Adaptive Weights for Improvements in Active Learning for Regression Using Greedy Sampling

Troy Russo
University of Washington
June 2025

Introduction

Active Learning (AL) optimizes labeling efficiency in regression problems by strategically selecting the most informative data points from an unlabeled pool. This project explored Greedy Sampling methods, specifically focusing on a novel modification called Weighted improved Greedy Sampling (WiGS).

Overview of Greedy Sampling

- **GSx (Input-space greedy):** Selects points maximally distant in input space.
- **GSy (Output-space greedy):** Picks points where model predictions differ most from true values.
- **iGS (Multiplicative approach):** Multiplies input-space and output-space distances to balance both.

Proposed Method (WiGS)

WiGS replaces the fixed multiplicative combination with a dynamic weighted additive score:

$$s_w(x_n) = (1 - w)\, z_{x_n} + w\, z_{y_n},$$

where z_{x_n} and z_{y_n} are the standardized input and output distances, respectively, and the weight w adjusts over iterations to shift emphasis between exploration and exploitation.

Empirical Findings

Experimental results using ridge regression across multiple datasets (e.g., AutoMPG, BostonHousing, Yacht) indicated only modest improvements from WiGS over GSx, GSy, and original iGS. While WiGS occasionally yielded slightly lower mean squared errors, these gains were small and not consistently significant, highlighting the limitations of manual weight tuning.

Conclusion and Future Directions

WiGS offers a flexible framework for balancing input- and output-space sampling in active learning for regression. However, manual weight schedules lack full adaptability. Future work will explore reinforcement learning—particularly Thompson sampling—to automatically tune weights on the fly, with the goal of achieving more robust performance gains.