

RESEARCH INTEREST

My research focuses on deep learning theory. I am interested in function approximation theory of neural networks and their statistical guarantees in learning applications. I am also interested in deep learning applications including large language models and generative models.

EDUCATION

2022 - present	Georgia Institute of Technology Ph.D. in Machine Learning
2018 - 2022	The University of Science and Technology of China (USTC) B.S. in Statistics

PUBLICATIONS

- **Robust Reinforcement Learning from Corrupted Human Feedback**
Alexander Bukharin, Ilgee Hong, Haoming Jiang, Zichong Li, Qingru Zhang, **Zixuan Zhang**, and Tuo Zhao. *Accepted by the 38th Annual Conference on Neural Information Processing Systems (NeurIPS 2024)*.
- **Nonparametric Classification on Low Dimensional Manifolds using Overparameterized Convolutional Residual Networks**
Zixuan Zhang^{*}, Kaiqi Zhang^{*}, Minshuo Chen, Mengdi Wang, Tuo Zhao and Yu-Xiang Wang. *Accepted by the 38th Annual Conference on Neural Information Processing Systems (NeurIPS 2024)*.
- **Effective Minkowski Dimension of Deep Nonparametric Regression: Function Approximation and Statistical Theories**
Zixuan Zhang, Minshuo Chen, Wenjing Liao and Tuo Zhao. *Accepted by the 40th International Conference on Machine Learning (ICML 2023)*.
- **Sequential Information Design: Markov Persuasion Process and Its Efficient Reinforcement Learning**
Jibang Wu^{*}, **Zixuan Zhang**^{*}, Zhe Feng, Zhaoran Wang, Zhuoran Yang, Michael I. Jordan, and Haifeng Xu. *In Proceedings of the 23rd ACM Conference on Economics and Computation (EC '22)*. Association for Computing Machinery, New York, NY, USA, 471-472.

RESEARCH EXPERIENCES

2024.3 - 2024.6	Robust Reinforcement Learning from Corrupted Human Feedback: Proposed a robust RLHF approach R^3M , which models the potentially corrupted preference data as sparse outliers. Formulated the robust reward learning as an regularized maximum likelihood estimation problem. Developed an efficient alternating optimization algorithm. Proved that R^3M can consistently learn the underlying reward and identify outliers. Conducted experiments on robotic control and natural language generation with large language models (LLMs), showing that R^3M improves robustness of the reward against several types of perturbations to the preference data.
-----------------	---

- 2023.6 - 2024.10 **Score approximation, estimation and distribution recovery of diffusion models on low-dimensional manifolds:** Develop a novel score decomposition as on-manifold score and projection term. Approximate and effectively estimate the score function with a properly chosen neural network architecture. Show that the generated distribution based on the estimated score function captures the data geometric structures and converges to the data distribution with mini-max optimal rate that depends on the intrinsic dimension.
- 2023.3 - 2023.4 **Natural Language Processing with Disaster Tweets:** Implemented four models, logistic regression, LSTM, Temporal Convolutional Networks, and BERT, to predict whether a real disaster is occurring based on the content of tweets. Investigated three parameter-efficient techniques, BitFit, Adapter Tuning, and LoRA, to reduce the computational resources for fine-tuning the pre-trained BERT.
- 2023.2 - 2023.5 **Nonparametric Classification using Overparameterized Convolutional Residual Networks:** Studied the performance of ConvResNeXts trained with weight decay from the perspective of nonparametric classification. Developed an approximation theory and statistical estimation results of overparameterized ConvResNeXts for Besov functions. Proved that ConvResNeXts can adapt to the function smoothness and low dimensional structures and efficiently learn the function without suffering from the curse of dimensionality. Conducted experimental studies of ConvResNets for ImageNet classifications with different resolutions to illustrate our theoretical findings.
- 2022.11 - 2023.1 **Effective Minkowski Dimension of Deep Nonparametric Regression:** Defined a new complexity notation, effective Minkowski dimension, to characterize intrinsic dimension of the data support. Established an approximation theory of deep neural networks for Holder functions. Derived the sample complexity that only depends on the effective Minkowski dimension. Illustrated our theoretical findings by considering nonparametric regression with an anisotropic Gaussian random design.
- 2021.9 - 2022.3 **Reinforcement Learning of Markov Persuasion Process:** Studied the Markov Persuasion Process (MPP) where a signal sender seeks to persuade a stream of myopic receivers to take actions that maximize the sender's cumulative utilities. Designed an online learning algorithm to learn the optimal signaling policy by interacting with the underlying MPP. Robustified the signaling policy with respect to the uncertainty of prior estimation to prevent receivers' detrimental equilibrium behavior. Deserved an upper regret bound of order $O(H^2\sqrt{T})$.
- 2021.7 - 2021.10 **Molecule Structure Recovery Based on Manifold Learning Methods:** Improved a new method of manifold learning which iteratively fitted the intuitive manifold and projected data points. Conducted simulation studies by rotating different parts of a certain molecule to obtain a manifold distributed dataset and verify the accuracy of the proposed algorithm. Applied our method on real molecule data and concluded that our method not only denoised data well but also maintained the intuitive structure of original data. Compared the results of different methods, including local-PCA, KDE, log-KDE, Mfit, and Moving Least Square, and reduced RMSE by more than 10%.

AWARDS

- Stewart Topper Fellowship, *Aug. 2022 & Dec. 2022*
- ISyE Fellowship, *Aug. 2022*
- Keshaosihua Scholarship, *Oct. 2021*
- Huawei Scholarship, *Oct. 2020*
- USTC Outstanding student scholarship, *Oct. 2019 & Oct. 2018*
- Second Prize in the 11th Chinese Mathematics Competitions, *Nov. 2019*

TEACHING EXPERIENCES

Teaching Assistant: Computational data analysis, Fall 2023

INDUSTRIAL EXPERIENCES

Research Scientist Intern at Microsoft, Summer 2024 & Fall 2024

SKILLS

Programming: Python, R, Matlab, C