Mingyuan "William" **Zhang**

PhD student researcher focusing on Machine Learning and Artificial Intelligence.

3401 Walnut Street, office 454C, Philadelphia, PA, USA

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

Research Interests

Machine learning and artificial intelligence, multiclass and multi-label classification, learning with noisy labels, and weakly supervised learning. I am looking for opportunities to apply theoretical insights to solve real-world problems.

Education

Ph.D. in Computer and Information Science. *2018 - Present*

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

- Advisor: Prof. Shivani Agarwal.
- Thesis: Statistical Machine Learning for Complex Classification Problems.
- Committee: Prof. Weijie Su (chair), Prof. Jake Gardner, Prof. Surbhi Goel, and Prof. Ambuj Tewari.

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.
- Honors in Statistics.

Publications

Preprints and Other Manuscripts.

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

[6] Multiclass Learning from Noisy Labels for Non-decomposable Performance Measures.
Mingyuan Zhang, Shivani Agarwal.
Under review.

[5] On the Minimax Regret in Online Ranking with Top-k Feedback.
Mingyuan Zhang, Ambuj Tewari.
Under review.

Conference Papers.

[4] 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. **Oral paper**.

[3] 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.

[2] 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**.

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

Skills

Programming Languages	Python, C++, Matlab, R.
Machine Learning Libraries	scikit-learn, PyTorch, TensorFlow.
Natural Languages	Mandarin Chinese (native proficiency), English (full professional proficiency)

Primary Research

Multiclass Learning from Noisy Labels Using Weighted Losses. *Feb. 2023 - Present*
Generalize the weighted loss method to multiclass settings to handle noisy labels. *University of Pennsylvania*

- Advised by Prof. Shivani Agarwal.
- We study how to generalize the weighted loss method proposed in Natarajan et al. (2013) from binary classification to multiclass settings. Doing so allows us to adapt multiclass versions of margin-based losses (including variants of hinge losses) to deal with noisy labels while preserving their convexity.
- I show how the multiclass weighted loss method can recover some methods in classification with a reject option.

Multiclass Learning from Noisy Labels for Non-decomposable Performance Measures. *May 2022 - Jan. 2023*
Optimize non-decomposable performance measures with noisy labels. *University of Pennsylvania*

- Advised by Prof. Shivani Agarwal.
- We study noisy label problems for *non-decomposable* performance measures which cannot be expressed as the expectation or sum of a loss on individual examples; such performance measures are defined by general (usually *non-linear*) functions of the confusion matrix of a classifier.
- We propose noise-corrected algorithms for two broad families of non-decomposable performance measures (monotonic convex and ratio-of-linear) for general multiclass class-conditional noise models.
- We also provide regret bounds for the algorithms, showing that they are consistent.
- Our experiments (implemented using scikit-learn and TensorFlow) demonstrate algorithms' effectiveness.

Multi-Label Learning from Noisy Labels. *Jul. 2021 - Apr. 2022*
Consistent and efficient learning from noisy labels in multi-label classification. *University of Pennsylvania*

- Advised by Prof. Shivani Agarwal.
- We study how to learn from noisy labels in multi-label problems.
- We propose a family of *Noise-Corrected Output Coding* algorithms for general low-rank multi-label losses (that include the common Hamming loss and F_1-measure as special cases) that apply to a broad family of class-conditional noise models, a widely studied class of noise models in multiclass settings.
- We show our algorithms are consistent and provide a quantitative regret transfer bound.
- We also propose an *In-Group Switching Noise* model for multi-label settings that naturally captures some of the correlation of noise among labels (tags).
- Our experiments (implemented using PyTorch) confirm our theoretical findings.

Learning from Noisy Labels with No Change to the Training Process. *Sep. 2020 - Jun. 2021*
Learn from class-conditional noisy labels without changing the training process. *University of Pennsylvania*

- Advised by Prof. Shivani Agarwal.
- We design a provably consistent algorithm for learning from class-conditional noisy labels for general cost-sensitive multiclass losses without changing the training process (previous methods require changing the training loss).
- We also give a quantitative regret transfer bound for our method by using strongly proper composite multiclass losses (for example, the composition of the cross entropy loss and the softmax function used in deep networks).
- We provide fixes and potential improvements for noise estimation methods that involve computing anchor points.
- We compare the performance of our plug-in method with two previously proposed methods on the MNIST and CIFAR10 data sets (experiments were implemented using TensorFlow). Our plug-in method is comparable to the other two methods, even though it requires no change to the training process.
- The manuscript of this work was published in ICML 2021.

