Preliminary Work

MAML

MAML

$\hat{\theta}^n$: model learned from task n

$\hat{\theta}^n$ depends on ϕ

Loss Function:

$$L(\phi) = \sum_{n=1}^{N} l^n(\hat{\theta}^n)$$

$l^n(\hat{\theta}^n)$: loss of task n on the testing set of task n

How to minimize $L(\phi)$? Gradient Descent

$$\phi \leftarrow \phi - \eta \nabla_\phi L(\phi)$$

Find ϕ achieving good performance **after training** 潛力

Model Pre-training

Widely used in transfer learning

Loss Function:

$$L(\phi) = \sum_{n=1}^{N} l^n(\phi)$$

Find ϕ achieving good performance 現在表現如何

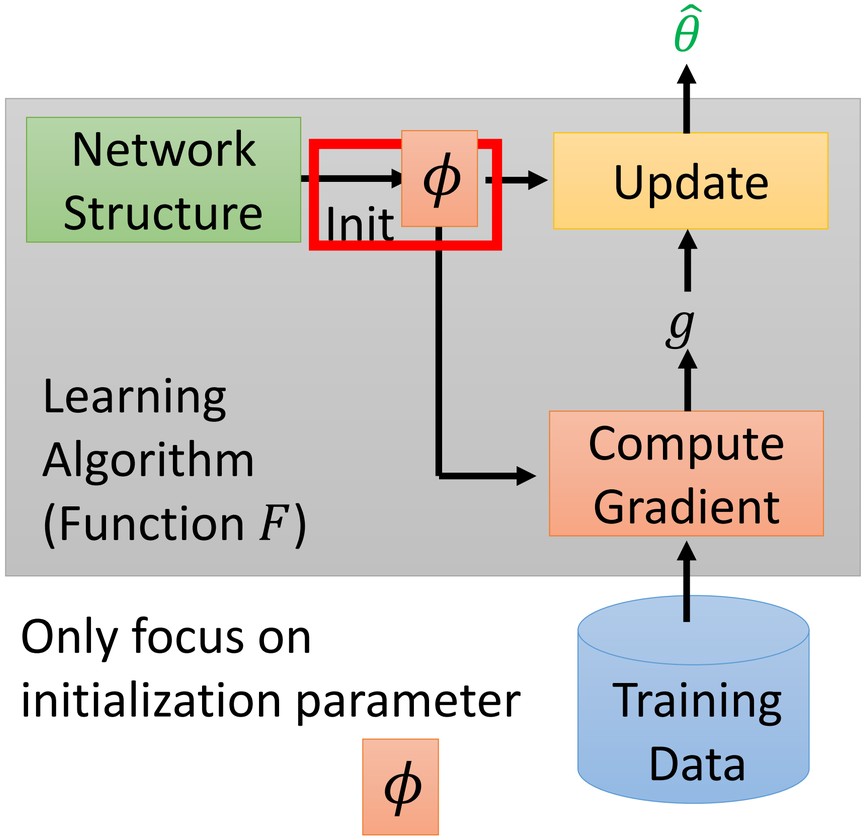
https://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2019/Lecture/Meta1%20(v6).pdf

Preliminary Work

MAML

MAML

- Fast … Fast … Fast …
- Good to truly train a model with one step. ☺
- When using the algorithm, still update many times.
- Few-shot learning has limited data.



$$L(\phi) = \sum_{n=1}^{N} l^n(\hat{\theta}^n)$$

$$\phi \leftarrow \phi - \eta \nabla_{\phi} L(\phi)$$

Considering one-step training:

$$\hat{\theta} = \phi - \varepsilon \nabla_{\phi} l(\phi)$$

https://speech.ee.ntu.edu.tw/~tlkagk/courses/ML_2019/Lecture/Meta1%20(v6).pdf