When MAML Learns Quickly Does It Generalize Well?

O. Taylan Turan*, David M.J. Tax*, and Marco Loog*
Pattern Recognition Laboratory, Delft University of Technology, Van Mourik Broekmanweg 6, Delft 2628 XE, The Netherlands
o.t.turan@tudelft.nl github.com/taylanot

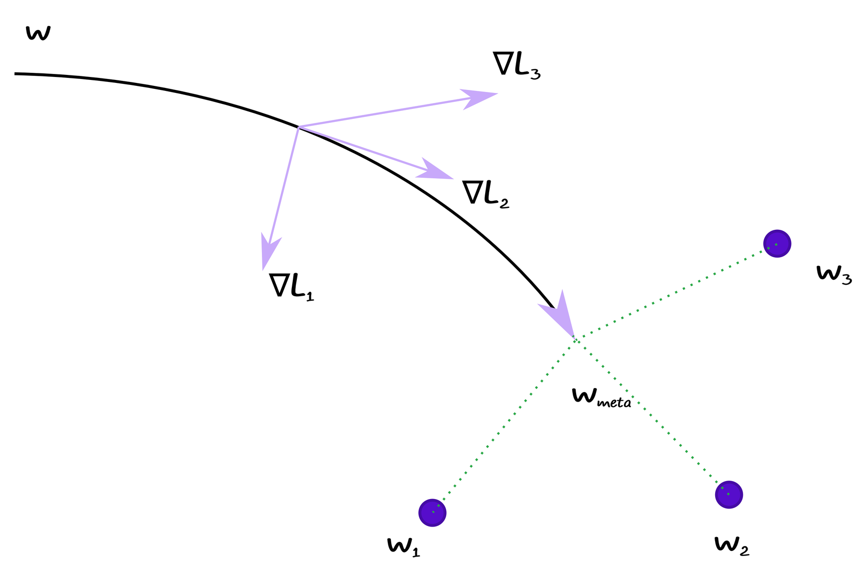
1. Introduction

- Learning-to-Learn paradigm: leverages similar learning problems (tasks) for a specific similar data-scarce learning problem (task).
- MAML: tackles meta-learning problem by providing an initialization for model parameters that facilitates quick adaptation and good generalization.

AIM : Investigating the effect of gradient step limitation.

2. MAML[1]

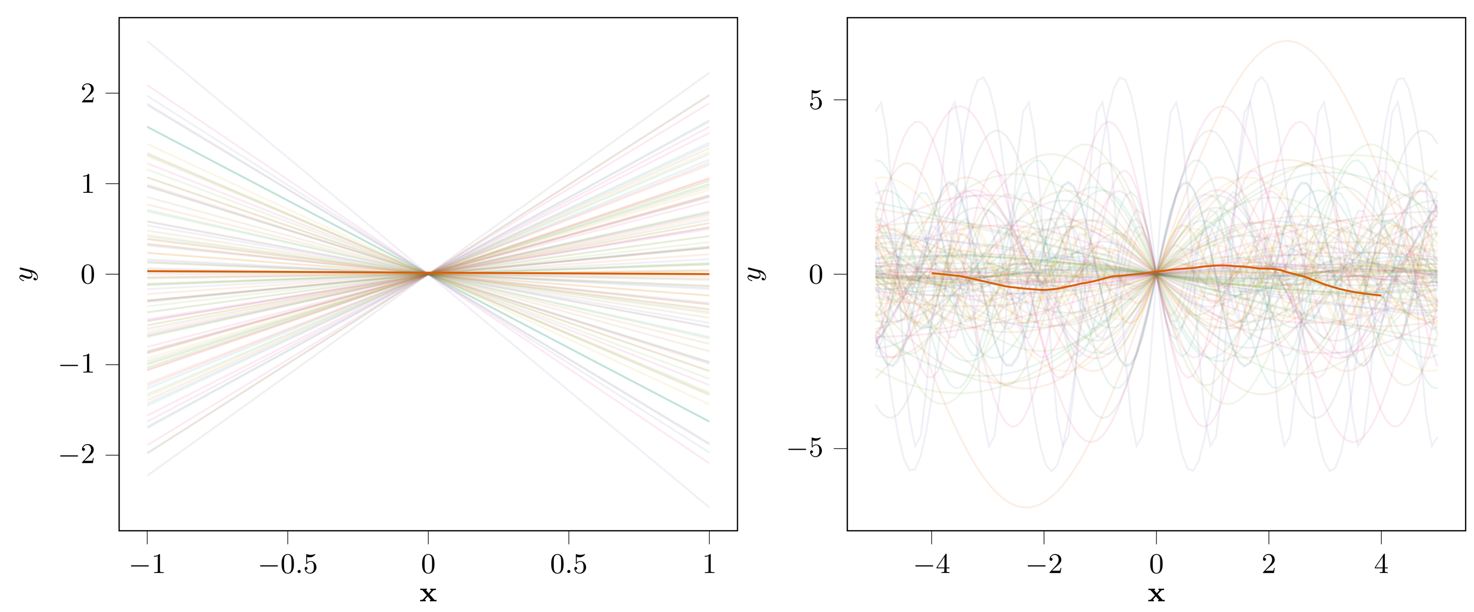
- From M tasks $\{\mathcal{T}_i\}_{i=0}^{M}$
- Learn a model initialization $\bar{\mathbf{w}}_{\text{meta}}$



