When MAML Can Adapt Fast
And How to Assist When it Cannot

Séb Arnold, Shariq Iqbal, Fei Sha
Meta-Learning with MAML

Objective:

$$
\min_{\theta} \mathbb{E}_\tau [\mathcal{L}_\tau (\theta - \alpha \nabla_\theta \mathcal{L}_\tau (\theta))]
$$

where:

- $\theta \triangleq$ parameters to learn,
- $\tau \triangleq$ task index, and
- $\mathcal{L}_\tau \triangleq$ the task-specific loss.
• Tasks: linearly separable

• Models: linear (shallow vs deep)
Insights

• **Theoretical** analysis:
  
  • Deep models are **required** for meta-learning.
  
  • Some parameters act like **implicit meta-optimizers**.

• **Empirical** analysis confirms on linear and deep models.
Solutions

1. Use deeper and larger models.

2. Just add a few linear layers.

3. Move parameters to KFO, our new meta-optimizer.
Solutions

1. Use deeper and larger models.

2. Just add a few linear layers.

3. Move parameters to KFO, our new meta-optimizer.
Solutions

1. Use deeper and larger models.

2. Just add a few linear layers.

3. Move parameters to KFO, our new meta-optimizer.
Results

- Extra linear layers improve shallow and deep meta-learning.
- Meta-optimizers (KFO) always perform best.
- Meta-optimizers are most beneficial with shallow models.
- + a few more…
Thank You

• Learn more:
  
  • Poster: ID 108 — Session 4: April 14 at 12:45-14:45 PDT
  
  • Web: sebarnold.net/projects/kfo
  
  • Code: github.com/Sha-Lab/kfo
  
  • Email: seb.arnold@usc.edu