Model-Agnostic Meta-Learning (MAML) for Fast Adaptation of Deep Networks

Spring 2022

Jishnu Jaykumar Padalunkal

Reading Group
Intelligent Robotics and Vision Lab
Erik Jonsson School of Engineering and Computer Science
University of Texas at Dallas

Jan 21, 2022

Table of Contents

Table of Contents

MAML

Model Agnostic Meta Learning[1]

- ICML 2017 Paper

[1] Chelsea Finn, Pieter Abbeel, and Sergey Levine. "Model agnostic meta learning for fast adaptation of deep networks". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1126–1135.

MAML

Model Agnostic Meta Learning[1]

- ICML 2017 Paper
- Authors
  - Chelsea Finn
  - Pieter Abbeel
  - Sergey Levin

[1] Chelsea Finn, Pieter Abbeel, and Sergey Levine. "Model agnostic meta learning for fast adaptation of deep networks". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1126–1135.

MAML

Model Agnostic Meta Learning[1]

- ICML 2017 Paper
- Authors
  - Chelsea Finn
  - Pieter Abbeel
  - Sergey Levin
- From
  - University of California, Berkeley
  - OpenAI

[1] Chelsea Finn, Pieter Abbeel, and Sergey Levine. "Model agnostic meta learning for fast adaptation of deep networks". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1126–1135.

MAML

Model Agnostic Meta Learning[1]

- ICML 2017 Paper
- Authors
  - Chelsea Finn
  - Pieter Abbeel
  - Sergey Levin
- From
  - University of California, Berkeley
  - OpenAI
- Proposes a **meta-learning** algorithm that is general and **model-agnostic**.

[1] Chelsea Finn, Pieter Abbeel, and Sergey Levine. "Model agnostic meta learning for fast adaptation of deep networks". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 1126–1135.

Intro. to Meta Learning

- **Meta learning** is a subfield of machine learning where
  - automatic learning algorithms are applied to metadata about machine learning experiments.[1]

[1] https://en.wikipedia.org/wiki/Meta_learning_(computer_science)

[2] https://bair.berkeley.edu/blog/2017/07/18/learning-to-learn/

Intro. to Meta Learning

- **Meta learning** is a subfield of machine learning where
  - automatic learning algorithms are applied to metadata about machine learning experiments.[1]
- As of 2017 the term had not found a standard interpretation, however the main goal is to use such metadata to understand how automatic learning can become flexible in solving learning problems, hence to improve the performance of existing learning algorithms or to learn (induce) the learning algorithm itself, hence the alternative term **learning to learn**.[1]

[1] https://en.wikipedia.org/wiki/Meta_learning_(computer_science)

[2] https://bair.berkeley.edu/blog/2017/07/18/learning-to-learn/

Intro. to Meta Learning

- **Meta learning** is a subfield of machine learning where
  - automatic learning algorithms are applied to metadata about machine learning experiments.[1]
- As of 2017 the term had not found a standard interpretation, however the main goal is to use such metadata to understand how automatic learning can become flexible in solving learning problems, hence to improve the performance of existing learning algorithms or to learn (induce) the learning algorithm itself, hence the alternative term **learning to learn**.[1]
- Some recent works include hyperparameter and neural network optimization, finding good network architectures, **few-shot image recognition**, and fast reinforcement learning.[2]

[1] https://en.wikipedia.org/wiki/Meta_learning_(computer_science)
[2] https://bair.berkeley.edu/blog/2017/07/18/learning-to-learn/

Meta Learning[1]

- Meta-learning, also known as “learning to learn”,
  - intends to design models that can learn new skills or adapt to new environments rapidly with a few training examples.

[1] https://lilianweng.github.io/lil-log/2018/11/30/meta-learning.html

Meta Learning[1]

- Meta-learning, also known as "learning to learn",
  - intends to design models that can learn new skills or adapt to new environments rapidly with a few training examples.
- There are three common approaches:
  - **Metric-based**: Learn an efficient distance metric
  - **Model-based**: Use (recurrent) network with external or internal memory
  - **Optimization-based**: Optimize the model parameters explicitly for fast learning. E.g. MAML

[1] https://lilianweng.github.io/lil-log/2018/11/30/meta-learning.html

Meta Learning

In meta-learning,

- the **model is trained by the meta-learner** to be able to learn on a large number of different tasks.
- the **goal of the trained model** is to **quickly learn a new task** from a **small amount of new data**.

Model Agnostic Nature

MAML begins with the phrase **Model Agnostic**.

Model Agnostic Nature

MAML begins with the phrase **Model Agnostic**.

Wrong interpretation

Can't be directly applied to all learning problems and models.

Model Agnostic Nature

MAML begins with the phrase **Model Agnostic**.

Wrong interpretation

Can't be directly applied to all learning problems and models.

