Meta-Learning Approaches for Regression, Classification and Graph Representation Learning

Haitz Sáez de Ocáriz Borde, Katherine Collins, and Ryan Crowley

Motivation

- Deep learning is often data-hungry
- Meta-learning advantages
  - Make predictions on new tasks with few examples
  - More similar to human learning
  - Increases generalization capacity
  - "Learning to learn"

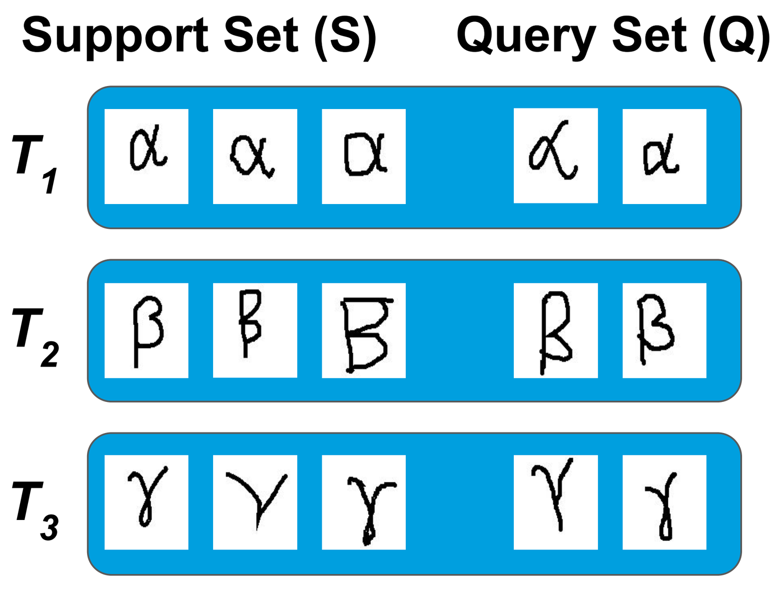

Figure 1: Meta-Learning Structure

Methods

Model-Agnostic Meta-Learning (MAML)

- Framework applicable to regression, classification, and RL[1]
- Loss function includes 1st and 2nd order derivatives

$$\theta_i \leftarrow \theta - \alpha \nabla_\theta L_{T_i}(f_\theta)$$

Equation 1: MAML Task-specific update

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{T_i \sim p(T)} L_{T_i}(f_{\theta_i'})$$

Equation 2: MAML Meta-update

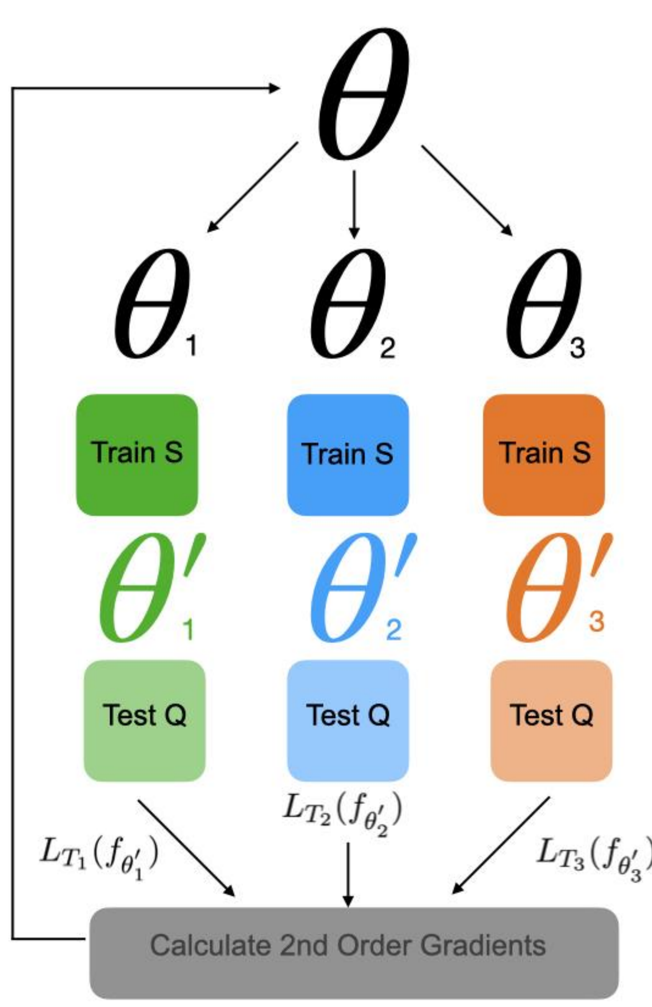

Figure 2: MAML Framework

MAML + Cosine Annealing (CA)