3. Experimental Setup

- Tasks: linear/nonlinear noisy ($\varepsilon \sim \mathcal{N}(0, \sigma^2)$) observations of functions $f(\mathbf{x})$

$$\mathbf{y} = f(\mathbf{x}) + \varepsilon, \quad (1)$$

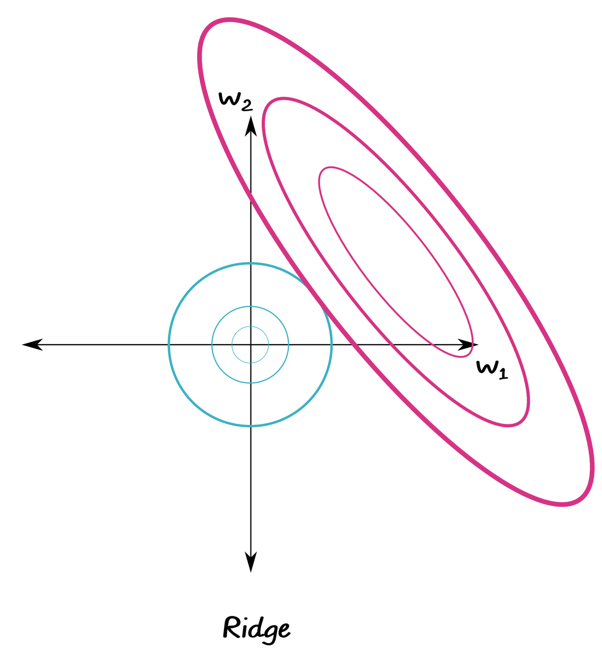


- Estimator: model $\hat{\mathcal{M}}$ trained with a given dataset $\mathcal{Z} := \{\mathbf{x}_i, y_i\}_{i=0}^{N}$
- Performance: expected error over the task distribution $p_{\mathcal{T}}$ and data distribution $p_{\mathcal{Z}}$

$$\mathcal{E} := \iiint (\hat{\mathcal{M}}(\mathbf{x}) - y)^2 p(\mathbf{x}, y) p_{\mathcal{Z}} p_{\mathcal{T}} d\mathbf{x} dy d\mathcal{Z} d\mathcal{T} \quad (2)$$

