INTEGRATING MACHINE LEARNING AND COMPUTATIONAL METHODS INTO SCHNIRELMANN DENSITY AND ADDITIVE CLOSURE

PU JUSTIN SCARFY YANG

ABSTRACT. This paper proposes a novel integration of machine learning (ML) and computational number theory to study Schnirelmann density and additive closure phenomena. We introduce experimental heuristics, model-based density estimations, and hybrid theoretical-computational conjecture generation for sparse sets.

1. Introduction

Schnirelmann density theory is traditionally developed through combinatorial and analytic tools. However, many of its problems, such as the minimal k for additive closure, exhibit behaviors that invite computational modeling. Here we present an approach that combines data-driven pattern recognition and numerical simulations to enhance and guide theoretical research.

2. Problem Statement

Definition 2.1 (Additive Closure Threshold Function)**.** Given a set $A \subseteq \mathbb{N}$ with density $\sigma(A) > 0$, define

$$k_{\min}(A) := \min\{k \in \mathbb{N} : kA = \mathbb{N}\}.$$

This function is known only up to general bounds. We aim to estimate $k_{\min}(A)$ via ML-driven methods and simulations.

3. Computational Experiments

3.1. **Synthetic Data Generation.** We generate synthetic sets $A \subseteq [1, N]$ under various models:

- Uniform random subsets
- Fixed density with periodic gaps
- Multiplicative constructions (e.g., sets of integers with specific modular constraints)

3.2. **Feature Selection.** We extract features from sets A such as:

- Empirical density over sliding windows
- Fourier coefficients of characteristic function
- Maximum gap
- Entropy of element spacing

Date: May 5, 2025.

3.3. **Model Training.** We apply regression models:

- Linear regression for interpretability
- Random forest and XGBoost for accuracy
- Neural networks for non-linear prediction

Target: predict $k_{\min}(A)$.

4. Machine-Learning-Guided Theorems

Proposition 4.1 (Heuristic Additive Closure Law)**.** *Let $A \subseteq [1, N]$ be a random set with empirical density δ, and maximum gap g. Then:*

$$k_{\min}(A) \lessapprox \frac{\log g}{\log(1/(1-\delta))}.$$

Remark 4.2. This is a heuristic inspired by Schnirelmann's exponential growth of density under summation, calibrated through empirical simulations.

5. Software Framework

We propose the following pipeline:

(1) Generate candidate sets A
(2) Compute features and $k_{\min}(A)$ via brute-force
(3) Train regression or classification model
(4) Export model into symbolic rules and conjectures

6. Future Work

- Formalize ML-discovered heuristics into rigorous theorems
- Apply reinforcement learning to explore new classes of additive sets
- Integrate symbolic reasoning (e.g., Lean, Coq) into the computational pipeline
- Public database of additive sets and their closure statistics