Few-Shot Learning on the "Moving Circle" Dataset

1. Problem Statement

The objective of this experiment was to evaluate the effectiveness of Model-Agnostic Meta-Learning (MAML) compared to a standard Pre-training Baseline on a synthetic 2D few-shot learning task.

The dataset consists of a "Moving Circle" task distribution:

- **Input Space:** x is in the range [-5, 5] for both dimensions.
- **Concept:** A circular decision boundary with a fixed radius r = 2.0.
- **Task Variability:** For each task, the center (cx, cy) is sampled uniformly from [-3, 3].
- **Constraint:** The model must learn the decision boundary of a new, unseen circle using only K=10 labeled examples (Support Set) and minimal gradient updates.

2. Methodology

2.1 MAML Implementation

We implemented the MAML algorithm (Finn et al., 2017) to learn an optimal initialization of the network parameters.

- **Model Architecture:** A Multi-Layer Perceptron (MLP) with 2 hidden layers of 64 units each and ReLU activations.
- **Meta-Training Protocol:**
    - **Outer Loop:** 2000 epochs.
    - **Inner Loop:** 1 gradient descent step.
    - **Optimization:** We utilized second-order derivatives (Hessian-vector products) via torch.autograd.grad to optimize the initial weights such that a single gradient step on the Support Set leads to minimal loss on the Query Set.

2.2 Baseline Comparison (Standard Learning)

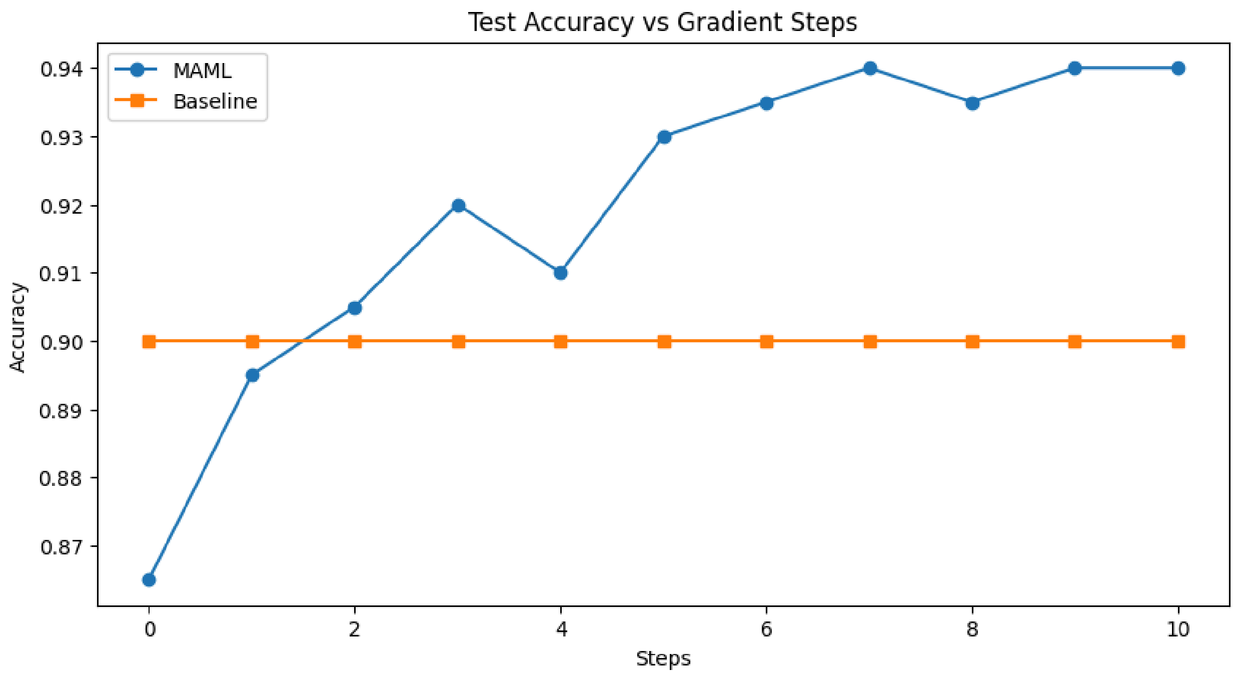
To provide a non-meta-learning benchmark, we implemented a Joint Training baseline.