- Proposed in compendium of MAML advances[2]
- Adjusts learning rate scheduling for meta-optimizer

Reptile

- MAML variant[3], similar to First Order MAML[1]
- Performs only Stochastic Gradient Descent (SGD)
- Less computationally expensive than MAML

$$\theta \leftarrow \theta + \beta(U_T^k(\theta) - \theta)$$

Equation 3: Reptile Update

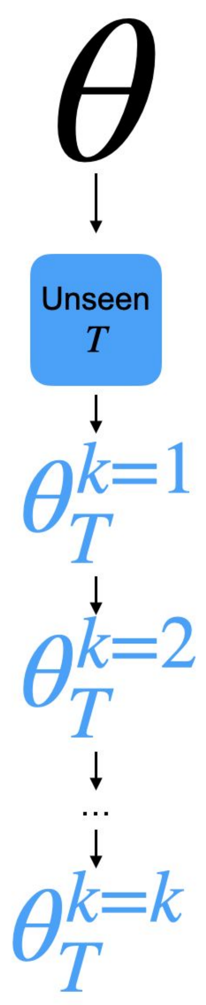

Figure 3: k-Shot Evaluation

Notation

θ = prior parameters, θ' = task-specific parameters
T = task, K = number of observations per task
k = number of gradient steps for unseen task

Sinusoidal Regression

Can we fit a novel sine wave from few observation?

