Mingyuan "William" **Zhang**

myz@seas.upenn.edu | www.mingyuanzhang.com | moshimowang | mingyuanzhang | Mingyuan Zhang

Summary

- Ph.D. student researcher working on **machine learning** and **artificial intelligence**.
- Conduct independent research on how to **design good algorithms with performance guarantees from first principles** for various machine learning problems.
- Methodologies include probability, statistics, machine learning, optimization, and programming.
- Produced **7** manuscripts (**6** of them as the lead author); published **5** of them in major machine learning and natural language processing conferences, including **top-tier** conferences such as **ICML**, **NeurIPS**, and **EMNLP**.
- Looking for **full-time job** opportunities to apply principled insights to solve impactful real-world problems in machine learning, data science, and/or quantitative research.

Education

Ph.D. in Computer and Information Science. *2018 - 2024 (Expected)*

University of Pennsylvania, Philadelphia, Pennsylvania, USA *GPA: 4.00/4.00*

- Advisor: Prof. Shivani Agarwal. Thesis: Statistical Machine Learning for Complex Classification Problems.

B.S. in Honors Mathematics, Honors Statistics, Computer Science, and Data Science. *2018*

University of Michigan, Ann Arbor, Michigan, USA *GPA: 3.92/4.00*

- Highest Honors in Mathematics and Honors in Statistics.

Research

Statistical Machine Learning for Complex Classification Problems. *Aug. 2018 - Present*

Design good algorithms with performance guarantees. Advised by Prof. Shivani Agarwal. *University of Pennsylvania*

- Design algorithms from first principles for machine learning problems involving additional complexities beyond standard settings; these include multi-label classification, learning with noisy labels, weakly supervised learning, optimizing complex performance measures, classification with a specific function class, and others.
- Prove performance guarantees for the designed algorithms to show the performance improves with more data.
- Design and execute experiments using tools such as scikit-learn, TensorFlow, and PyTorch in order to validate the theories and benchmark the proposed algorithms against others.
- Produced 5 first-authored manuscripts; published 4 of them in major conferences on machine learning, including top-tier conferences such as ICML and NeurIPS.

Foreseeing the Benefits of Incidental Supervision. *Nov. 2019 - June 2020*

Measure the quality of incidental/weak supervision signals. Advised by Prof. Dan Roth. *University of Pennsylvania*

- We proposed a unified measure to quantify the benefits of various incidental/weak signals for a given target task.
- We published this work in EMNLP 2021, a top-tier conference on natural language processing.

Publications

Multiclass Learning from Noisy Labels for Non-decomposable Performance Measures. *[Link]*

Mingyuan Zhang, Shivani Agarwal.

In *Proceedings of the 27th International Conference on Artificial Intelligence and Statistics* (**AISTATS**), 2024.

Foreseeing the Benefits of Incidental Supervision. *[Link]*

Hangfeng He, **Mingyuan Zhang**, Qiang Ning, Dan Roth.

In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing* (**EMNLP**), 2021.

Learning from Noisy Labels with No Change to the Training Process. *[Link]*

Mingyuan Zhang, Jane Lee, Shivani Agarwal.

In *Proceedings of the 38th International Conference on Machine Learning* (**ICML**), 2021.

Bayes Consistency vs. H-Consistency: The Interplay between Surrogate Loss Functions and the Scoring Function Class. *[Link]*

Mingyuan Zhang, Shivani Agarwal.

In *Advances in Neural Information Processing Systems* (**NeurIPS**), 2020. **Spotlight paper**.

Convex Calibrated Surrogates for the Multi-Label F-Measure. *[Link]*
Mingyuan Zhang, Harish G. Ramaswamy, Shivani Agarwal.
In *Proceedings of the 37th International Conference on Machine Learning* (**ICML**), 2020.

Multi-Label Learning from Noisy Labels.
Mingyuan Zhang, Shivani Agarwal.
Under review.

On the Minimax Regret in Online Ranking with Top-k Feedback. *[Link]*
Mingyuan Zhang, Ambuj Tewari.
Preprint, 2023.

Multiclass and Multi-Label Learning with General Losses.
H. G. Ramaswamy, **M. Zhang**, B. S. Babu, S. Agarwal, A. Tewari, R. C. Williamson.
In preparation.

Skills

Programming Languages	Python, C++, Matlab, R.
Machine Learning Libraries	scikit-learn, TensorFlow, PyTorch, Hugging Face, OpenAI API.
Natural Languages	English, Mandarin Chinese.

Courses

Graduate level: Real Analysis (A), Probability Theory (A), Discrete Stochastic Processes (A), Numerical Linear Algebra (A+), Combinatorial Theory (A+), Complex Variables (A), Applied Functional Analysis (A), Nonlinear Programming (A+), Statistical Inference (A), Linear Models (A), Analysis of Multivariate and Categorical Data (A), Statistical Learning (A), Time Series Analysis (A-), Machine Learning (A), Information Theory (A+), Statistical Signal Processing (A).

Undergraduate level: Intermediate Microeconomics Theory (A+), Intermediate Macroeconomics Theory (A), Game Theory (A+), Theoretical Statistics (A+), Statistical Computing Methods (A+), Numerical Methods (A+), Programming and Data Structures (A+), Data Structures and Algorithms (A+), Algorithms (A), Randomized Algorithms (A+), Database Management Systems (A), Computer Vision (A), Information Retrieval (A).

Honors, Awards, Certificates

2017-2018	**Outstanding Achievement in Mathematics Awards**, University of Michigan	*USA*
2023	**Deep Learning Specialization**, DeepLearning.AI	*Coursera*
2023	**DeepLearning.AI TensorFlow Developer Specialization**, DeepLearning.AI	*Coursera*
2023	**TensorFlow: Advanced Techniques Specialization**, DeepLearning.AI	*Coursera*
2023	**Natural Language Processing Specialization**, DeepLearning.AI	*Coursera*
2023	**Machine Learning Engineering for Production (MLOps) Specialization**, DeepLearning.AI	*Coursera*
2024	**Generative AI with Large Language Models**, DeepLearning.AI	*Coursera*

Teaching & Service

Head Teaching Assistant — Machine learning. *Spring of 2020, 2021, 2022*
CIS 520: A graduate-level machine learning course. *University of Pennsylvania*

- Managed the TA team ($\sim$ 10 TAs) for a large class ($\sim$ 100 students) to help students learn effectively, held weekly office hours, made homework and exam questions, and administrated and graded exams.

Grader — Various math courses. *2015 - 2018*
Various linear algebra and probability courses. *University of Michigan*

- Graded weekly assignments for six (undergraduate and graduate level) linear algebra and probability courses.

Reviewers — Various machine learning conferences and journals. *2021 - 2024*
NeurIPS, ICLR, AISTATS, JMLR, IEEE PAMI.

- Reviewed more than 30 machine learning manuscripts in total.