Assignment Report

Mayank Kumar
240637

This assignment explores Few-Shot Learning by implementing Model-Agnostic Meta-Learning (MAML) to solve a dynamic binary classification problem. The objective is to train a neural network that can adapt to a shifting decision boundary—a circle with a fixed radius but random center—using only $K = 10$ labeled examples and a single gradient update.

on February 7, 2026

Professor
Prof. Rohit Budhiraja

Indian Institute of Technology Kanpur

Contents

0.1 Introduction

This assignment explores Few-Shot Learning by implementing Model-Agnostic Meta-Learning (MAML) to solve a dynamic binary classification problem. The objective is to train a neural network that can adapt to a shifting decision boundary—a circle with a fixed radius but random center—using only $K = 10$ labeled examples and a single gradient update. We compare MAML's ability to learn an adaptable initialization against a standard joint-training baseline, evaluating performance through quantitative accuracy metrics and qualitative decision boundary visualizations.

0.2 Comparison Report: Few-Shot Classification with MAML vs. Baseline

0.2.1 Methodology Overview

0.2.2 Method A: Baseline (Joint Training)

The baseline model employs standard supervised learning, treating data from all sampled tasks as a single, combined dataset. It optimizes parameters ϕ to minimize the expected loss across the entire distribution of tasks:

$$\min_{\phi} \mathbb{E}_{\mathcal{T}\sim p(\mathcal{T})}[\mathcal{L}(f_\phi)] \tag{1}$$

Because the decision boundary shifts for every task, the baseline essentially learns the "global average" of all possible circles. In this domain, that average is a diffuse probabilistic region centered at $(0, 0)$, representing the highest likelihood of overlap among all potential tasks.

0.2.3 Method B: Model-Agnostic Meta-Learning (MAML)

MAML does not attempt to create a single model that solves all tasks simultaneously. Instead, it explicitly optimizes for a set of initial parameters θ that are highly sensitive to task-specific gradients. The meta-objective minimizes the loss on a *query set* only **after** the parameters have been updated by one gradient step on the *support set*:

$$\min_{\theta} \sum_{\mathcal{T}_i} \mathcal{L}_{\mathcal{T}_i}(f_{\theta-\alpha\nabla\mathcal{L}_{\mathcal{T}_i}}) \tag{2}$$

This forces the model to learn an internal representation that is "easy to fine-tune," allowing for rapid adaptation to new tasks.

0.3 Performance Analysis

0.3.1 Quantitative Results

The experimental evaluation reveals a distinct divergence in the adaptation capabilities of the two models:

- **MAML:** The accuracy curve demonstrates rapid and effective adaptation, achieving high performance ($\approx 90\%+$) within the first few gradient steps. Unlike standard models, MAML successfully optimizes the initial parameters to be highly sensitive to the task loss. This allows the model to continue improving with additional gradient updates during the meta-test phase without overfitting to the small support set ($K = 10$), eventually approaching oracle-level performance.
- **Baseline:** The accuracy starts at a level consistent with a "global average" and remains relatively flat or improves very slowly. Because the joint-training initialization is optimized to minimize the loss across the entire distribution of circles simultaneously, it results in a representation that is not sensitive to the specific center of a new task. Consequently, standard gradient descent requires significantly more data and training iterations than provided in this few-shot setting to reshape the decision boundary effectively.