- Observations per wave are (x,y) points

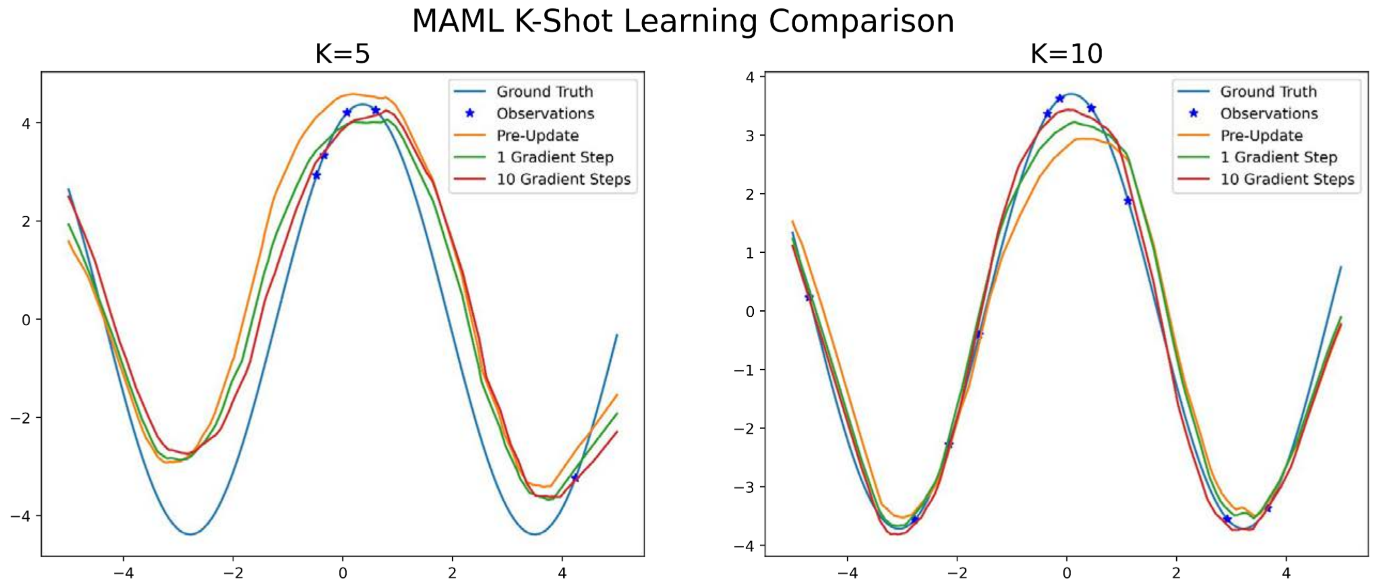

Figure 4: Sine Wave Regression Adaptation

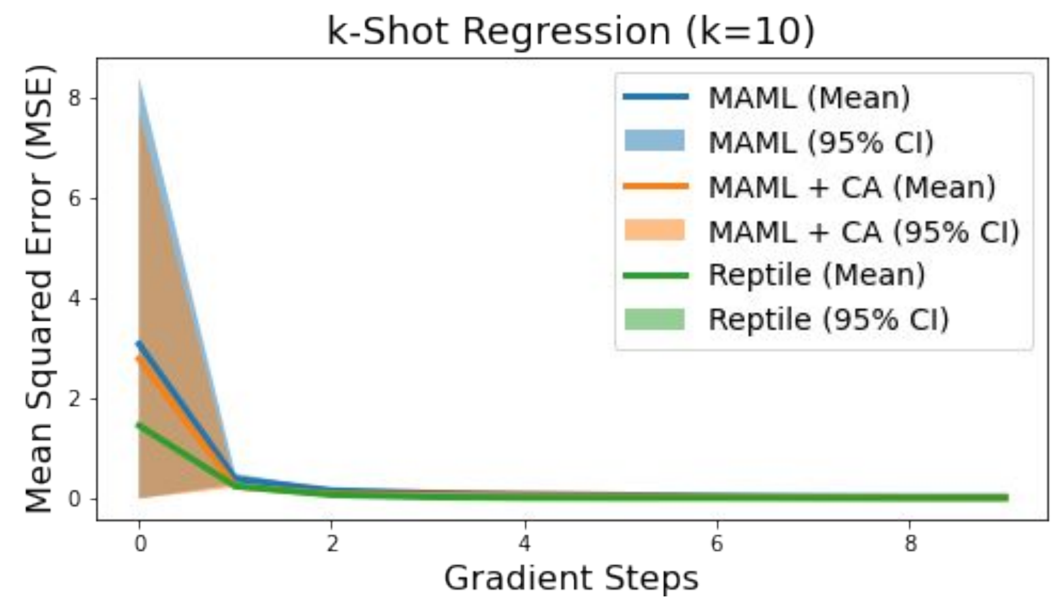

Figure 5: Meta-Learning Algorithm Comparison (K = 10)

Pushing MAML's Limits

Can we scale to sine waves in multiple dimensions?

	Pre-Update	1 Gradient Step	5 Gradient Steps
MAML	4.24 (+/- 0.54)	1.11 (+/- 0.26)	0.31 (+/- 0.03)
MAML + CA	4.30 (+/- 0.41)	2.36 (+/- 0.34)	1.63 (+/- 0.18)
Reptile	4.27 (+/- 0.50)	0.97 (+/- 0.16)	0.16 (+/- 0.04)

Table 1: Meta-Learning Algorithm Comparison for Multi-Dim (D = 20) Sine Regression (MSE Loss)

Can we handle out-of-distribution functions? Noise?