Correct interpretation

Can be directly applied to any learning problem and model that is trained with a **gradient descent** procedure.

Table of Contents

Motivation

Humans

- Learning quickly is a hallmark of human intelligence, whether it involves recognizing objects from a few examples or quickly learning new skills after just minutes of experience.

Motivation

Humans

- Learning quickly is a hallmark of human intelligence, whether it involves recognizing objects from a few examples or quickly learning new skills after just minutes of experience.
- i.e. can act and adapt intelligently to a wide variety of new, unseen situations **quickly**.

Motivation

Humans

- Learning quickly is a hallmark of human intelligence, whether it involves recognizing objects from a few examples or quickly learning new skills after just minutes of experience.
- i.e. can act and adapt intelligently to a wide variety of new, unseen situations **quickly**.

Our artificial agents should be able to do the same

Our artificial agents should be able to do the same.

Learning and adapting quickly from only a few examples, and continuing to adapt as more data becomes available. **This kind of fast and flexible learning is challenging**.

- since the agent must integrate its prior experience with a small amount of new information, while avoiding overfitting to the new data.
- Moreover, the form of prior experience and new data will depend on the task.
- Hence, according to the authors, for the greatest applicability, the mechanism for learning to learn (or meta-learning) should be general to the task and the form of computation required to complete the task.

Our artificial agents should be able to do the same.

Learning and adapting quickly from only a few examples, and continuing to adapt as more data becomes available. **This kind of fast and flexible learning is challenging**.

- since the agent must integrate its prior experience with a small amount of new information, while avoiding overfitting to the new data.
- Moreover, the form of prior experience and new data will depend on the task.
- Hence, according to the authors, for the greatest applicability, the mechanism for learning to learn (or meta-learning) should be general to the task and the form of computation required to complete the task.

And finally MAML was born.

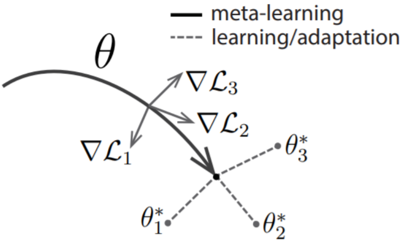
Table of Contents

General MAML

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   Note: the meta-update is using different set of data.
    Update θ ← θ − β∇_θ Σ_{T_i~p(T)} L_{T_i}(f_{θ'_i})
end while
```

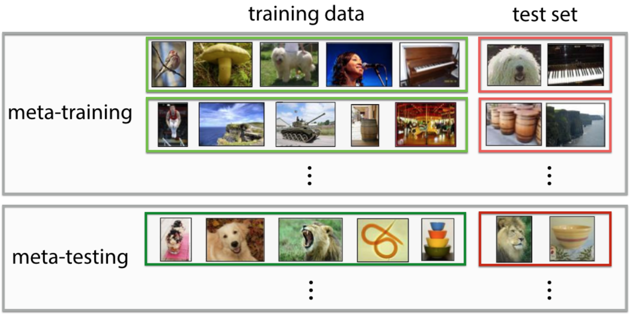


Sample Task (T) - Image Classification

- T_1: Cat/Dog, T_2: Car/Truck, T_3: Fruit/Vegetable, T_4: Digits, T_5: Alphabets, and many more...
- θ represents parameters of the meta-learner, say f_θ.
- Algorithm-1 image taken from

`https://lilianweng.github.io/lil-log/2018/11/30/meta-learning.html`

MAML for Few-Shot Supervised Learning



- N-way K-Shot problem (Image[2])
  - N=Num. of classes
  - K=Samples per class
  - Support (training data) and query (test set) set up for training and test data

[2] https://bair.berkeley.edu/blog/2017/07/18/learning-to-learn/

MAML for Few-Shot Supervised Learning

Algorithm 2 MAML for Few-Shot Supervised Learning

