Semi-Supervised Regression with Heteroscedastic Pseudo-Labels

Xueqing Sun[1], Renzhen Wang[1*], Quanziang Wang[1], Yichen Wu[2], Xixi Jia[3], Deyu Meng[1]

[1] Xi'an Jiaotong University, [2] City University of Hong Kong, [3] Xidian University

xqsun@stu.xjtu.edu.cn, rzwang@xjtu.edu.cn

NeurIPS 2025

1. Background: Semi-Supervised Regression (SSR)

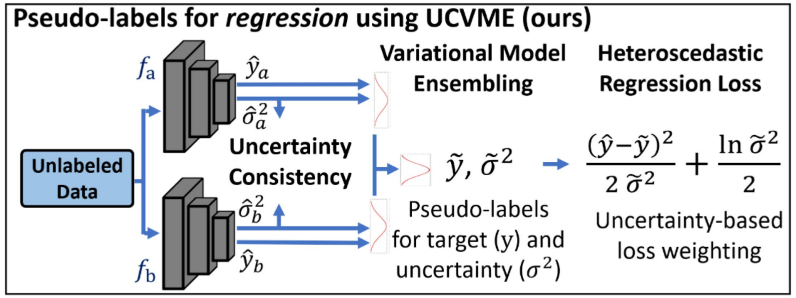
- Semi-Supervised classfication tasks: outputs are discrete and can be sharpened to encourage high-confidence predictions
  - Pseudo-labeling
  - Consistency regularization
- Semi-Supervised regression tasks: continous outputs and lack of well-defined decision boundaries
  - Consistency-based methods: design different constraints to ensure prediction consistency in SSR, such as RankUp[1], UCVME[2], CLSS[3], ...
  - Uncertainty-based methods: provide a potential way to improve the quality of pseudo-labels for unlabeled data, such as SSDKL[4], SimRegMatch[5], ...

[1] RankUp: Boosting semi-supervised regression with an auxiliary ranking classifier.In NeurIPS, 2024.
[2] Semi-supervised deep regression with uncertainty consistency and variational model ensembling via bayesian neural networks. In AAAI, 2023.
[3] Semi-supervised contrastive learning for deep regression with ordinal rankings from spectral seriation. In NeurIPS, 2023.
[4] Semi-supervised deep kernel learning: Regression with unlabeled data by minimizing predictive variance. In NeurIPS, 2018.
[5] Deep semi-supervised regression via pseudo-label filtering and calibration. In Applied Soft Computing, 2024.

1. Background: Consistency-based Methods

XI'AN JIAOTONG UNIVERSITY

- UCVME[2]: generates high-quality pseudo-labels and uncertainty estimates for heteroscedastic regression.
  - uncertainty-based loss weighting
  - variational model ensembling method



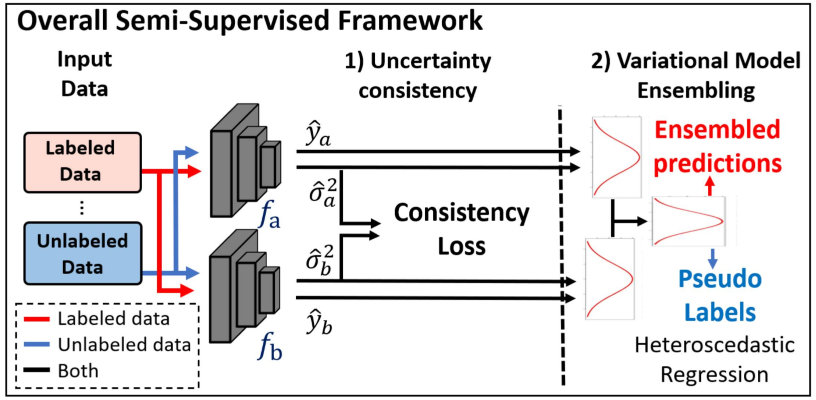
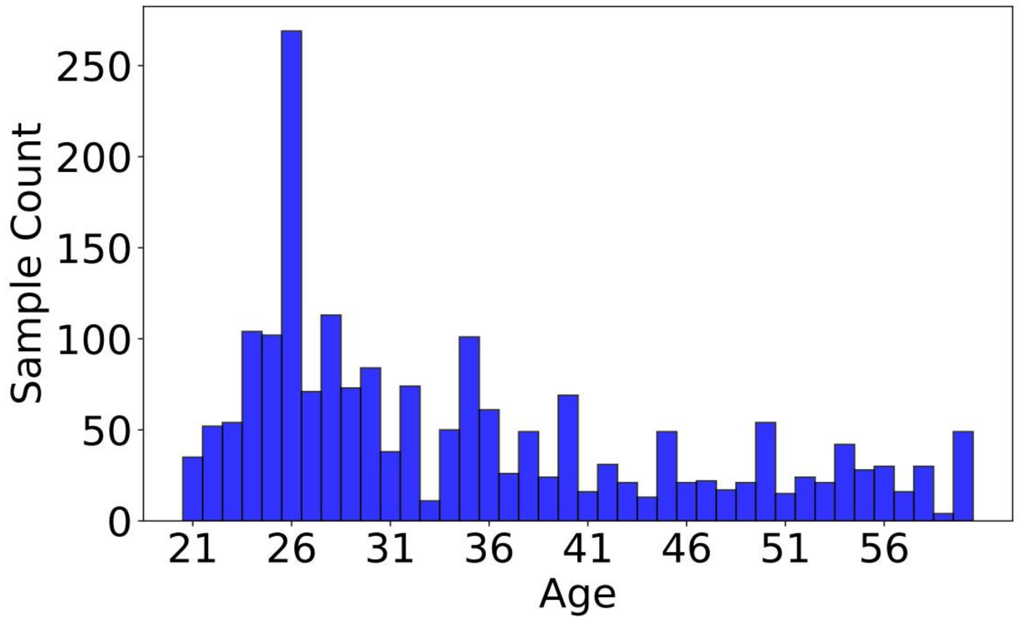