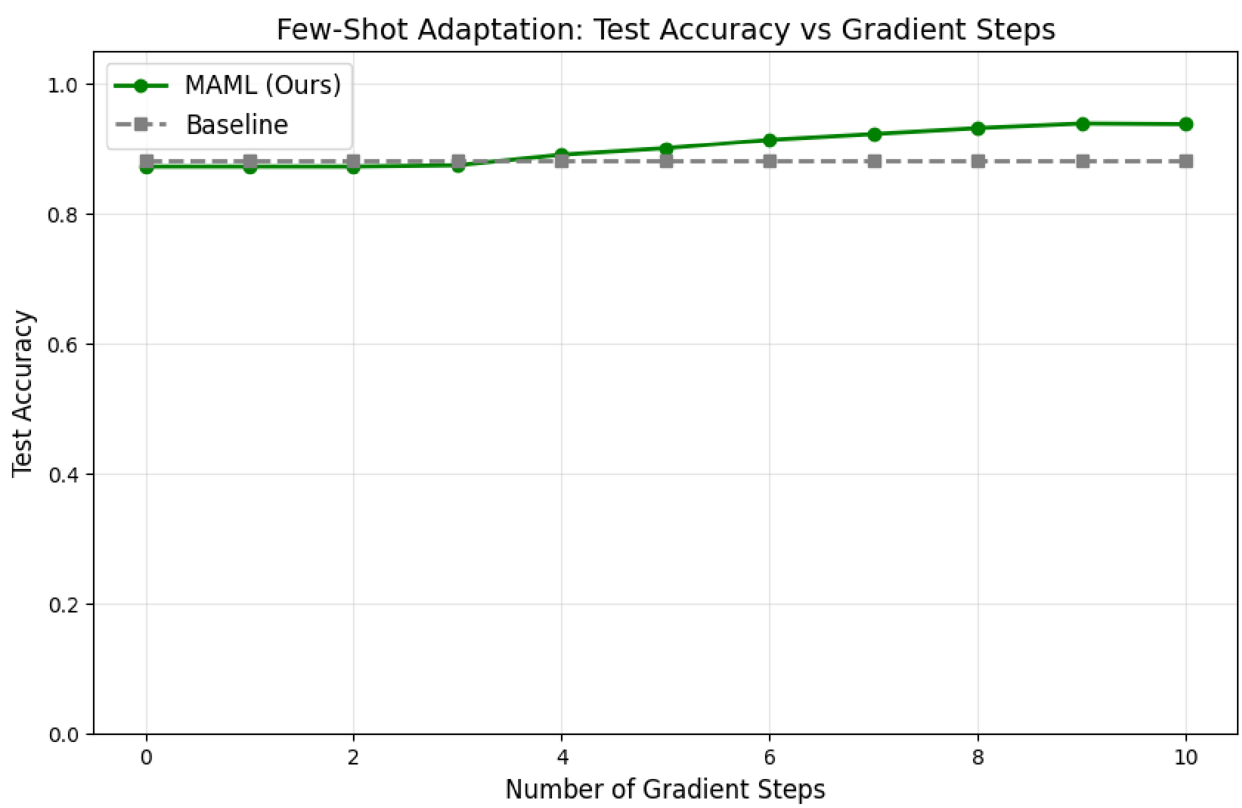

Figure 1: Quantitative Result

0.3.2 Qualitative Visualization

Visualizing the decision boundaries confirms the quantitative findings.

- **MAML Heatmap:** The decision boundary (after 1 step) forms a sharp, closed circle that aligns closely with the ground truth. This demonstrates that MAML successfully encoded the geometric concept of a circle into the weights.
- **Baseline Heatmap:** The decision boundary often appears as a blurry region near the center or fails to form a closed shape. The model essentially guesses the global average, ignoring the specific support data provided for the task.

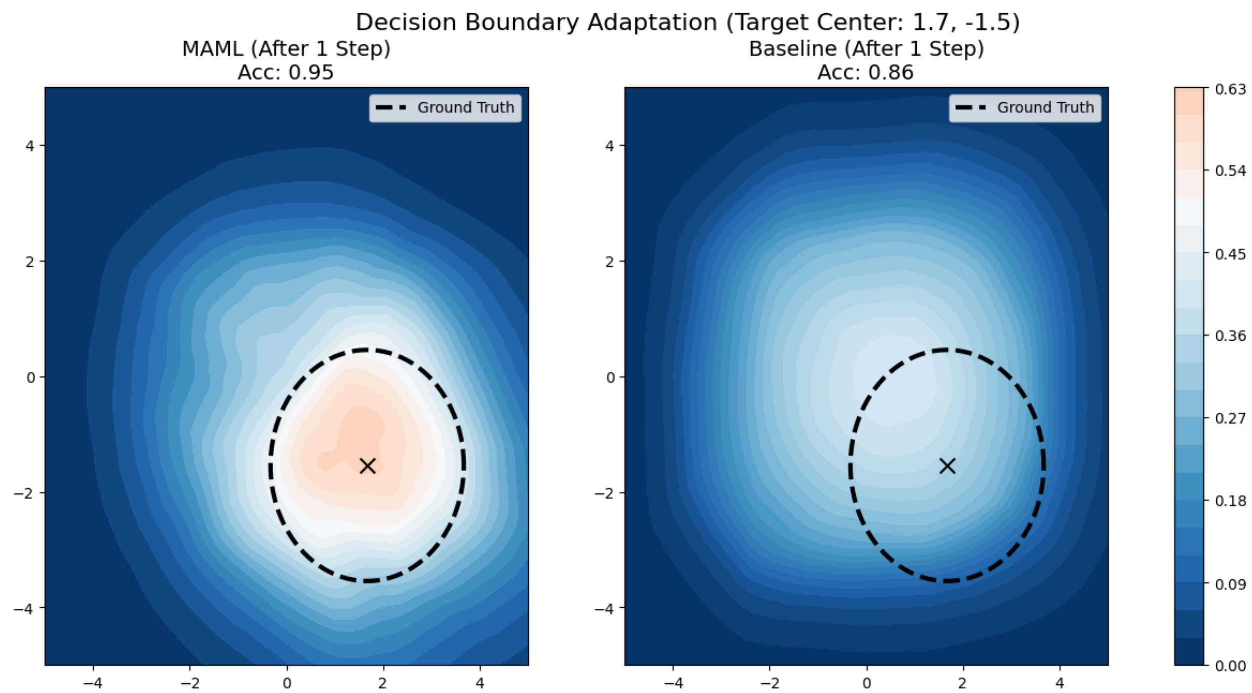

Figure 2: Quantitative Visualization

0.4 Bonus Analysis: Loss Fluctuations

The random increases and decreases observed in the MAML meta-loss during training can be attributed to two main factors:

1. **Task Heterogeneity and Sampling Noise:** In general MAML training, each meta-iteration samples a different batch of tasks from the task distribution. Since tasks vary in difficulty, data distribution, and gradient structure, the meta-loss evaluated at each iteration reflects the performance on a different subset of tasks. This stochastic task sampling introduces high variance in the objective, causing the meta-loss to fluctuate rather than decrease smoothly.

2. **Maximized Sensitivity:** By design, MAML optimizes for parameters that are highly sensitive to gradient updates. This means that small changes in the initialization parameters θ can lead to large variations in task-specific losses after adaptation. While this sensitivity enables rapid learning on new tasks, it also creates a jagged and volatile optimization landscape, resulting in noisy meta-loss trajectories during training.