Model Agnostic Meta Learning

Chelsea Finn, Pieter Abbeel, Sergey Levine

Presented by Quan Zhang

Finn, Chelsea, Pieter Abbeel, and Sergey Levine. "Model-agnostic meta-learning for fast adaptation of deep networks." 2017.

Introduction

- A good machine learning model often requires training with a large number of samples.
- Humans, in contrast, learn new concepts and skills much faster and more efficiently.
- Kids who have seen cats and birds only a few times can quickly tell them apart.
- People who know how to ride a bike are likely to discover the way to ride a motorcycle fast with little or even no demonstration.
- Learning new concepts and skills fast with a few training examples?
- **Meta learning: learning to learn**

Question: How to avoid overfit?

Meta learning: learning to learn

- We expect a good meta-learning model capable of well adapting or generalizing to new tasks and new environments that have never been encountered during training time.
- The adaptation process has a limited exposure to the new task configurations (often during test)
- Eventually, the adapted model can complete new tasks
- This is learning to learn

Few shot learning: find a sweet spot

Training data to kids:



Impression: 1- to 1.5-foot long, furry, cute face, big eyes, pointy ears, whiskers, body like other mammals (four legs, one tail)

Few shot learning: find a sweet spot

Training data to kids:

Impression: 1- to 1.5-foot long, furry, cute face, big eyes, pointy ears, whiskers, body like other mammals (four legs, one tail)

What is this?

And these?

Human’s learning

- Have seen different kinds of animals
- Very small training data for each kind
- Fast adaptation
- No overfit at all

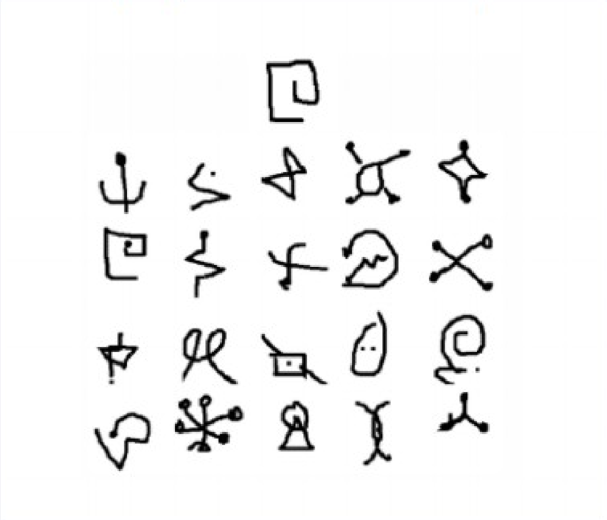
Meta learning

K-shot N-class classification task (often very few shots)

Model agnostic meta learning (MAML)

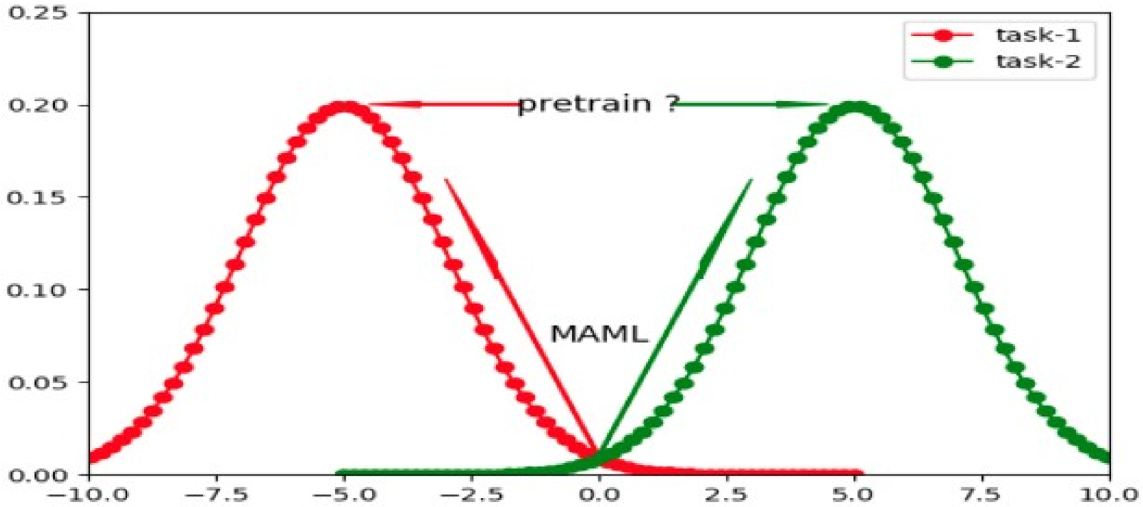
Few-Shot Learning Problem

1. Requires Meta-Learning to generalize well
2. Example problem: 1-shot, 20-way classification



- MAML: general to different problem (supervised learning, reinforcement learning…), agnostic to different model architectures (CNN, RNN…)

MAML In a Nutshell: finding a sweet spot for many tasks



MAML searches for a point maximizing improvement of k-step fine-tuning for all tasks

MAML algorithm

1. The meta-learner maintain the model parameter θ
2. The task searcher perform SGD to find better $\theta \rightarrow \theta'_i$ on each task $\mathcal{T}_i$
3. Update θ according to the **loss on** θ'_i summing over all taks

Algorithm 1 Model-Agnostic Meta-Learning

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters

```
randomly initialize θ
while not done do
    Sample batch of tasks T_i ~ p(T)
    for all T_i do
        Evaluate ∇_θ L_{T_i}(f_θ) with respect to K examples
        Compute adapted parameters with gradient descent: θ'_i = θ − α∇_θ L_{T_i}(f_θ)
    end for
    Update θ ← θ − β∇_θ Σ_{T_i~p(T)} L_{T_i}(f_{θ'_i})
end while
```

MAML to fit sine curves

1. Sin functions with different amplitude ranging $[0.1, 5.0]$
2. Each task (sin functions) has $K = 10$ samples
3. The task searcher adopt single step