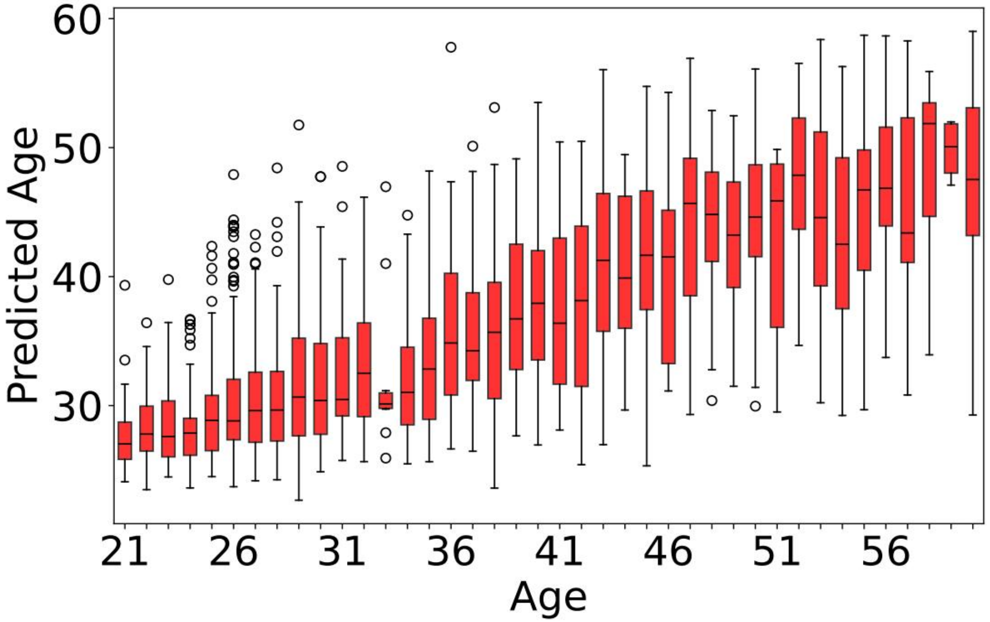
Figure 2: Semi-supervised deep regression framework for our UCVME method. UCVME improves overall pseudo-label quality and assigns greater sample weights to pseudo-labels with low uncertainty.

[2] Semi-supervised deep regression with uncertainty consistency and variational model ensembling via bayesian neural networks. In AAAI, 2023.

2. Motivation

- Motivation: assigning appropriate uncertainty values to pseudo-labels, reflecting their varying degrees of error during training.





Left: Histogram of true labels for the unlabeled data. **Right:** Box-plot of pseudo-labels generated by UCVME.

