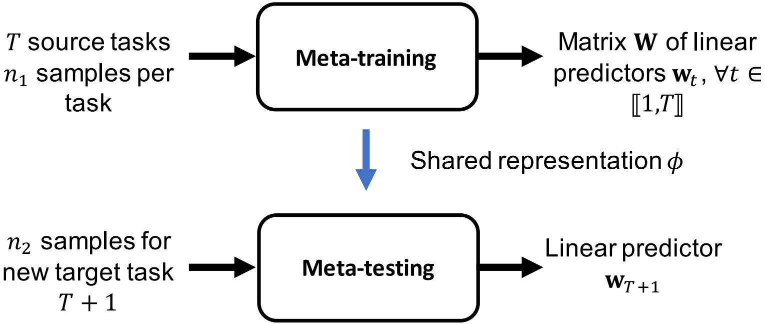
Putting Theory to Work: From Learning Bounds to Meta-Learning Algorithms

Quentin Bouniot †‡ Ievgen Redko ‡ Romaric Audigier † Angélique Loesch † Amaury Habrard ‡

† Université Paris-Saclay, CEA, List, F-91120, Palaiseau, France ‡ Université de Lyon, UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, Saint-Etienne, France

Meta-Learning 101



Goal: Minimize **excess risk** $\mathrm{ER} = \mathcal{L}(\phi^*, \mathrm{w}^*_{T+1}) - \mathcal{L}(\phi, \mathrm{w}_{T+1})$,

- True risk $\mathcal{L}$
- optimal weights ϕ^*
- w^*_{T+1} ideal target linear predictor.

Assumption 1: Diversity of the source tasks

Optimal predictors $\mathbf{W}^* = [\mathbf{w}_1^*, \ldots, \mathbf{w}_T^*]$ **cover** all the **directions evenly**

Assumption 2: Constant classification margin

Norm of $\{\mathbf{w}_t^*\}_{t \in [\![1,T]\!]}$ **should not increase** with T

$$\boxed{\textbf{if satisfied}, \ \mathrm{ER}(\phi, \mathbf{w}_{T+1}) \leq O\left(\frac{1}{n_1 T} + \frac{1}{n_2}\right)}$$ [2, 5]

✓ All source and target data are useful to decrease the excess risk

From Theory to Practice

Ensuring Assumption 1: Add **spectral** or **entropic** regularization

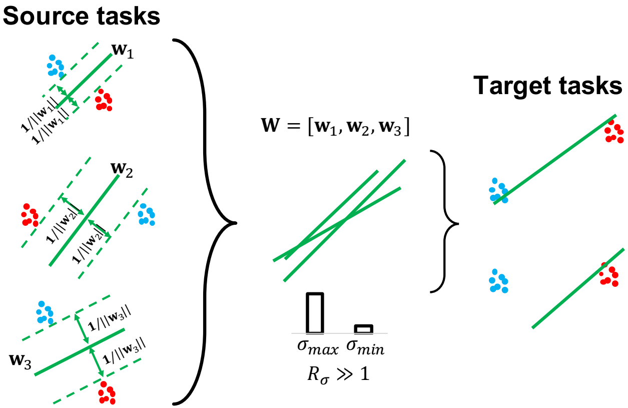
$$R_\sigma(\mathrm{W}) = \frac{\sigma_{max}(\mathrm{W})}{\sigma_{min}(\mathrm{W})} \quad \text{or} \quad H_\sigma(\mathrm{W}) = -\sum_{i=1}^{N} \mathrm{softmax}(\sigma(\mathrm{W}))_i \cdot \log \mathrm{softmax}(\sigma(\mathrm{W}))_i$$

✓ Regularizing with $R_\sigma(\mathrm{W})$ or $H_\sigma(\mathrm{W})$ leads to a better coverage of the searched space

Ensuring Assumption 2: Add **norm regularization/normalization** for linear predictors

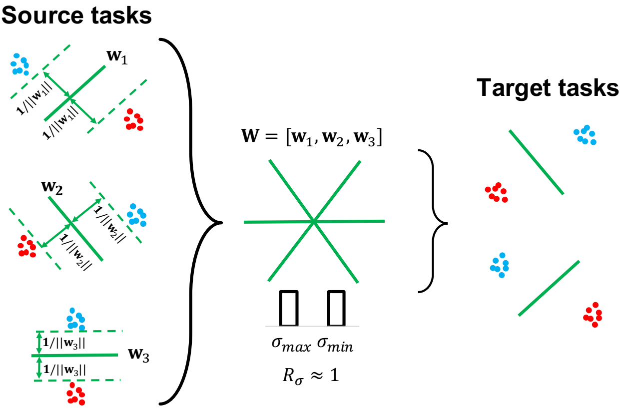
✓ **Normalizing predictors** ensure **constant margin** that **does not change** with T

