EXPECTED LOSS OF A FUTURE EMPIRICAL RISK MINIMIZER

Author1, Author2
Affiliation
Univ
City
{Author1, Author2}email@email

Author3
Affiliation
Univ
City
email@email

ABSTRACT

This works aims to investigate the generalization via the expected loss calculation of, arguably the most impactful meta-learning algorithm Model-Agnostic Meta Learning (MAML) [1], which claims to have "good" generalization capabilities for a given task in few-shot image classification, few-shot regression and reinforcement learning scenarios. However, the claimed generalization properties of this method remains rather illusive due to non-convex problem settings that it is being utilized. This work aims illustrates first in a linear and convex, then in a non-linear and non-convex supervised learning setting. Throughout the work comparison with the Bayes error is used as a baseline and models with convex losses are used as benchmarks with or without meta-information for the problems being considered. We aim to showcase the gain of performance that comes with the MAML, if there is any.

***K*eywords** First keyword · Second keyword · More

1 Introduction

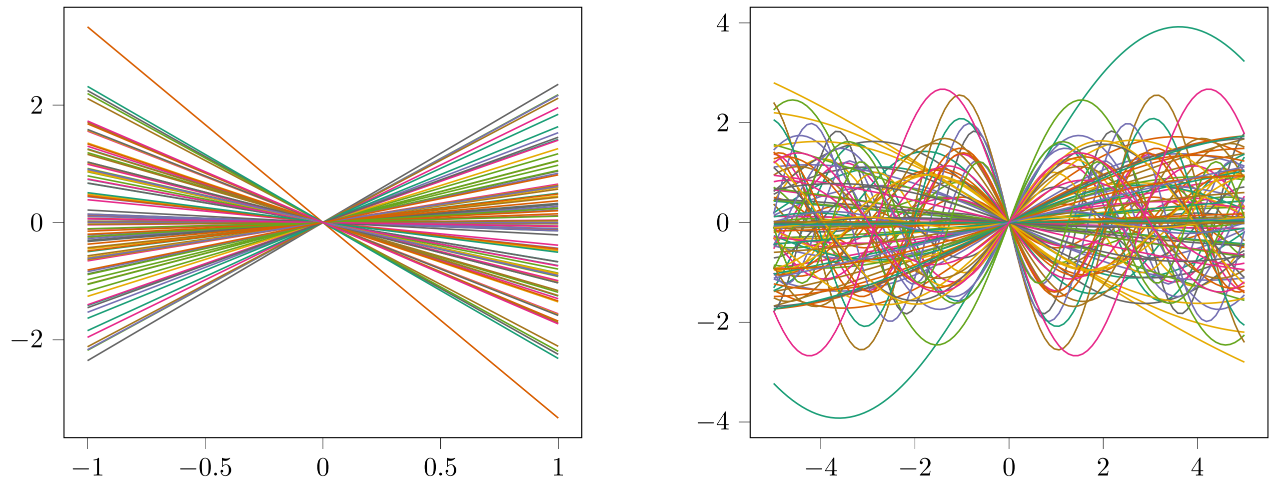
Learning-to-learn(meta-learning) is a hot research field which treats the learning of the traditional machine learning task as the learning problem. The utility that comes with this specific learning enhances the capability of the machine related prediction tasks by means of increased efficiency of data utilization resulting from the context detection capability of the algorithms presented under this paradigm.

Model-Agnostic Machine Learning (MAML)[1] is arguably one of the most impactful meta-learning algorithm that is proposed recently. The reason for this method to gain this much traction can be associated with it being non-parametric in the meta level. In a broad sense, this algorithm seeks to find an intermediate model in the environment of tasks. The way to obtain this model is to look at the future loss of a possible gradient steps from the tasks observed. This makes MAML highly attractive for the gradient based methods as the implementation effort that goes into any model that rely on gradients are minimal. Given the fact that most deep learning frameworks rely on the gradient descent the research avenue that [1] opens up is quite wide.

Although, there are multiples of works that built upon the ideology presented in [1], the claimed generalization aspect is investigated in a limited extended only with convex problem assumptions. In a similar fashion this paper aims to investigate the generalization of MAML with the aim to find out to what extend the claims regarding generalization hold. The main research questions investigated in this paper are:

- *What is the extend of MAML's generalization capabilities?*
- *Is MAML algorithm really model agnostic?*
- *Is the generalization performance of MAML, merely and artifact of non-convex problem setting?*

Authors' of this paper acknowledges the fact that in the definition of meta-learning given in [1] includes the term "quickly learn" which indicates that there is computational budget limitation. That is why the biggest chunk of interest



(a) $y = \mathbf{x}^{\mathrm{T}}\mathbf{a}$

(b) $y = \sin(\mathbf{x} + \boldsymbol{\phi})^{\mathrm{T}}\mathbf{a}$

Figure 1: 100 sample tasks drawn from $p_{\mathbf{a}}$ for linear and $p_{\mathbf{a},\boldsymbol{\phi}}$ nonlinear tasks

lies in the small sample and limited gradient steps cases in both linear and non-linear problem setting. However, the effect of number of gradient steps is investigated as well.

2 Related Work

3 Experimental Setting

Throughout this work uppercase bold letters (*e.g.*$\mathbf{X}$), lowercase bold letters (*e.g.*$\mathbf{x}$) and lowercase letters (*e.g.*x) are used for matricies, vectors and scalars. Moreover, the vectors are assumed to be stored in columns.

3.1 Linear Regression Problem

Consider the conventional linear regression setting in $\mathbb{R}^D$

$$y = \mathbf{x}^{\mathrm{T}}\mathbf{a} + \varepsilon \tag{1}$$

where, $y \in \mathbb{R}$, $\mathbf{x} \in \mathbb{R}^D$, $\mathbf{a} \in \mathbb{R}^D$ and $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Assuming that the each realization of $\mathbf{a}$ corresponds to a task

4 Results and Discussion

5 Conclusion

Acknowledgments

References

[1] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. *34th International Conference on Machine Learning, ICML 2017*, 3:1856–1868, 2017.