3. Method: The proposed framework

- Overall framework includes two core components
  - Heteroscedastic Pseudo-Labels: aims to learn the uncertainty or noise in these pseudo-labels
  - Bi-level Optimization: aims to learn the optimal parameters

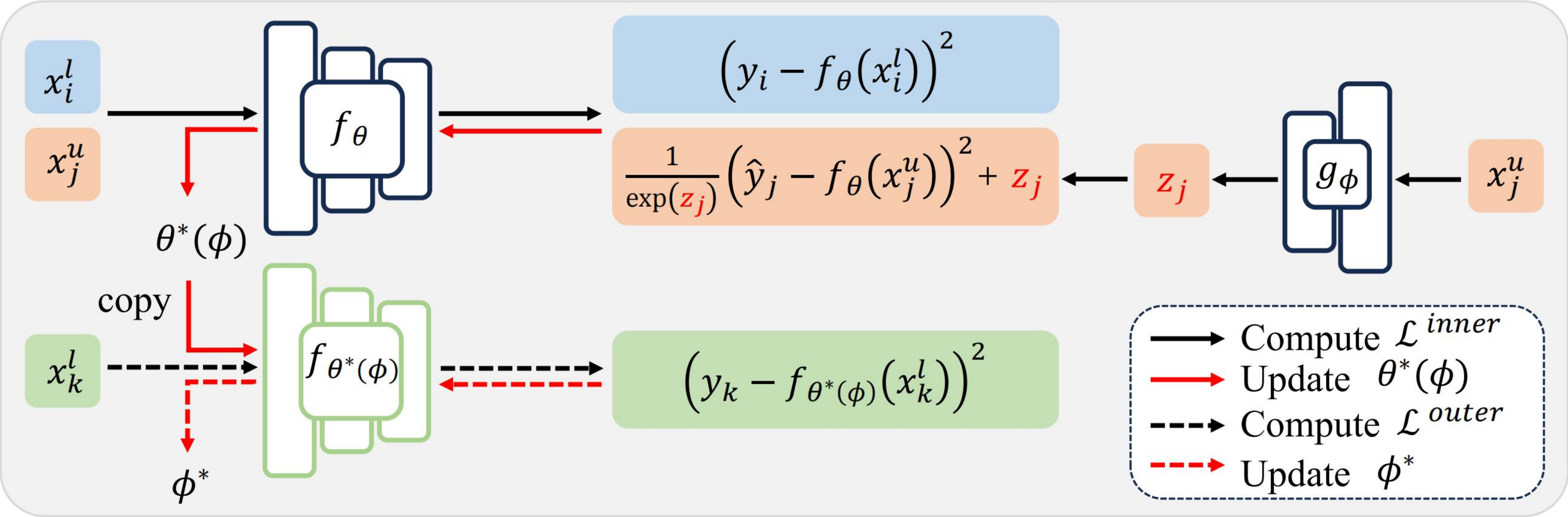


Figure 2: **Method Overview.** The proposed bi-level optimization framework consists of two main steps: (1) Inner-loop update, which updates the regression model using $\mathcal{L}^{inner}$ as defined in eq. (5), where $z_j = \log \sigma_j^2$; (2) Outer-loop update, which updates the uncertainty-learner using $\mathcal{L}^{outer}$ as defined in eq. (6). Note that we assume a batch size of 1 for better visualization.

3. Method: Heteroscedastic Pseudo-Labels

- Heteroscedastic Pseudo-Labels: treats the pseudo-labels for each sample as heteroscedastic

$$\hat{y}_j = f_\theta(x_j^u) + \epsilon_j, \quad \epsilon_j \sim \mathbb{N}(0, \sigma_j^2)$$

$$-\log p(\hat{y}_j | x_j^u) \propto \frac{\left(\hat{y}_j - f_\theta(x_j^u)\right)^2}{\sigma_j^2} + \log(\sigma_j^2)$$

$$\mathcal{L}_u = \sum_{x_j^u \in \mathcal{B}_u} \frac{1}{\sigma_j^2}\left(\hat{y}_j - f_\theta(x_j^u)\right)^2 + \sum_{x_j^u \in \mathcal{B}_u} \log(\sigma_j^2)$$

- Uncertainty-learner g_ϕ : dynamically assign uncertainty values to pseudo-labels

