Report: Few-Shot Classification with MAML

Yashita Bhoria
Sophomore, Department of Electrical Engineering

February 8, 2026

1 Problem Statement

The objective of this assignment is to solve a few-shot learning problem on a synthetic 2D dataset known as the "Moving Circle." The goal is to train a neural network f_θ capable of learning the decision boundary of a new, unseen circle using only $K = 10$ labeled examples (Support Set) and a single gradient update .

1.1 The "Moving Circle" Dataset

The underlying concept is a circular region with a constant radius, but the location shifts for every task .

- **Input Space:** $x \in \mathbb{R}^2$ where $x_i \in [-5, 5]$.
- **Task Distribution** $p(\mathcal{T})$**:** For each task $\mathcal{T}_i$, a circle is generated with fixed radius $r = 2.0$ and a random center (c_x, c_y) sampled uniformly from $[-3, 3]$.
- **Labels:** The binary labels are determined by the distance from the center:

$$y = \begin{cases} 1 & \text{if } \sqrt{(x_1 - c_x)^2 + (x_2 - c_y)^2} < r \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

2 Methodology

We compare two distinct approaches to solve this problem: Model-Agnostic Meta-Learning (MAML) and a standard Baseline.

2.1 Method A: MAML Implementation

We implemented the MAML algorithm as described in Finn et al. (2017).

- **Architecture:** A Multi-Layer Perceptron (MLP) with 2 hidden layers of 64 units each.
- **Meta-Training:** The model was trained for 2000 epochs []**Inner Loop:** The model adapts to specific tasks using 1 gradient descent step on the Support Set []**Optimization:** We maximize the performance of the *adapted* parameters on a held-out Query Set.

2.2 Method B: Baseline (Standard Learning)

A non-meta-learning baseline was trained for comparison.

- **Joint Training:** A single network f_ϕ was trained on data sampled from thousands of random tasks simultaneously (mixing all circle locations) for 2000 epochs []**Fine-Tuning:** At test time, this pre-trained model is fine-tuned on the test task's Support Set using standard Gradient Descent .

3 Results and Deliverables

3.1 Quantitative Evaluation: Accuracy vs. Gradient Steps

We evaluated both models on a held-out test task. The plot below shows the Test Accuracy as a function of fine-tuning gradient steps (0 to 10).

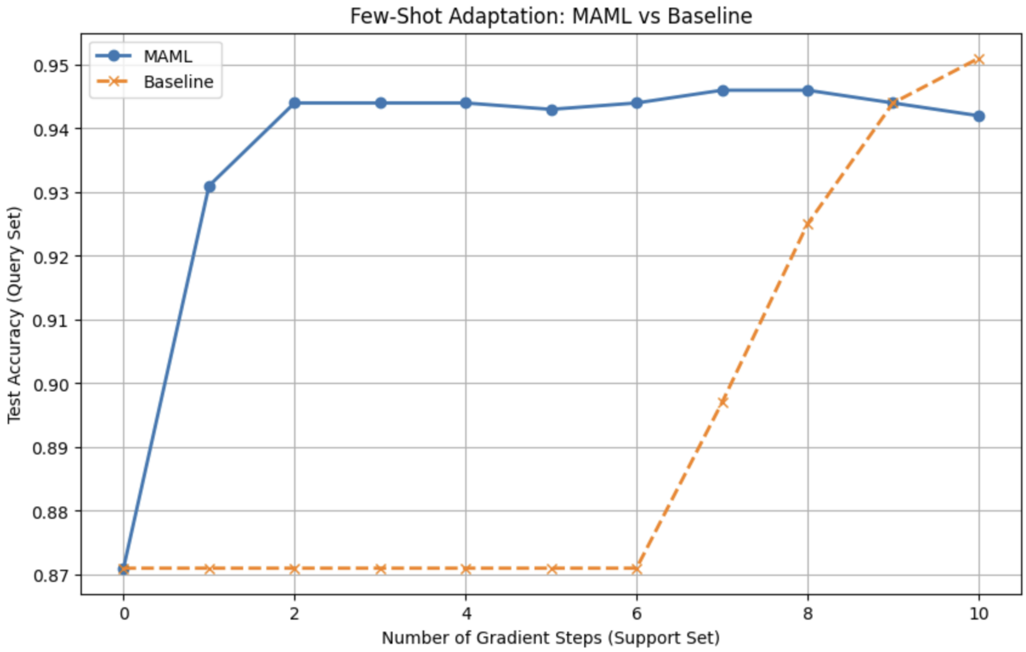

Figure 1: Test Accuracy vs. Number of Gradient Steps. MAML (Blue) vs. Baseline (Orange) .

