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

Kentaro Hoffman Tyler McCormick Simon Nguyen Troy Russo

University of Washington
trusso96@uw.edu

June 6, 2025

Abstract

This project presents a novel methodological improvement to the improved Greedy Sampling (iGS) approach for active learning in regression problems. While the original iGS method by Wu *et al.* multiplicatively combines distance metrics from input space (GSx) and output space (GSy), we propose an adaptive weighting mechanism that provides greater flexibility and performance.

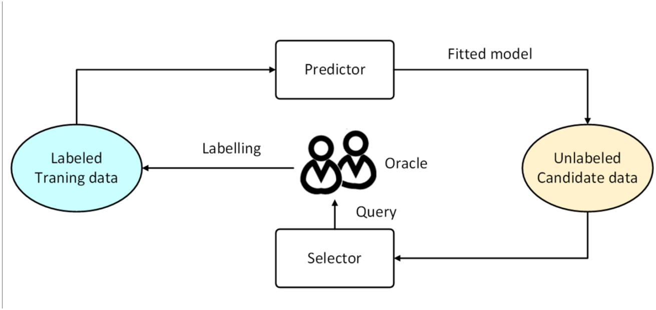
- Introduce a weighted linear combination of normalized distance metrics to account for the changing informativeness between input and output spaces at each iteration.
- Replace the fixed multiplicative approach with a dynamic weight parameter that can adjust throughout the active learning process.
- Explore several weight adaptation strategies, with future plans to select weights via reinforcement learning.

This work addresses the dynamic nature of informativeness in active learning for regression, offering a flexible framework to balance input- and output-space contributions as the model evolves.

Motivation

- **Sparse labeled data:** Only $K \ll N$ samples may be labeled—each label must be maximally informative.
- **Avoiding redundancy:** Repeated or overly similar queries fail to explore new regions, reducing efficiency.

Active Learning Loop



Existing Strategies & This Talk

Existing Strategies:

- Random sampling (baseline)
- GSx: Input-space greedy
- GSy: Output-space greedy
- iGS: Multiplicative GSx $\times$ GSy (Wu *et al.*, 2018)
- Limited work on regression post-2018

This Talk: Build off iGS by *adaptively* weighting informativeness from input and output spaces.

Notation & Setup

Unlabeled Set:

$$U_0 = \{x_n\}_{n=1}^{N} \xrightarrow[\text{add more data at iteration } k]{} U_k = \{x_n\}_{n=k+1}^{N}$$

Labeled Set:

$$L_0 = \{\} \xrightarrow[\text{append one labeled point each iteration } k]{} \mathcal{L}_k = \{(x_m, y_m)\}_{m=1}^{k}$$

Train f_k on $\mathcal{L}_k$, then select $x^* \in U_k$.

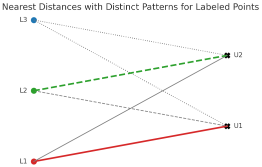
Spaces & Illustration

Unlabeled Pool as Matrix:

$$\begin{pmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{d1} & \cdots & x_{dN} \end{pmatrix} \quad [f(x_1), \ldots, f(x_N)]$$

Labeled Data & True Labels:

$$\begin{pmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{d1} & \cdots & x_{dk} \end{pmatrix} \quad [y_1, \ldots, y_k]$$



Colored/thick lines: $\min_m \|U_n - L_m\|$.

Gray lines: other pairwise distances.

Distance Matrices

$$\mathbf{D}_x = \begin{pmatrix} d_{x1}^1 & \cdots & d_{xN}^1 \\ \vdots & \ddots & \vdots \\ d_{x1}^k & \cdots & d_{xN}^k \end{pmatrix}, \quad \mathbf{d}_x = \begin{pmatrix} \min_m d_{x1}^m \\ \vdots \\ \min_m d_{xN}^m \end{pmatrix}$$

$$\mathbf{D}_y = \begin{pmatrix} d_{y1}^1 & \cdots & d_{yN}^1 \\ \vdots & \ddots & \vdots \\ d_{y1}^k & \cdots & d_{yN}^k \end{pmatrix}, \quad \mathbf{d}_y = \begin{pmatrix} \min_m d_{y1}^m \\ \vdots \\ \min_m d_{yN}^m \end{pmatrix}$$

$$\mathbf{D}_{xy} = \begin{pmatrix} d_{x1}^1\, d_{y1}^1 & \cdots & d_{xN}^1\, d_{yN}^1 \\ \vdots & \ddots & \vdots \\ d_{x1}^k\, d_{y1}^k & \cdots & d_{xN}^k\, d_{yN}^k \end{pmatrix}, \quad \mathbf{d}_{xy} = \begin{pmatrix} \min_m (d_{x1}^m\, d_{y1}^m) \\ \vdots \\ \min_m (d_{xN}^m\, d_{yN}^m) \end{pmatrix}$$

where $d_{xn}^m = \|x_n - x_m\|$, $d_{yn}^m = |f(x_n) - y_m|$, and $d_{xyn} = \min_m (d_{xn}^m\, d_{yn}^m)$.

