MAML (Model Agnostic Meta Learning)

Assignment 4: Few Shot Classification with MAML

Report by

Rohith Rajeev Nambiar

1 Introduction

The objective of this assignment is to implement Model Agnostic Meta Learning (MAML) and compare its performance with a standard joint training baseline on a synthetic few shot classification task.

The task is based on a "Moving Circle" dataset in which the decision boundary is a circle with a fixed radius $r = 2.0$. For each task, the center (c_x, c_y) of the circle is sampled uniformly from the region $[-3, 3] \times [-3, 3]$. The model must learn to adapt to a new, unseen circle location using only a small labeled support set and a limited number of gradient updates. The goal is to evaluate how effectively meta-learning enables fast adaptation to new tasks.

2 Procedure

2.1 MAML Implementation

We implemented the MAML algorithm to train a model that can quickly adapt to new tasks with minimal data.

- **Model Architecture:** A multi-layer perceptron (MLP) with two hidden layers and ReLU activation functions was used as the base model.
- **Meta Training:** During training, batches of tasks were sampled from the moving circle distribution. For each task, the model first performed a gradient update on the support set to obtain adapted parameters. These adapted parameters were then evaluated on a query set.
- **Meta Update:** The loss computed on the query set was used to update the original model parameters. This process trains the model to find an initialization that can adapt quickly to new tasks.

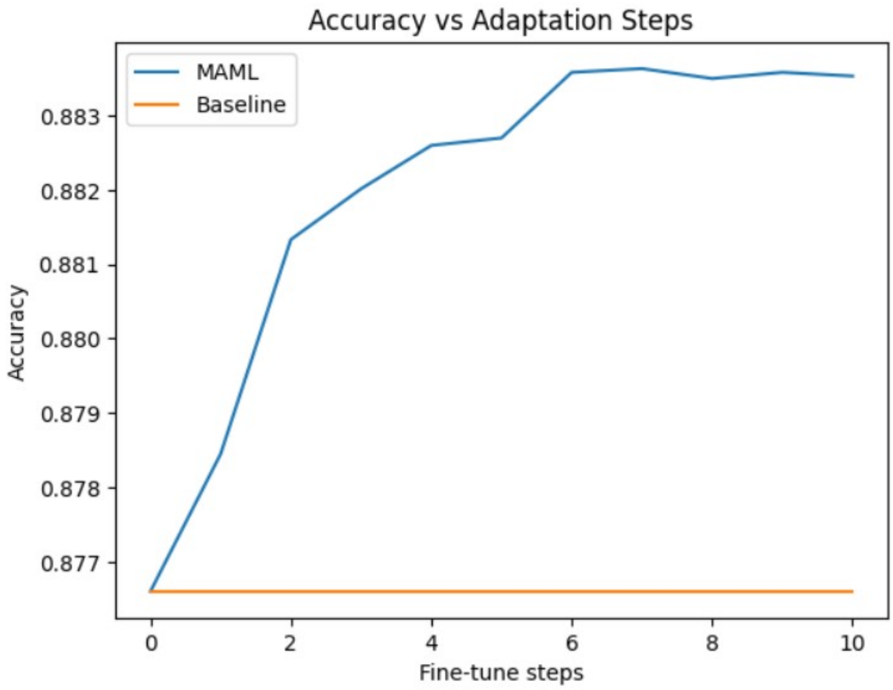
2.2 Baseline Method

As a comparison, a standard joint training baseline was implemented.

- The baseline model was trained using data sampled from many tasks at once, without explicit meta learning.
- At test time, the trained model was fine tuned on the support set of a new task using standard gradient descent.
- This baseline helps evaluate the benefit of meta learning compared to traditional supervised training.

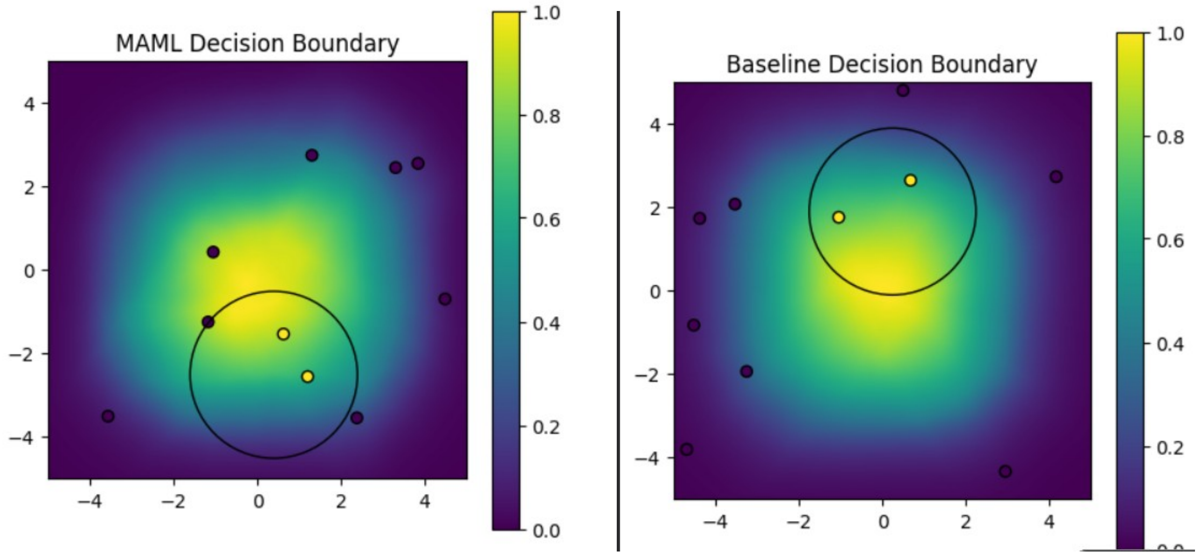
3 Results

Both models were evaluated on unseen test tasks by measuring classification accuracy after a small number of gradient updates.



The MAML model showed strong performance immediately after adaptation, achieving high accuracy with very few gradient steps. This indicates that the meta-learned initialization was well suited for rapid task-specific learning.

In contrast, the baseline model improved more slowly. Since it was trained to perform well on average across tasks, it required more updates to adjust to a specific new circle location.



Overall, the results demonstrate that MAML is more effective for few-shot learning scenarios, where fast adaptation to new tasks is essential.

4 Bonus Question

Why does the meta-loss not decrease uniformly with every epoch?

The meta-loss in a meta-learning setup such as MAML does not decrease smoothly or uniformly with each epoch because the training process is stochastic and task dependent.

First, each training iteration samples a new task from a distribution of tasks. In the moving circle dataset, the position of the circle changes randomly from task to task. Some sampled tasks are easier than others, while some are more difficult. As a result, the loss computed at each epoch can vary significantly depending on the specific task sampled, leading to visible fluctuations in the meta loss curve.

Second, unlike standard supervised learning where the model optimizes a fixed dataset, meta learning optimizes performance across a continuously changing set of tasks. The objective function is therefore not stationary. The model is not minimizing loss on a single fixed problem, but instead learning an initialization that works well after adaptation on many different tasks. This causes natural variability in the measured loss from epoch to epoch.

Finally, the meta-loss is typically estimated using relatively small support and query sets. Because the evaluation is based on a limited number of samples, statistical noise can have a noticeable effect. A few misclassified points in the query set can cause a sudden increase in loss, even if the overall learning trend is improving.

For these reasons, the meta loss curve often appears noisy and does not decrease uniformly, even though the overall long term trend may still show improvement.