Without regularization



× Linear predictors cover **only part of the space** or **over-specialize** to the tasks

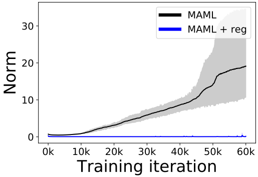
With proposed regularization

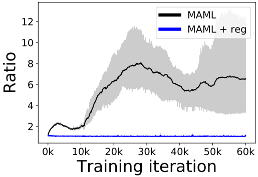


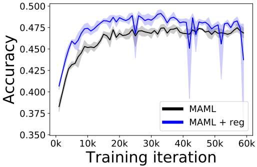
✓ **Assumption 1** makes sure that **linear predictors are complementary**

✓ **Assumption 2 avoids under- or over-specialization** to the tasks

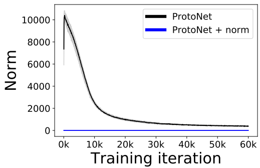
Tracking **the ratio** and **the norm**

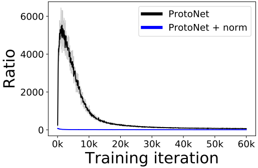


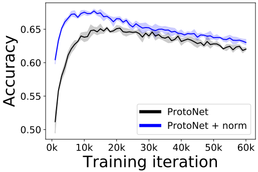




× MAML [3] does not verify the assumptions







✓ PROTONET [4] naturally verifies the assumptions

Accuracy gap with regularization

Dataset	Episodes	MAML[3]	PROTONET[4]	BASELINE[1]	BASELINE++[1]
Omniglot	20-way 1-shot	+3.95*	+0.33*	−13.2*	−7.29*
	20-way 5-shot	+1.17*	+0.01	+0.66*	−2.24*
miniImageNet	5-way 1-shot	+1.23*	+0.76*	+1.52*	+0.39
	5-way 5-shot	+1.96*	+2.03*	+1.66*	−0.13
tieredImageNet	5-way 1-shot	+1.42*	+2.10*	+5.43*	+0.28
	5-way 5-shot	+2.66*	+0.23	+1.92*	−0.72

✓ Enforcing the assumptions leads to **better generalization** when not verified naturally

Conclusions

- ✓ **Practical ways** to enforce **theoretical** assumptions
- ✓ Some models **naturally fulfill** them
- ✓ Our regularization allows to **learn faster** with **better generalization**

References

[1] Wei-Yu Chen, Yu-Chiang Frank Wang, Yen-Cheng Liu, Zsolt Kira, and Jia-Bin Huang.
A closer look at few-shot classification.
In *ICLR*, 2019.

[2] Simon S. Du, Wei Hu, Sham M. Kakade, Jason D. Lee, and Qi Lei.
Few-Shot Learning via Learning the Representation, Provably.
In *arXiv:2002.09434*, 2020.

[3] Chelsea Finn, Pieter Abbeel, and Sergey Levine.
Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks.
In *International Conference on Machine Learning*, 2017.

[4] Jake Snell, Kevin Swersky, and Richard S. Zemel.
Prototypical Networks for Few-shot Learning.
In *Advances in Neural Information Processing Systems*, 2017.

[5] Nilesh Tripuraneni, Chi Jin, and Michael I. Jordan.
Provable Meta-Learning of Linear Representations.
In *arXiv:2002.11684*, 2020.

Arxiv paper:

Contact:

quentin.bouniot@cea.fr

@QBouniot

https://qbouniot.github.io

4th Workshop on Meta-Learning at NeurIPS 2020, Vancouver, Canada
11/12/2020

Quentin Bouniot, Ievgen Redko, Romaric Audigier, Angélique Loesch, Amaury Habrard

Thank you for listening !

Commissariat à l'énergie atomique et aux énergies alternatives
Institut List | CEA SACLAY NANO-INNOV | BAT. 861 – PC142
91191 Gif-sur-Yvette Cedex - FRANCE
www-list.cea.fr