**Our Contributions**

1. **Theory**: Analysis of 2 popular optimization algos

2. **Extensive experiments**: text, vision, and speech settings

**Personalized Federated Learning**

Model on client $i = (u, v_i)$

$$\text{Objective: } \min_{u, v_1, \ldots, v_n} \frac{1}{n} \sum_{i=1}^{n} F_i(u, v_i)$$

- $u$: shared parameters
- $v_i$: personal parameters

**Optimization Algorithms**

1. Client sampling + model broadcast
2. Local updates
3. Aggregate updates

**Theorem**

For smooth, nonconvex functions and client sampling, we have the rates:

- **Alternating update**: $\frac{\sigma^2_{v_i}}{\sqrt{t}}$
- **Simultaneous update**: $\frac{\sigma^2_{u}}{\sqrt{t}}$

where $\sigma^2_{v_i} < \sigma^2_{u}$ under typical scenarios

**Key technical challenge**: Dependent random variables due to random sampling of clients

**Methodology**: virtual full participation

**Experimentally**, small but consistent trend of alternating > simultaneous

**Best personalization architecture** depends on nature of data heterogeneity

**Personalization Architectures**

- **Personalized output layer**
- **Personalized input layer**
- **Personalized adapters**

**2. Experiments**

- **Top-1 Error**
- **Word Error Rate**
- **Misclassification Error**

Choose your personalization architecture wisely!

**Code**

krishnap25
KrishnaPillutla
krishnap25.github.io

**References**

- Collins et al. ICML (2021)
- Singhal et al. NeurIPS (2021)
- Arivazhagan et al. (2019)
- Liang et al. (2019)