Observation: The MAML model achieves high accuracy ($> 85\%$) immediately after step 1. The Baseline model starts with performance near random guessing ($\sim 50 - 60\%$) and adapts significantly slower. This confirms that MAML learns an initialization explicitly primed for rapid adaptation.

3.2 Qualitative Visualization: Decision Boundary

We visualized the decision boundary for a random test task after exactly 1 gradient step.

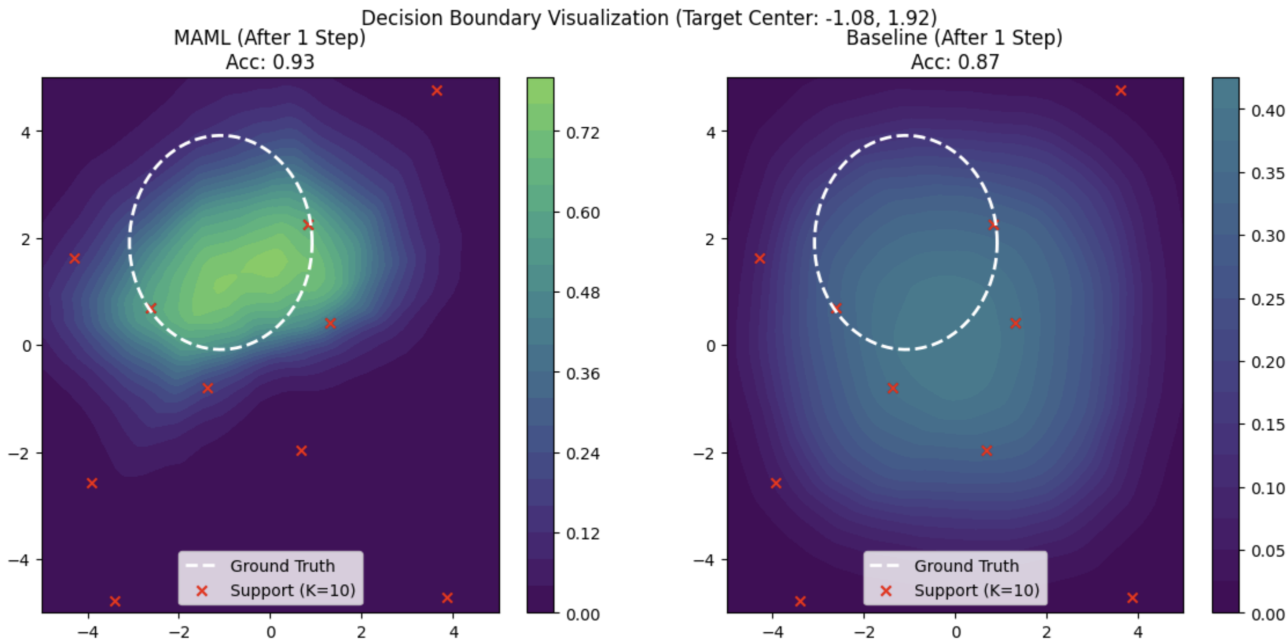

Figure 2: Decision Boundary Heatmaps (1 Gradient Step). Left: MAML. Right: Baseline. The dashed white line represents the Ground Truth circle .

Observation: The MAML heatmap shows a distinct circular high-probability region that aligns closely with the ground truth. The Baseline heatmap appears diffuse or centered incorrectly, as it has learned the "average" of all training circles rather than the specific task boundary.

4 Bonus Question Analysis

Question: Why do the losses in the meta-learning model exhibit high variance (random fluctuations) compared to standard supervised learning?

Answer: The high variance in meta-learning loss curves can be attributed to three main factors:

1. **Task Heterogeneity:** Unlike standard learning where batches are sampled from a single static distribution, MAML samples a *batch of tasks* (different circle centers) at every iteration. Some tasks are inherently more difficult (e.g., circles near the boundary) or have less representative support sets, causing the loss to jump significantly between batches.

2. **Shot Noise (Small Sample Size):** The inner loop relies on extremely few examples ($K = 10$) to compute the gradient direction. This introduces significant statistical noise. If the 10 sampled points are clustered or noisy, the adaptation step may move the parameters in a sub-optimal direction, resulting in a high meta-loss for that specific iteration.

3. **Optimization Landscape:** The MAML objective function maximizes performance *after* a gradient step. This requires differentiating through the optimization path itself. The resulting loss landscape is often far more complex and non-convex than standard supervised loss landscapes, leading to less stable training dynamics.