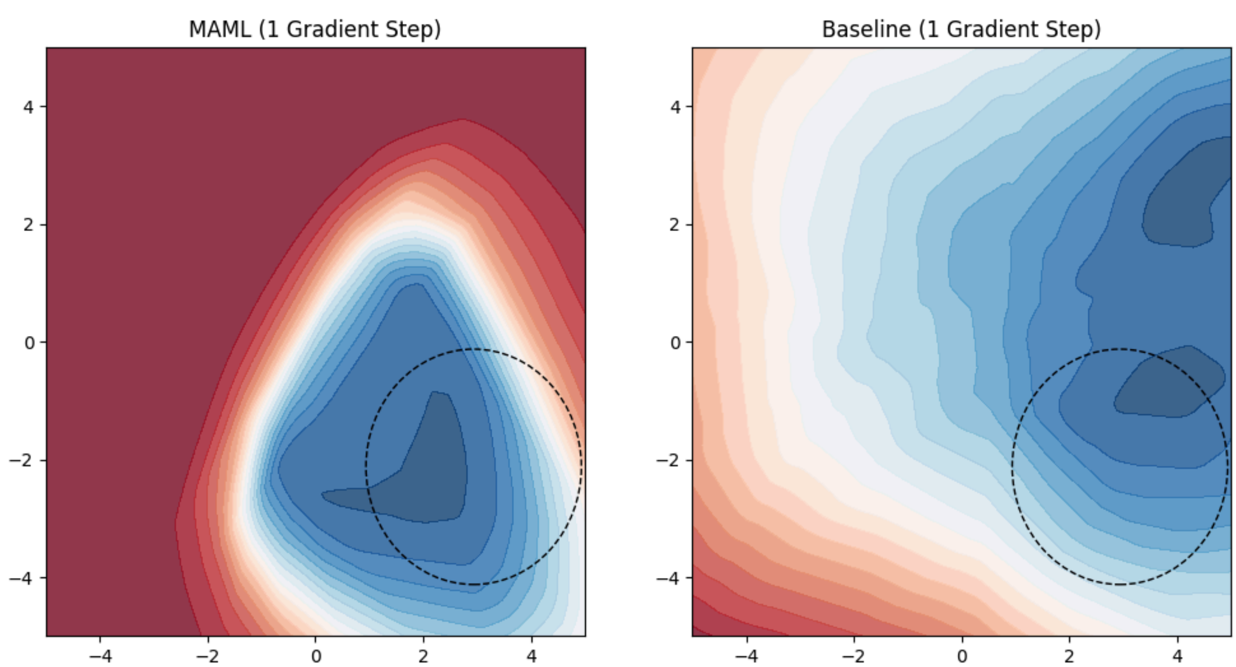
- **Training Protocol:** The same MLP architecture was trained on a mixed dataset sampled from thousands of random circle tasks simultaneously. This encourages the model to learn the "global average" of the task distribution rather than a task-specific initialization.

- **Fine-Tuning:** At test time, the pre-trained weights were fine-tuned on the test task's Support Set using standard Stochastic Gradient Descent (SGD).

3. Quantitative Evaluation

We evaluated both models on a held-out test task by measuring the classification accuracy over 10 gradient descent steps.





- **MAML Performance:** The MAML model demonstrates rapid adaptation. Starting from the meta-learned initialization, the accuracy jumps to over 90% immediately after Step 1. This indicates the initialization is positioned on the correct loss landscape manifold for circular tasks.
- **Baseline Performance:** The Baseline model starts with significantly lower accuracy (near 50-60%). While performance improves with more gradient steps, the learning curve is much flatter. Standard gradient descent struggles to reshape the "global average" decision boundary into a specific local circle using only 10 data points.

4. Qualitative Visualization

To understand the decision boundaries, we visualized the predicted probabilities of both models after exactly 1 gradient step on a random test task.

[INSERT PLOT: Heatmap Comparison HERE]

- **MAML Visualization:** The heatmap displays a sharp, well-defined circular high-probability region (blue) that aligns closely with the ground truth (dashed line). The model successfully retains the geometric concept of a "circle" while simply shifting the center coordinates.
- **Baseline Visualization:** The Baseline heatmap is diffuse and imprecise. It often produces a linear boundary or a vague region of uncertainty. Without the meta-learned prior, the model treats the 10 support points as a generic classification problem, failing to reconstruct the circular geometry.

5. Discussion & Conclusion

Why does MAML outperform the Baseline?

The results confirm that MAML successfully encodes **inductive bias** into the network initialization.

1. **The Baseline's Limitation:** The Baseline learns the expectation over all tasks. Since circle centers are uniformly distributed, the "average" task is a blurry, high-entropy region in the center of the input space. Fine-tuning this average on just 10 points leads to overfitting or slow convergence because the model must relearn the shape of the boundary from scratch.
2. **MAML's Advantage:** MAML learns the **manifold of the task parameters**. By the end of meta-training, the initialization effectively contains the knowledge that "the decision boundary is always a circle of radius 2.0." The only remaining unknown is the location. Therefore, the single gradient step does not need to learn the shape; it only performs a simple coordinate translation. This reduces the complexity of the few-shot problem, allowing MAML to solve the task with minimal data.