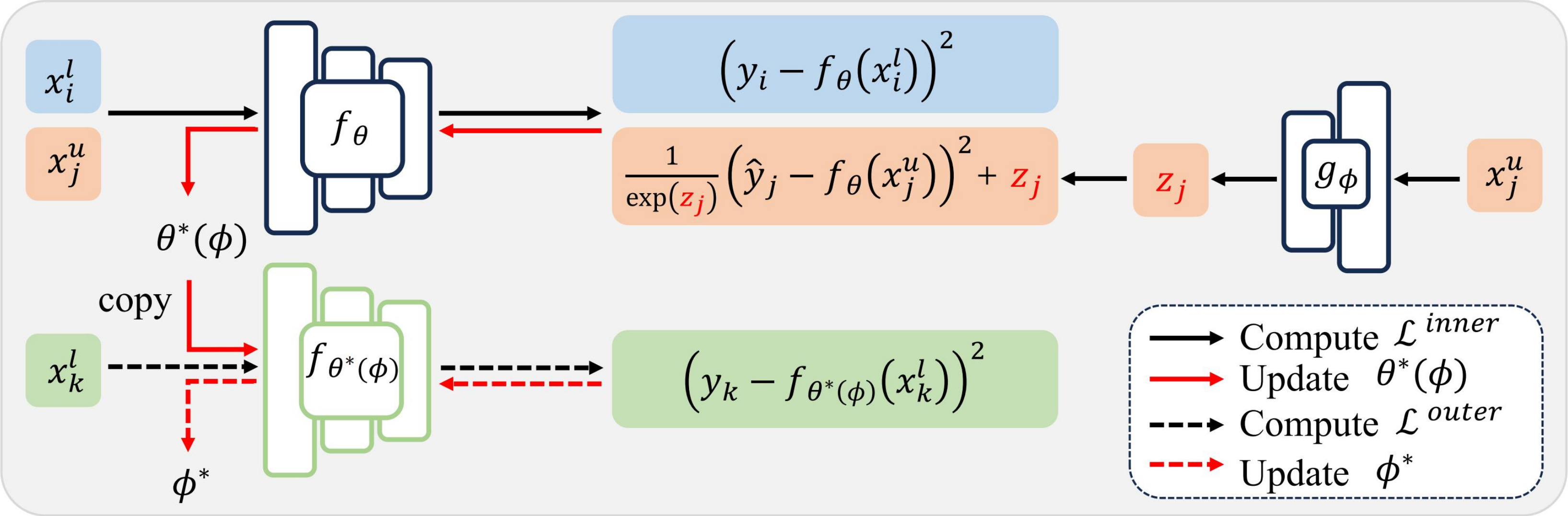
$$z_j := \log \sigma_j^2 = g_\phi(x_j^u)$$

- Jointly optimization $\{\theta, \phi\}$:

$$\{\theta^*, \phi^*\} = \arg\min_{\theta, \phi} \sum_{x_i^l \in \mathcal{B}_l} \left(y_i - f_\theta(x_i^l)\right)^2 + \lambda \sum_{x_j^u \in \mathcal{B}_u} \frac{1}{\exp(z_j)}\left(\hat{y}_j - f_\theta(x_j^u)\right)^2 + z_j$$

3. Method: Bi-level Optimization

- Bi-level Optimization:
  - Inner-loop for optimizing f_θ: using labeled and pseudo-labeled data with uncertainty to better learn sample representations.
  - Outer-loop for optimizing g_ϕ: generating well-calibrated uncertainty for f_θ, protecting it from incorrect pseudo-labels while maintaining generalization on labeled data.
  - Training algorithm:
    - update θ: $\theta^{t+1}(\phi^t) = \theta^t - \alpha \cdot \nabla_\theta \mathcal{L}^{inner}(\theta^t, \phi^t)$
    - update ϕ: $\phi^{t+1} = \phi^t - \beta \cdot \nabla_\phi \mathcal{L}^{outer}\left(\theta^{t+1}(\phi^t)\right)$



4. Experiments: IMDB-WIKI

- **IMDB-WIKI (age):** 191509(train data), 11012(test data), 11012(val data)

