Introduction

Unimodal Task Distribution

Multimodal Task Distribution

Real-world task distributions are often multimodal

- Have a rich structure (e.g., multiple modes)
- Some knowledge can be transferable across modes/tasks

Model-agnostic meta-learning (MAML) [1]

- Seek a common initialization parameter for all the modes

An ensemble of MAMLs (Multi-MAML)

- Mode labels are often not available
- Prevent sharing related knowledge among modes/tasks

Background

Model-Agnostic Meta-Learning [1]

- Meta-learn a parameter initialization that can be fine-tuned for new tasks in few gradient update steps

Model-Agnostic Meta-Learning Objective

- Inner loop: \[ \theta_{t} = \theta - \alpha \nabla_{\theta} \mathcal{L}_{t}(f(x, \theta); \mathcal{D}_{t}^{\text{train}}) \]
- Outer loop: \[ \theta' = \theta - \beta \nabla_{\theta} \sum_{t=1}^{T} \mathcal{L}_{t}(f(x, \theta_{t}); \mathcal{D}_{t}^{\text{val}}) \]

Our Approach

Intuition

- Modulation network: identify task modes and modulate the initialization accordingly
- Task network: further gradient adaptation via MAML steps

Experiment - Regression

(a) M3MAML, post modulation vs. other prior models

(b) MMAMLL, post adaptation vs. other posterior models

Experiment - Classification

Experiment - Reinforcement Learning

Experiment - Learned Task Embeddings