Mid-Evaluation Report

From Classical Machine Learning to
Model-Agnostic Meta-Learning

Saksham Agrahari
Department of Electrical Engineering
Indian Institute of Technology Kanpur

January 9, 2026

Abstract

This report presents a structured study that begins with classical supervised machine learning and progresses toward Model-Agnostic Meta-Learning (MAML). The discussion first formulates machine learning as a parameter optimization problem and examines its behavior across regression and classification tasks. Observations from repeated task-level learning motivate the need for learning frameworks that support efficient adaptation.

Meta-learning is then introduced as a formal extension in which tasks, rather than individual data points, constitute the primary learning units. The MAML algorithm is studied in detail through its mathematical formulation, emphasizing task-specific adaptation and meta-level optimization. The report further analyzes supervised regression and classification under the meta-learning framework and concludes with a discussion of open challenges and future directions.

1 Classical Machine Learning

Classical supervised machine learning formulates learning as the problem of optimizing model parameters using a fixed dataset associated with a single task. Given sufficient labeled data, models are trained to minimize a loss function that measures prediction error. This paradigm has been successfully applied across a wide range of domains, including signal processing, pattern recognition, and data-driven modeling.

In this setting, learning is typically performed independently for each task. Once a task changes, models are either retrained from scratch or fine-tuned using task-specific data. While effective in static environments, this approach does not explicitly account for repeated learning across related tasks.

—

1.1 Learning as an Optimization Problem

Let a model be represented by a function f_θ, parameterized by θ. Given a dataset $\mathcal{D} = \{(x_j, y_j)\}$, learning is defined as:

$$\min_{\theta} \; L(\theta), \tag{1}$$

where $L(\theta)$ is a loss function measuring the discrepancy between predictions and true outputs.

Optimization is commonly performed using gradient descent:

$$\theta \leftarrow \theta - \eta \nabla_\theta L(\theta), \tag{2}$$

where η denotes the learning rate. This update rule forms the foundation of most machine learning algorithms.

—

1.2 Regression and Classification

In regression tasks, outputs are continuous-valued, and loss functions such as mean squared error are used. In classification tasks, outputs are discrete labels, and loss functions such as cross-entropy are commonly employed. Although the loss functions differ, both settings rely on the same underlying optimization mechanism.

Empirical observations from learning tasks show that performance is strongly influenced by data availability, initialization, and model capacity.

—

2 Limitations of Classical Learning

Despite its success, classical machine learning exhibits limitations when learning must be repeated across multiple related tasks. Each task is treated independently, resulting in redundant computation and inefficient use of prior learning experience.

In low-data regimes, models trained from scratch exhibit unstable optimization and poor generalization. Fine-tuning alleviates this issue to some extent but does not explicitly optimize for adaptability. These limitations motivate learning frameworks that account for task-level structure.

—

3 Meta-Learning Framework

Meta-learning addresses the inefficiency of task-isolated learning by explicitly modeling learning across a distribution of tasks. In this framework, the objective is not to optimize performance on a single task, but to acquire parameters that facilitate rapid learning on new tasks.

Tasks are assumed to be drawn from an underlying distribution $p(T)$, and learning proceeds by leveraging shared structure across tasks.

—

3.1 Meta-Learning Problem Set-Up

In the meta-learning formulation, a task represents a complete learning problem characterized by its own data distribution and loss function. Learning is defined at the task level rather than the data level. During training, tasks are sampled from $p(T)$, and the model adapts to each task using limited task-specific data.

The effectiveness of this adaptation is then used to guide updates to the model's initial parameters.

—

3.2 A Model-Agnostic Meta-Learning Algorithm

Model-Agnostic Meta-Learning proposes a general algorithm applicable to any model trained using gradient descent. The algorithm does not impose restrictions on model architecture or loss function form, requiring only differentiability with respect to parameters.

Learning is decomposed into two nested processes: task-specific adaptation and meta-level optimization. By optimizing initial parameters based on post-adaptation performance, the algorithm ensures that standard gradient descent becomes an effective adaptation mechanism.

—

4 Mathematical Formulation of MAML

Let f_θ denote a model parameterized by θ. For a given task T_i, model performance is measured using a task-specific loss function $L_{T_i}(f_\theta)$.

—

4.1 Task Adaptation

Task-specific adaptation is performed using gradient descent:

$$\theta_i' = \theta - \alpha \nabla_\theta L_{T_i}(f_\theta), \tag{3}$$

where α denotes the learning rate. This update rule is identical to classical learning and represents how the model adapts to a specific task.

—

4.2 Meta-Objective

After adaptation, the adapted parameters θ_i' are evaluated:

$$\min_\theta \sum_{T_i \sim p(T)} L_{T_i}(f_{\theta_i'}). \tag{4}$$

This objective explicitly optimizes parameters for their ability to support learning, rather than immediate performance.

—

4.3 Meta-Gradient and Higher-Order Effects

Optimizing the meta-objective requires differentiating through the adaptation step, resulting in higher-order derivatives. These derivatives capture how changes in parameters influence the learning process itself.

—

5 Learning Domains Under MAML

5.1 Regression

In regression tasks, each task corresponds to learning a continuous-valued function. Meta-learning captures shared structure across tasks while allowing rapid task-specific adaptation. This setting illustrates how initialization influences convergence speed and generalization.

—

5.2 Classification

In classification tasks, each task involves learning decision boundaries from limited labeled data. Meta-learning optimizes representations that can be efficiently adjusted through gradient descent, enabling effective learning under few-shot constraints.

—

6 Discussion

Model-Agnostic Meta-Learning provides a unified framework that retains standard optimization procedures while modifying the learning objective. Its generality makes it applicable across learning domains, but computational cost and stability remain important considerations.

—

7 Future Work

Future work includes studying first-order approximations to reduce computational complexity, analyzing behavior under task distribution shift, and developing theoretical guarantees for convergence and generalization in meta-learning frameworks.

—

References

[1] C. Finn, P. Abbeel, and S. Levine, *Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks*, Proceedings of the International Conference on Machine Learning (ICML), 2017.