```
Require: p(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
        Sample K datapoints D = {x^(j), y^(j)} from T_i
        Evaluate ∇_θ L_{T_i}(f_θ) using D and L_{T_i} in Equation (2)
        or (3)
        Compute adapted parameters with gradient descent:
        θ'_i = θ − α∇_θ L_{T_i}(f_θ)
        Sample datapoints D'_i = {x^(j), y^(j)} from T_i for the
        meta-update
    end for
    Update θ ← θ − β∇_θ Σ_{T_i~p(T)} L_{T_i}(f_{θ'_i}) using each D'_i
    and L_{T_i} in Equation 2 or 3
end while
```

MAML for Few-Shot Supervised Learning

$$\min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\theta - \alpha \nabla_{\theta} \mathcal{L}(\theta, \mathcal{D}^{S}_{\mathcal{T}_i}), \mathcal{D}^{Q}_{\mathcal{T}_i}) = \min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\theta', \mathcal{D}^{Q}_{\mathcal{T}_i})$$

Figure: https://www.youtube.com/watch?v=ItPEBdD6VMk

MAML for Few-Shot Supervised Learning

f_θ Meta Model

$$\min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\theta - \alpha \nabla_\theta \mathcal{L}(\theta, \mathcal{D}^S_{\mathcal{T}_i}), \mathcal{D}^Q_{\mathcal{T}_i}) = \min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\theta', \mathcal{D}^Q_{\mathcal{T}_i})$$

Figure: `https://www.youtube.com/watch?v=ItPEBdD6VMk`

MAML for Few-Shot Supervised Learning

f_θ Meta Model

$$\min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\theta - \alpha \nabla_\theta \mathcal{L}(\theta, \mathcal{D}^S_{\mathcal{T}_i}), \mathcal{D}^Q_{\mathcal{T}_i}) = \min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\theta', \mathcal{D}^Q_{\mathcal{T}_i})$$

Tasks

Figure: `https://www.youtube.com/watch?v=ItPEBdD6VMk`

MAML for Few-Shot Supervised Learning

f_θ Meta Model

Loss function

$$\min_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\theta - \alpha \nabla_\theta \mathcal{L}(\theta, \mathcal{D}^S_{\mathcal{T}_i}), \mathcal{D}^Q_{\mathcal{T}_i}) = \min_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\theta', \mathcal{D}^Q_{\mathcal{T}_i})$$

Tasks

Figure: `https://www.youtube.com/watch?v=ItPEBdD6VMk`

MAML for Few-Shot Supervised Learning

f_θ Meta Model

Loss function

$$\min_\theta \sum_{\mathcal{T}_i\sim p(\mathcal{T})} \mathcal{L}(\theta - \alpha\nabla_\theta \boxed{\mathcal{L}(\theta, \mathcal{D}^S_{\mathcal{T}_i})}, \mathcal{D}^Q_{\mathcal{T}_i}) = \min_\theta \sum_{\mathcal{T}_i\sim p(\mathcal{T})} \mathcal{L}(\theta', \mathcal{D}^Q_{\mathcal{T}_i})$$

Tasks

Figure: `https://www.youtube.com/watch?v=ItPEBdD6VMk`

MAML for Few-Shot Supervised Learning

f_θ Meta Model

Updated Parameters

Loss function

$$\min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\theta - \alpha \nabla_\theta \boxed{\mathcal{L}(\theta, \mathcal{D}^S_{\mathcal{T}_i})}, \mathcal{D}^Q_{\mathcal{T}_i}) = \min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\theta', \mathcal{D}^Q_{\mathcal{T}_i})$$