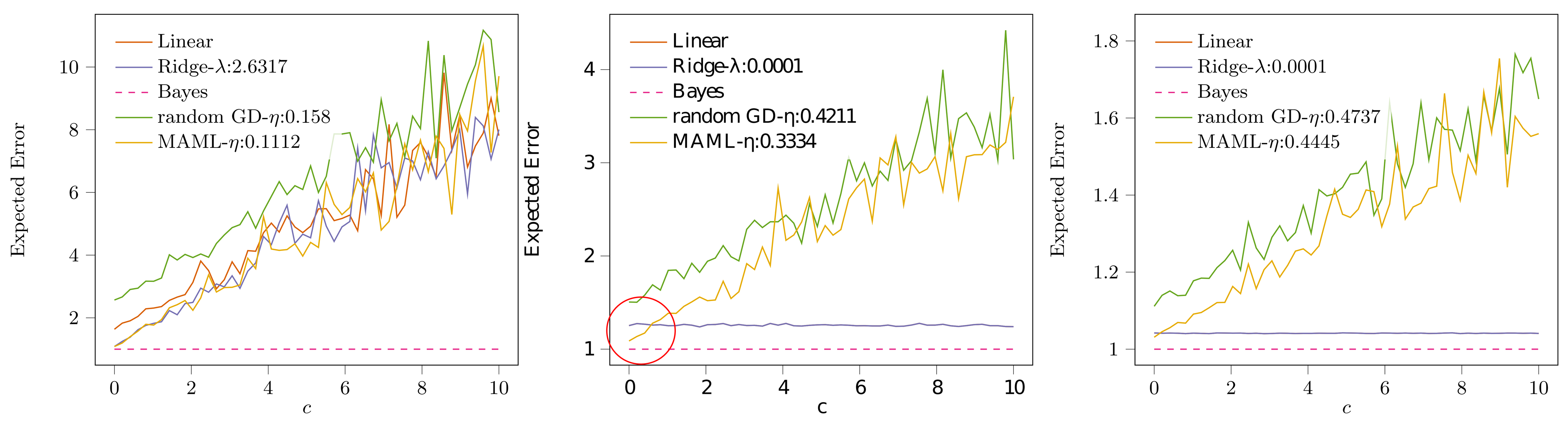
4. Baselines

- Linear/Kernelized Ridge Regression
- Randomly Initialized Gradient Descent

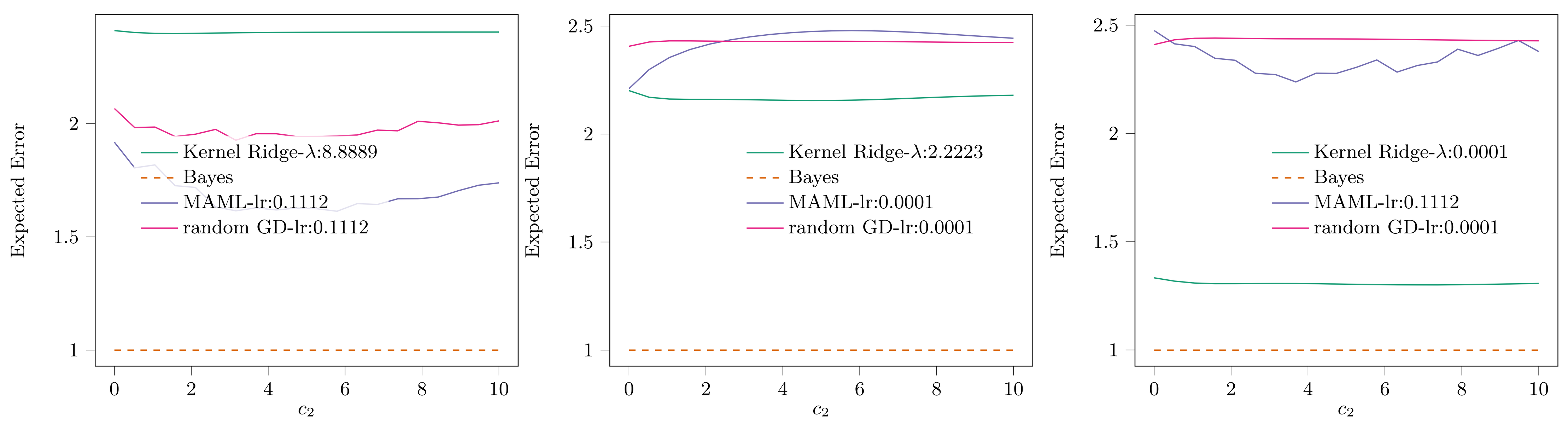


5. Results for Task Variance ($N = 1, \quad 10, \quad 50$)

- Linear problem: $f(x) := \mathbf{x}^\mathsf{T}\mathbf{a}$



- Nonlinear problem: $f(x) := \sin(\mathbf{x} + \boldsymbol{\phi})^\mathsf{T}\mathbf{a}$



6. Conclusions

- Given enough data single-task learners can outperform MAML in expectation in most of the cases
- Task variance highly influences the performance of MAML in expectation.

9. Experimentation

8. References

[1] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. *arXiv:1703.03400 [cs]*, July 2017.