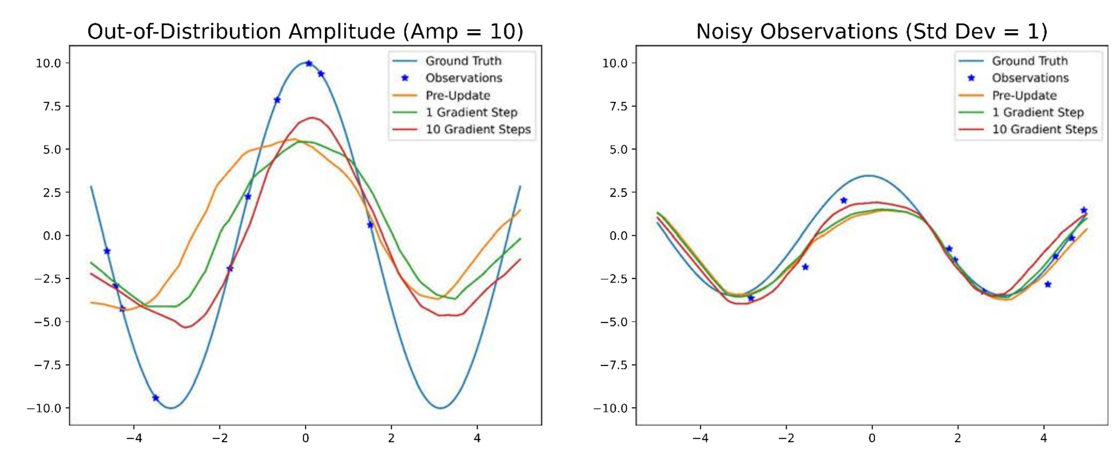

Figure 6: MAML Sine Wave Adaptation Under Data Shifts

Key Takeaways

- MAML replicated, compared to Reptile and MAML + CA
- High variability in MAML depending on seeds
- Scales to multidimensional output
- Struggles with distribution shifts

Extension to Graphs

Can we extend these meta-learning frameworks to graph representation learning?

- Compare GNN types for meta-learning
- Equivariant Message Passing GNNs[5] for improved performance

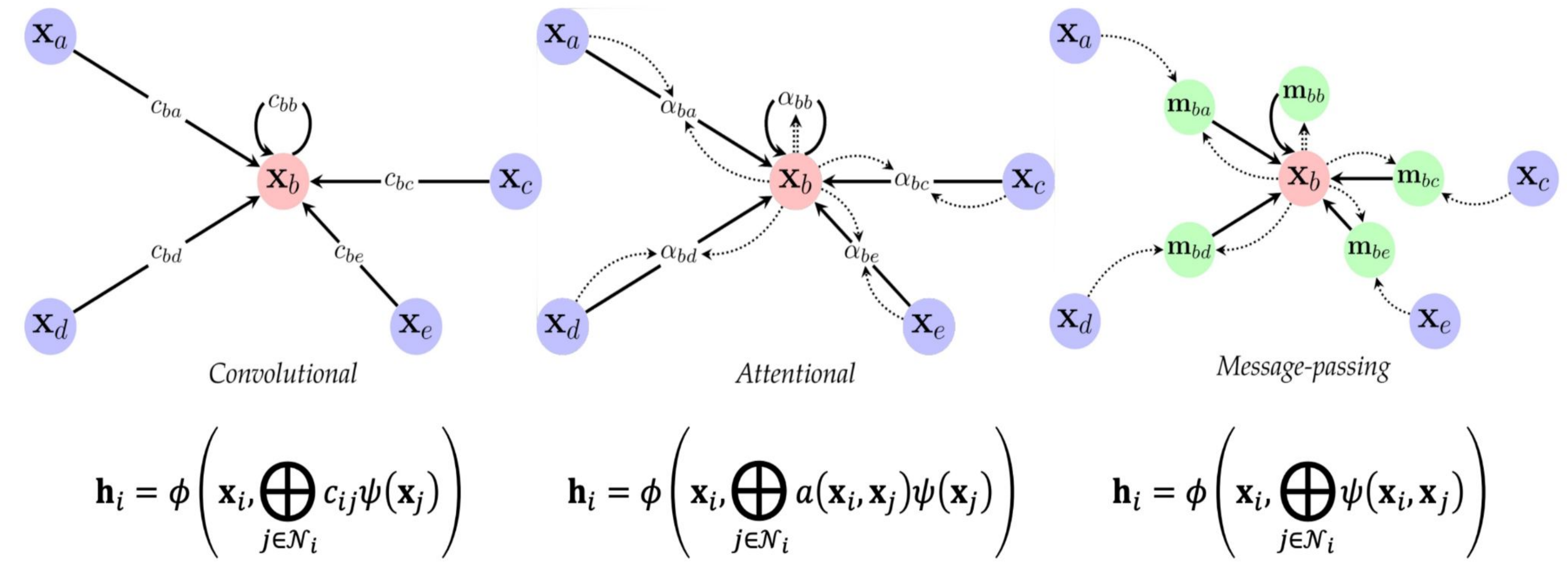

Figure 7: The flavors of GNN layers. Figure from Bronstein 2021[4]

