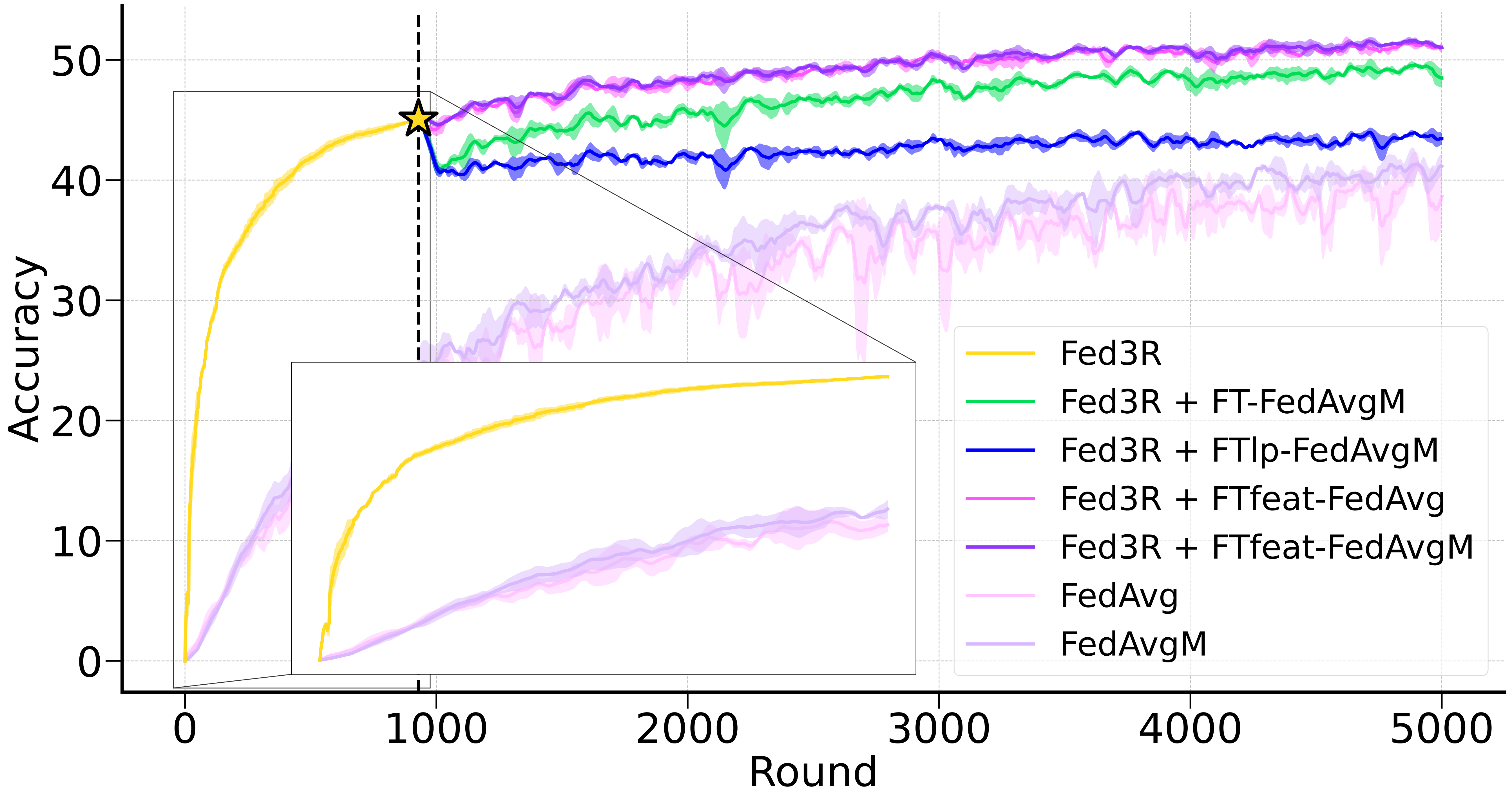


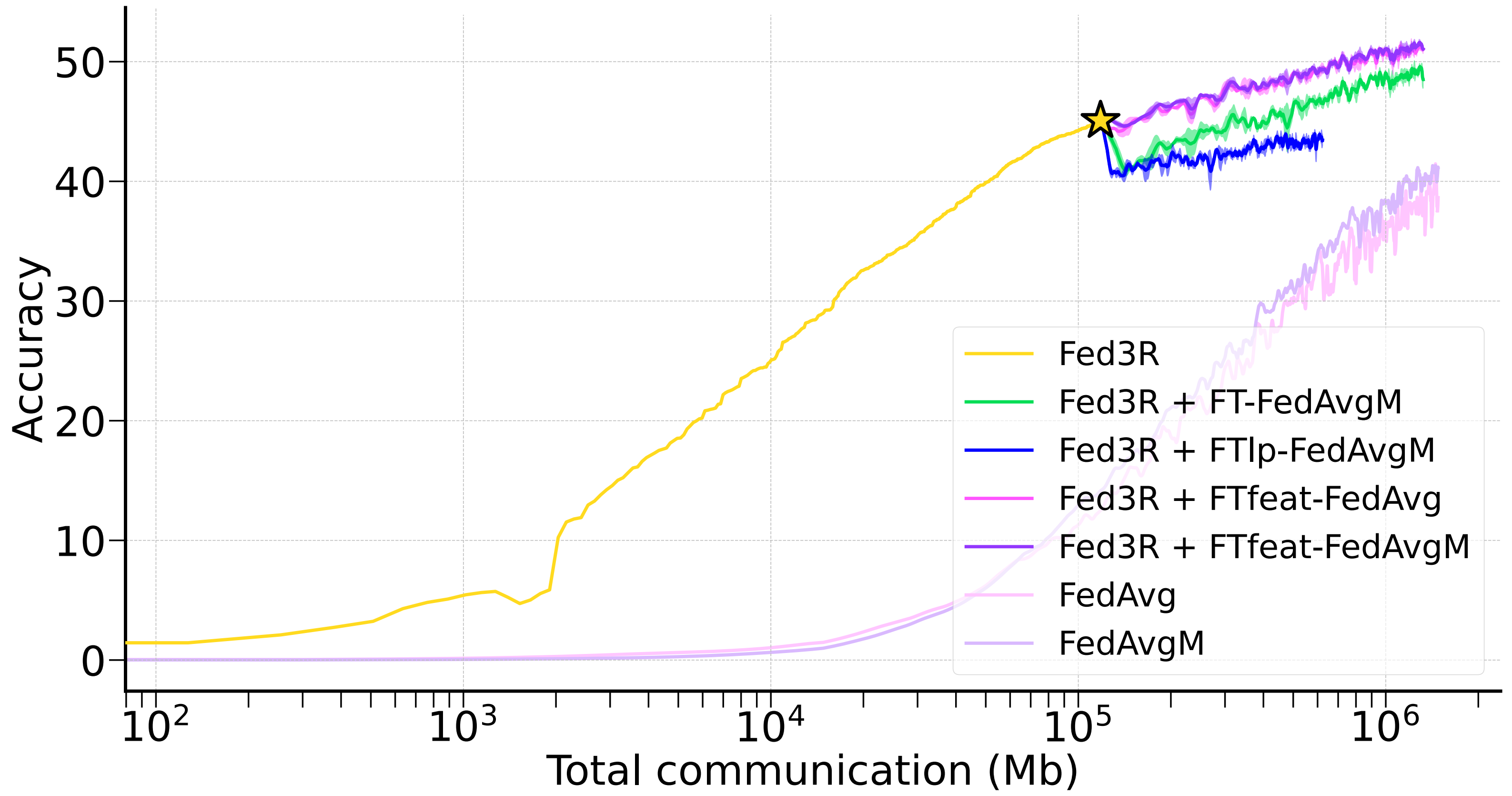
Results

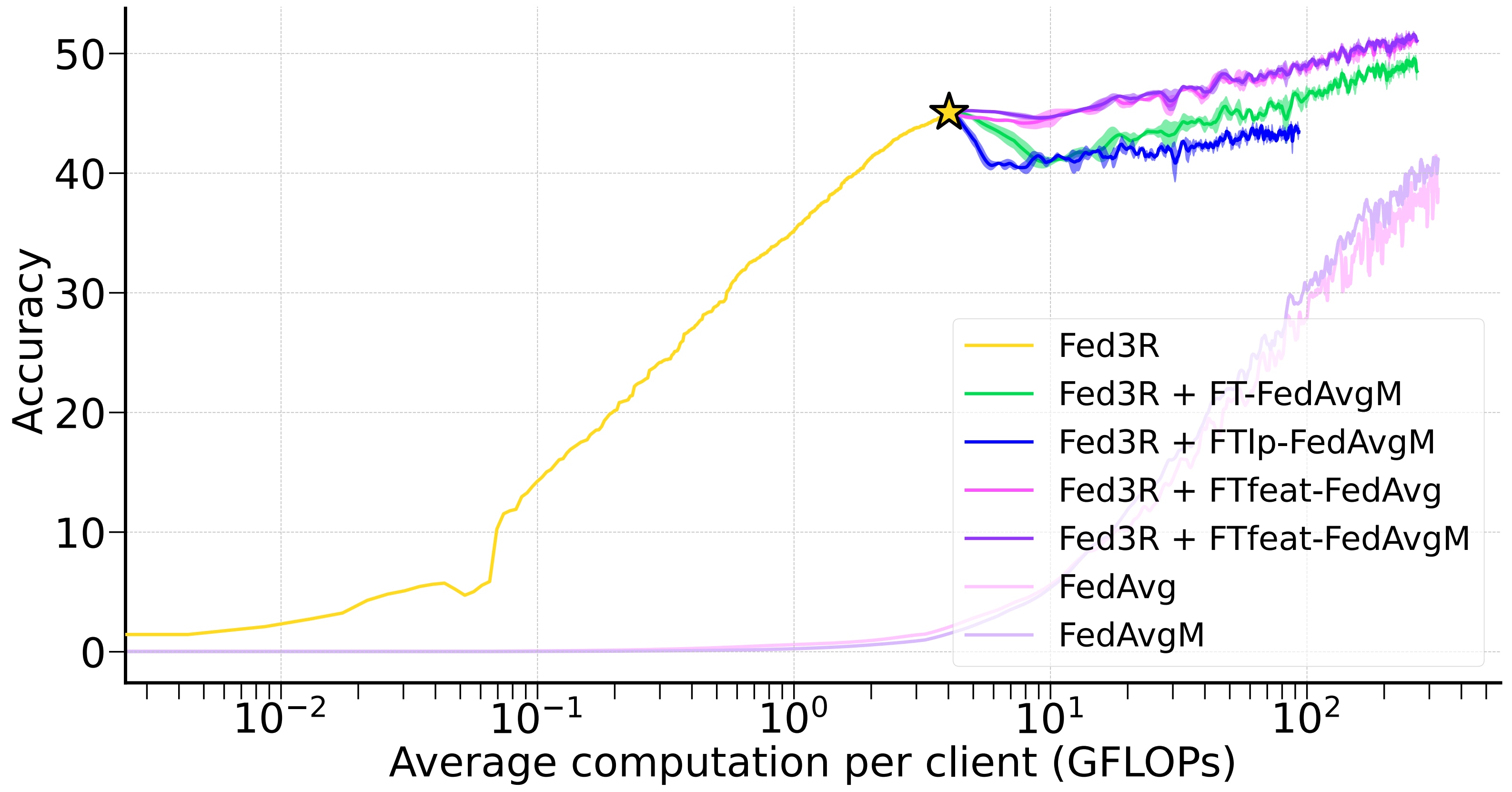
ICML International Conference On Machine Learning

Fed3R+FT vs. baselines

on iNaturalist-Users120K - 9275 clients; 1203 classes







Thank you for your attention!

Join us at the Poster Session 5, 11:30 a.m. — 1 p.m. CEST, Hall C 4-9 #2507

Exhibition Congress Center, Wien (AU), July 25, 2024

Eros Fanì - eros.fani@polito.it
Polytechnic University of Turin (**PoliTO**), Italy

Accelerating Heterogeneous Federated Learning with Closed-form Classifiers

Eros Fanì, Raffaello Camoriano, Barbara Caputo, Marco Ciccone

SCAN ME