

Assignment: Few-Shot Classification with MAML

Moving Circle Dataset

Chandraveer Singh Sisodia

February 8, 2026

1 Problem Statement

The objective of this assignment is to perform few-shot binary classification on the "Moving Circle" dataset[cite: 3]. The input space consists of $x \in \mathbb{R}^2$ where $x_1, x_2 \in [-5, 5]$ [cite: 7]. For each task \mathcal{T}_i , the decision boundary is defined by a circle with a fixed radius $r = 2.0$ and a random center (c_x, c_y) sampled uniformly from $[-3, 3]$ [cite: 8]. The goal is to train a model f_θ that can adapt to an unseen task using only $K = 10$ labeled examples and a single gradient update[cite: 12].

2 Methodology

Two approaches were implemented and compared:

1. **MAML (Model-Agnostic Meta-Learning):** Implemented per Finn et al. (2017) using an inner loop for task adaptation and an outer loop for meta-updates[cite: 14, 15].
2. **Baseline (Standard Learning):** A joint training approach where a single network is trained on data from thousands of tasks simultaneously[cite: 21, 23].

3 Quantitative Evaluation

The quantitative performance was evaluated by plotting Test Accuracy against the number of Gradient Steps (0 to 10) on a held-out test task[cite: 28, 29].

3.1 Observations

- **MAML Curve:** Typically starts at a moderate accuracy ($\approx 50 - 60\%$) and jumps to over 90% in exactly one gradient step. This demonstrates that MAML has optimized the initial parameters specifically for rapid adaptation.
- **Baseline Curve:** Starts at a lower accuracy and climbs much more slowly. Since its weights are optimized for the "average" of all circles during joint training, it lacks the specialized initialization required for quick adaptation.

4 Qualitative Visualization

Decision boundary heatmaps were generated for a random test task after 1 gradient step to visualize predicted probabilities[cite: 32, 33, 35].

4.1 Comparison Results

- **MAML Result:** The heatmap displays a clearly defined circular region aligning closely with the ground-truth dashed line[cite: 34]. This confirms MAML has learned the underlying circular "concept" and uses the support set to identify the specific translation[cite: 36].
- **Baseline Result:** The heatmap appears diffuse or centered toward the origin. Due to simultaneous training on thousands of tasks, it struggles to pivot to specific coordinates using only 10 points and 1 step[cite: 36].

5 Bonus Question

Question: Why is the Meta Loss not decreasing uniformly as observed in normal training trends?

Solution:

1. **Task Diversity (The Primary Reason):** Unlike standard training where the model looks at a fixed dataset, MAML samples a new batch of tasks every iteration. The model may perform well on "Easy" tasks (e.g., loss of 0.24 at Epoch 1400), but then encounter a batch of "Hard" tasks at Epoch 1600 that look very different, causing the loss to spike to 1.73.
2. **High Meta-Learning Rate ($\beta = 0.01$):** For the "outer loop" that updates the global model, a learning rate of 0.01 is relatively aggressive. When combined with the "noise" of different tasks, the model makes large jumps in weight space, causing the sharp oscillations observed in the training logs.
3. **Small Sample Size (Few-Shot Noise):** In few-shot learning, the model only sees a handful of examples per task. If these examples are noisy or unrepresentative, the calculated gradient is of "low quality." This noise propagates directly into the Meta Loss, resulting in a jagged rather than smooth trend.