

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** for various machine learning problems. Methodologies include probability, statistics, optimization, and programming.
- Produced **7** manuscripts (**6** of them as the lead author); **4** of them have been published in **top-tier** machine learning and natural language processing conferences, including **ICML**, **NeurIPS**, and **EMNLP**.

Education

Ph.D. in Computer and Information Science.

2018 - 2024

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 for machine learning problems involving additional complexities beyond standard settings; these include multiclass classification with a specific function class, multi-label classification, learning with noisy labels, weakly supervised learning, and optimizing complex performance measures.
- 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 algorithms against others.
- Five manuscripts have been produced; three of them have been published in top-tier conferences on machine learning, including ICML and NeurIPS.

Foreseeing the Benefits of Incidental Supervision.

Nov. 2019 - June 2020

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

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

Learning to Rank with Top-k Feedback.

Sep. 2017 - Aug. 2018

Rank items in a streaming fashion with partial feedback. Advised by Prof. Ambuj Tewari. University of Michigan

- Derived minimax regret rates for several widely used performance measures in learning to rank with partial feedback settings; these regret rates provided theoretical insights into the problem's difficulty.
- Proposed an efficient algorithm that achieved the minimax regret rate for the Precision@n measure.
- A manuscript was produced based on this work.

Publications

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.

Multiclass Learning from Noisy Labels for Non-decomposable Performance Measures.

Mingyuan Zhang, Shivani Agarwal.

Under review.

Multi-Label Learning from Noisy Labels.

Mingyuan Zhang, Shivani Agarwal.

Under review.

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.

On the Minimax Regret in Online Ranking with Top-k Feedback.

[Link]

Mingyuan Zhang, Ambuj Tewari.

Preprint.

Skills

Programming Languages Python, C++, Matlab, R.
Machine Learning Libraries scikit-learn, TensorFlow, PyTorch.
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	Natural Language Processing Specialization , DeepLearning.AI	Coursera
2023	Machine Learning Engineering for Production (MLOps) Specialization , DeepLearning.AI	Coursera
2023	Practical Data Science on the AWS Cloud Specialization , 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 (~ 10 TAs) for a large class (~ 100 students), held weekly office hours, made homework and exam questions, and administrated and graded exams in order to help students learn effectively.

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.

2021-2023

NeurIPS, ICLR, AISTATS, JMLR and IEEE PAMI.