Tasks

Figure: `https://www.youtube.com/watch?v=ItPEBdD6VMk`

MAML for Few-Shot Supervised Learning

f_θ Meta Model

Loss function

Updated Parameters

$$\min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\underbrace{\theta - \alpha \nabla_\theta \boxed{\mathcal{L}(\theta, \mathcal{D}^{S}_{\mathcal{T}_i})}}_{\text{Gradient Descent}}, \mathcal{D}^{Q}_{\mathcal{T}_i}) = \min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\theta', \mathcal{D}^{Q}_{\mathcal{T}_i})$$

Tasks

Figure: https://www.youtube.com/watch?v=ItPEBdD6VMk

MAML for Few-Shot Supervised Learning

f_θ Meta Model

Loss function

Updated Parameters

$$\min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\underbrace{\theta - \alpha \nabla_\theta \boxed{\mathcal{L}(\theta, \mathcal{D}^S_{\mathcal{T}_i})}}_{\text{Gradient Descent}}, \mathcal{D}^Q_{\mathcal{T}_i}) = \min_{\theta} \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}(\theta', \mathcal{D}^Q_{\mathcal{T}_i})$$

Support Set

Tasks

Query Set

Figure: `https://www.youtube.com/watch?v=ItPEBdD6VMk`

MAML for Few-Shot Supervised Learning

Figure: `https://www.youtube.com/watch?v=ItPEBdD6VMk`

MAML for Few-Shot Supervised Learning

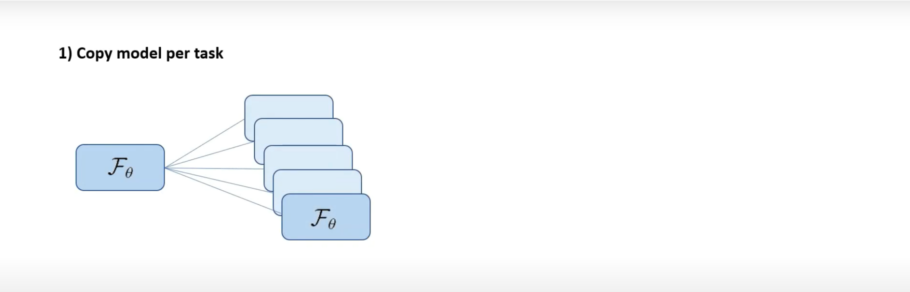


Figure: `https://www.youtube.com/watch?v=ItPEBdD6VMk`

MAML for Few-Shot Supervised Learning

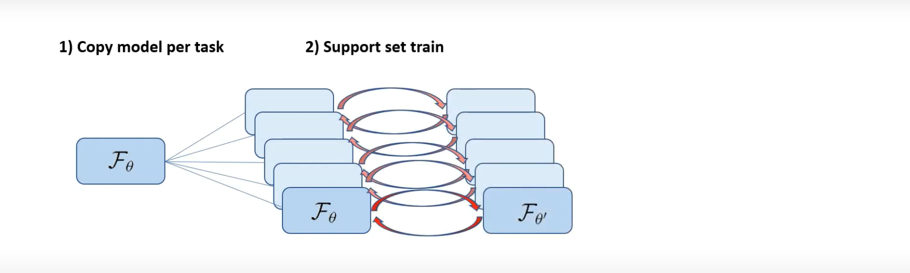


Figure: https://www.youtube.com/watch?v=ItPEBdD6VMk

MAML for Few-Shot Supervised Learning

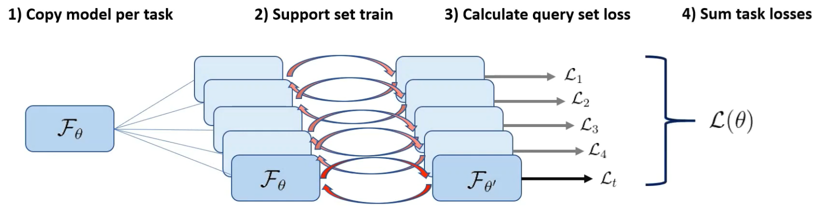


Figure: https://www.youtube.com/watch?v=ItPEBdD6VMk

MAML for Few-Shot Supervised Learning

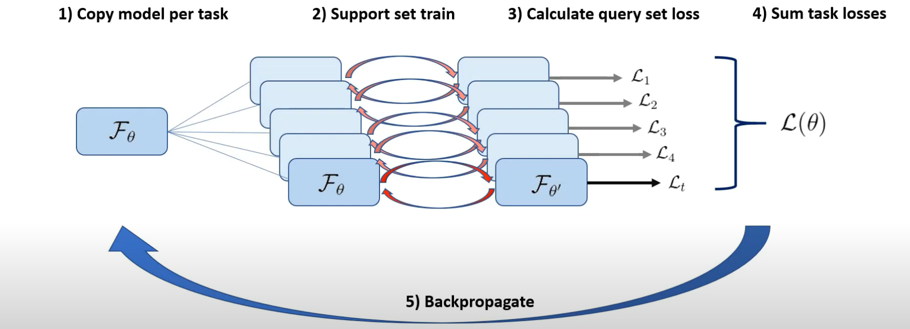


Figure: https://www.youtube.com/watch?v=ItPEBdD6VMk

Table of Contents

Results of Few-shot classification

Omniglot (Lake et al., 2011)	5-way Accuracy 1-shot	5-way Accuracy 5-shot	20-way Accuracy 1-shot	20-way Accuracy 5-shot
MANN, no conv (Santoro et al., 2016)	82.8%	94.9%	–	–
MAML, no conv (ours)	**89.7 ± 1.1%**	**97.5 ± 0.6%**	–	–
Siamese nets (Koch, 2015)	97.3%	98.4%	88.2%	97.0%
matching nets (Vinyals et al., 2016)	98.1%	98.9%	93.8%	98.5%
neural statistician (Edwards & Storkey, 2017)	98.1%	99.5%	93.2%	98.1%
memory mod. (Kaiser et al., 2017)	98.4%	99.6%	95.0%	98.6%
MAML (ours)	**98.7 ± 0.4%**	**99.9 ± 0.1%**	**95.8 ± 0.3%**	**98.9 ± 0.2%**

MiniImagenet (Ravi & Larochelle, 2017)	5-way Accuracy 1-shot	5-way Accuracy 5-shot
fine-tuning baseline	28.86 ± 0.54%	49.79 ± 0.79%
nearest neighbor baseline	41.08 ± 0.70%	51.04 ± 0.65%
matching nets (Vinyals et al., 2016)	43.56 ± 0.84%	55.31 ± 0.73%
meta-learner LSTM (Ravi & Larochelle, 2017)	43.44 ± 0.77%	60.60 ± 0.71%
MAML, first order approx. (ours)	**48.07 ± 1.75%**	**63.15 ± 0.91%**
MAML (ours)	**48.70 ± 1.84%**	**63.11 ± 0.92%**

Thank you