Method	$\gamma = 5\%$		$\gamma = 10\%$		$\gamma = 20\%$	
	MAE↓	R^2↑	MAE↓	R^2↑	MAE↓	R^2↑
Fully-Supervised	7.974 ± 0.043	0.724 ± 0.002	7.974 ± 0.043	0.724 ± 0.002	7.974 ± 0.043	0.724 ± 0.002
Supervised	10.172 ± 0.077	0.610 ± 0.004	9.248 ± 0.052	0.657 ± 0.002	8.647 ± 0.099	0.690 ± 0.005
Mean Teacher	9.492 ± 0.051	0.647 ± 0.002	8.633 ± 0.093	0.689 ± 0.002	8.191 ± 0.066	0.711 ± 0.002
Temporal Ensembling	11.335 ± 0.114	0.532 ± 0.007	9.517 ± 0.064	0.639 ± 0.004	9.577 ± 0.126	0.639 ± 0.007
SSDKL	10.116 ± 0.073	0.611 ± 0.004	9.488 ± 0.031	0.641 ± 0.002	9.056 ± 0.043	0.656 ± 0.003
TNNR	10.069 ± 0.088	0.612 ± 0.005	9.309 ± 0.052	0.654 ± 0.003	8.640 ± 0.033	0.688 ± 0.001
SimRegMatch	9.908 ± 0.097	0.628 ± 0.004	9.110 ± 0.166	0.665 ± 0.007	8.587 ± 0.094	0.693 ± 0.006
UCVME	9.730 ± 0.156	0.633 ± 0.007	8.920 ± 0.039	0.673 ± 0.004	8.309 ± 0.117	0.698 ± 0.003
CLSS	9.906 ± 0.058	0.621 ± 0.007	9.251 ± 0.107	0.656 ± 0.006	8.781 ± 0.070	0.681 ± 0.003
RankUp	10.251 ± 0.072	0.599 ± 0.005	8.836 ± 0.047	0.676 ± 0.003	8.216 ± 0.022	0.703 ± 0.001
Ours	9.177 ± 0.061	0.664 ± 0.003	8.539 ± 0.065	0.695 ± 0.003	8.166 ± 0.071	0.712 ± 0.002

Our proposed method achieves new state-of-the-art,
especially in scenarios where the labeled data is **scarce.**

4. Experiments: Ablation

- How our algorithm adjusts uncertainty

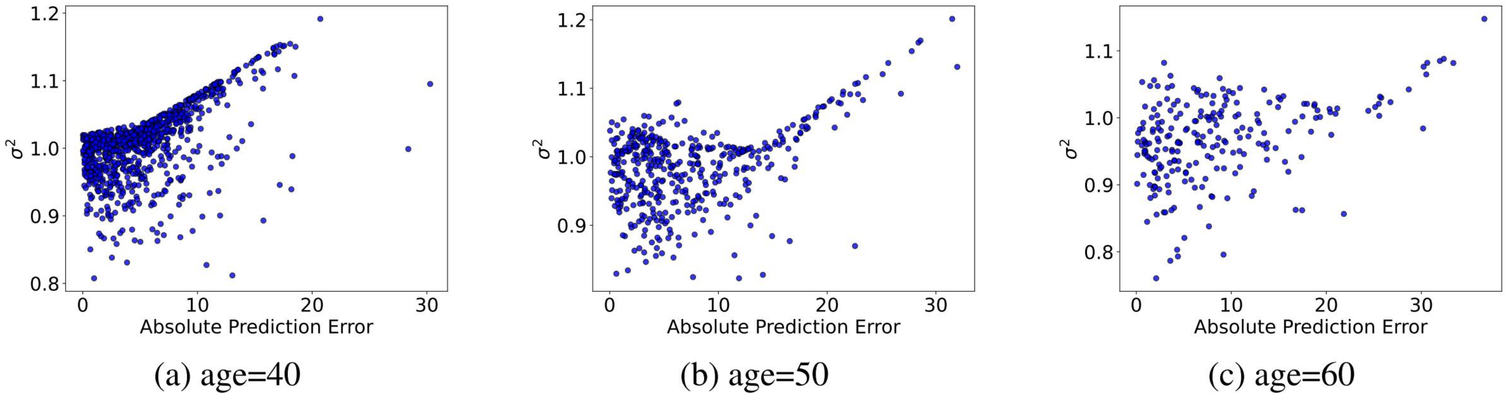


Figure 5: Correlation analysis between estimated uncertainty σ^2 and absolute prediction error on IMDB-WIKI with $\gamma = 10\%$ under different age.

- Ablation

Table4: Ablation study. BL, UL and BLO refer to Baseline, Uncertainty-learner and Bi-level Optimization.

Components			γ = 5%		γ = 10%	
BL	UL	BLO	MAE↓	R^2↑	MAE↓	R^2↑
✓	×	×	9.512	0.651	8.864	0.683
✓	✓	×	9.914	0.630	9.562	0.651
✓	✓	✓	9.177	0.664	8.539	0.695

5. Summary

- We propose an uncertainty-aware pseudo-labeling framework for SSR tasks
  - Heteroscedastic Pseudo-Labels
  - Bi-level Optimization

Paper (arXiv):

https://arxiv.org/abs/2510.15266

Code:

https://github.com/sxq11/Heteroscedastic-Pseudo-Labels

Contach us:

xqsun@stu.xjtu.edu.cn, rzwang@xjtu.edu.cn

Thanks for your attention!