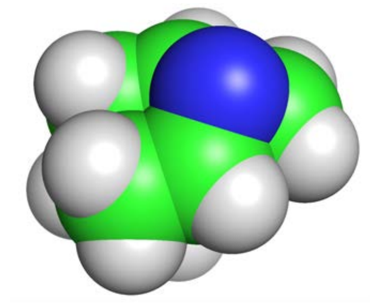

Figure 8: QM9 Molecule (PyMOL[6])

	Pre-Update	1 Gradient Step	5 Gradient Steps
Convolutional	3514.3 (+/- 557692.8)	153.8 (+/- 42812.9)	17.8 (+/- 1898.6)
Attention	156.2 (+/- 3363.1)	25.4 (+/- 474.1)	1.8 (+/- 4.7)
Equivariant Message Passing	658.1 (+/- 1057556.7)	0.4 (+/- 0.5)	1e-6 (+/- 2e-11)

Table 2: Meta-Learning for QM9 Graph Regression (MSE Loss)

Key Takeaways

- Meta-learning algorithms are applicable to Graph Representation Learning
- More general graph layers improve meta-learning performance
- Equivariant Message Passing allows modeling of intermolecular forces

Classification

Can we learn to classify images and graphs of novel classes from few examples?

- Same algorithm, updated loss and neural network
- CNN for Omniglot
- GNN classification for Enzymes TUDataset

Figure 9: Sample Omniglot Characters

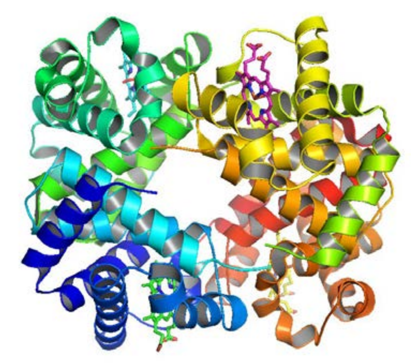

Figure 10: Protein Structure (PyMOL)

Discussion

Additional MAML Limitations

- Requires extensive tuning to perform well on real-world tasks
- Prohibitively slow in low compute settings
- Implementation hurdles around second-order derivatives

Future Work

- Explore additional improvements to MAML (MAML++[2])
- Investigate ensemble methods for meta-learning

UNIVERSITY OF CAMBRIDGE

References

1. C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in Proceedings of the 34th International Conference on Machine Learning, vol. 70. PMLR, 06–11 Aug 2017, pp. 1126–1135. [Online]. Available: https://proceedings.mlr.press/v70/finn17a.html
2. A. Antoniou, H. Edwards, and A. J. Storkey, "How to train your MAML," CoRR, vol. abs/1810.09502, 2018. [Online]. Available: http://arxiv.org/abs/1810.09502.
3. A. Nichol, J. Achiam, and J. Schulman, "On first-order meta-learning algorithms," CoRR, vol. abs/1803.02999, 2018. [Online]. Available: http://arxiv.org/abs/1803.02999
4. M. M. Bronstein, J. Bruna, T. Cohen, P. Velickovic, Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges (2021).
5. Satorras, Víctor Garcia, Emiel Hoogeboom, and Max Welling. "E (n) equivariant graph neural networks." International Conference on Machine Learning. PMLR, 2021.
6. Schrödinger, LLC, "The PyMOL molecular graphics development component, version 1.8," November 2015.

Presented at the Advanced Machine Learning (AML) Poster Presentation March 21, 2022, Cambridge, UK