Sampling Methods

GSx: Input-Space Greedy

$$d_{xn} = \min_m \|x_n - x_m\|, \quad x^* = \arg\max_n d_{xn}.$$

GSy: Output-Space Greedy

Train f_k, predict $\hat{y}_n$;

$$d_{yn} = \min_m |\hat{y}_n - y_m|, \quad n^* = \arg\max_n d_{yn}, \quad x^* = x_{n^*}.$$

iGS: Multiplicative

Compute

$$d_{xyn}^m = d_{xn}^m \cdot d_{yn}^m, \quad d_{xyn} = \min_{1 \le m \le k} d_{xyn}^m, \quad x^* = \arg\max_{x_n \in U_k} d_{xyn}.$$

WiGS Formulation & Schedule

Standardize distances:

$$z_{xn} = \frac{d_{xn} - \mu_x}{\sigma_x}, \quad z_{yn} = \frac{d_{yn} - \mu_y}{\sigma_y}.$$

Weighted score:

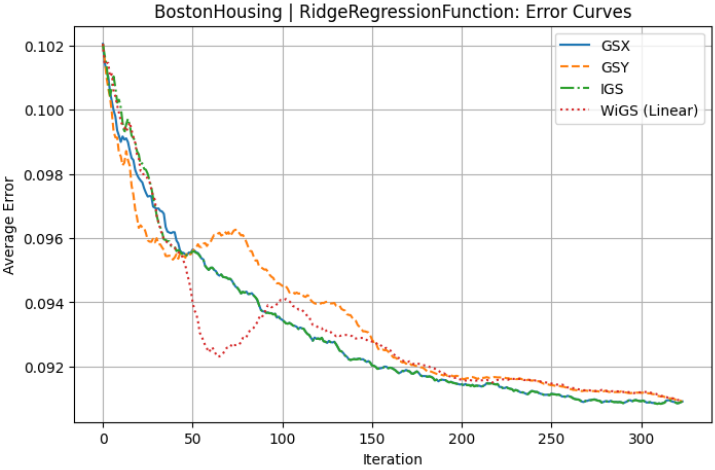
$$s_w(x_n) = (1 - w)\, z_{xn} \;+\; w\, z_{yn}, \quad w_t = \frac{t+1}{T}.$$

Select $x^* = \arg\max_{x_n \in U_k} s_w(x_n)$.

Experimental Setup

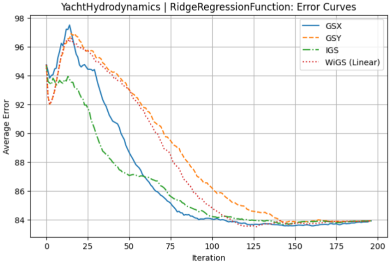
- **Datasets (9):**
  - AutoMPG, BostonHousing, Concrete*, Yacht*, WineQualityRed, ThirdData, YachtHydro, WageData
- **Model:** Ridge regression, $\alpha = 0.01$.
- **Methods:** Random, GSx, GSy, iGS, WiGS.
- **Protocol:** 100 runs, 200 iterations per run.
- **Metrics:** MSE learning curves, Wilcoxon signed-rank test ($\alpha = 0.05$).

Results: Boston Housing MSE



Other Results

WageData MSE



YachtHydro MSE

Summary of Contributions & Future Work

Our results show that the novel WiGS:

$$s_w = (1 - w)\, z_x \; + \; w\, z_y, \quad w = \frac{i+1}{T},$$

a standardized, weighted additive scoring function with a simple linear schedule, offers a promising path for balancing exploration and exploitation in active learning for regression.

- **Limitations:** A static, hand-tuned schedule can't fully adapt to changing data.
- **Future Work:**
  - Use reinforcement learning to learn optimal weight policies on the fly.
  - Extend adaptive WiGS to consistently improve performance across diverse regression settings.

Future Work: Reinforcement Learning

- Model weight selection as a multi-armed bandit problem. Each "arm" corresponds to a candidate weight w.
- Maintain a posterior distribution over w. At each iteration, sample w from the posterior (Thompson sampling).
- Define the reward as the reduction in validation MSE after querying with weight w. Use this observed reward to update the posterior.
- Thompson sampling naturally balances exploration (trying new weights) and exploitation (focusing on high-reward weights), allowing w to adapt as the regression model improves.

References

[1] Wu, D., Lin, C.-T., & Huang, J. (2019). Active Learning for Regression Using Greedy Sampling. *Information Sciences*, 474, 90–105. 10.1016/j.ins.2018.09.060