Bayes Consistency vs. H-Consistency: The Interplay between Surrogate Loss Functions and the Scoring Function Class. *Mar. 2020 - Aug. 2020*

Explore the relationship between Bayes consistency and $\mathcal{H}$-consistency. *University of Pennsylvania*

- Advised by Prof. Shivani Agarwal.
- We investigate the conundrum posed in Long and Servedio, (2013) regarding Bayes consistency and $\mathcal{H}$-consistency.
- Our findings show it is not just the surrogate loss that is important, but the interplay between the surrogate loss and the class of scoring functions over which the loss is minimized also matters.
- The manuscript of this work was published in NeurIPS 2020 as a spotlight paper.

Convex Calibrated Surrogates for the Multi-Label F-Measure. *Sep. 2019 - Feb. 2020*

How to design a good algorithm to optimize the multi-label F-measure. *University of Pennsylvania*

- Advised by Prof. Shivani Agarwal.
- We show a provably consistent algorithm for the multi-label F-measure.
- Our algorithm is based on the theory of convex calibrated surrogate losses.
- We also show a quantitative regret transfer bound for the surrogate loss we designed.
- We observe good practical performance in our experiments (implemented using PyTorch).
- The manuscript of this work was published in ICML 2020.

Design of Calibrated Surrogates for Multiclass and Multi-Label Learning Problems. *Sep. 2018 - Aug. 2019*

How to design good algorithms for multiclass and multi-label learning problems. *University of Pennsylvania*

- Advised by Prof. Shivani Agarwal.
- We study how to design consistent and efficient algorithms for general multiclass and multi-label learning problems.
- I designed and ran both synthetic and real data experiments (using PyTorch) to benchmark the proposed algorithms with popular algorithms in multi-label and label ranking settings.
- I proved lower and upper bounds for the calibration functions of the proposed algorithms.
- The journal version of this work is in preparation.

Academic Service

Reviewers.

Various conferences and journals.

- Conference on Neural Information Processing Systems (NeurIPS), 2021-2023.
- International Conference on Learning Representations (ICLR), 2022-2023.
- Journal of Machine Learning Research (JMLR).
- Transactions on Pattern Analysis and Machine Intelligence (PAMI).

Teaching

Head Teaching Assistant — Machine Learning. *Spring of 2020, 2021, 2022*

CIS 520: A graduate level machine learning course. *University of Pennsylvania*

- Managed TA group ($\sim$ 10 TAs) for a large class ($\sim$ 100 students), held office hours, made homework and exam questions, and graded exams.

Grader — Various Math Courses. *2015 - 2018*

Various linear algebra and probability courses. *University of Michigan*

- Graded weekly assignments for six different levels of linear algebra and probability courses.

Tutor — Linear Algebra. *May 2015 - Jun. 2015*

MATH 217: An introductory linear algebra course. *University of Michigan*

- Assisted students with linear algebra and rigorous proofs.

Courses

Graduate level: Measure Theory (A), Probability Theory (A), Stochastic Processes (A), Numerical Linear Algebra (A+), Combinatorial Theory (A+), Complex Variables (A), Applied Functional Analysis (A), Nonlinear Programming (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 Microeconomic Theory (A+), Intermediate Macroeconomics Theory (A), Game Theory (A+), Theoretical Statistics (A+), Statistical Computing Methods (A+), Programming and Data Structures (A+), Algorithms (A), Database Management Systems (A), Computer Vision (A), Information Retrieval (A)

Honors & Awards

2017-2018 **Outstanding Achievement in Mathematics Awards**, University of Michigan *USA*

2017-2018 **James B. Angell Scholar**, University of Michigan *USA*

2015 **James B. Angell Scholar**, University of Michigan *USA*

2014 **William J. Branstrom Freshman Prize**, University of Michigan *USA*

2013-2018 **University Honors**, University of Michigan *USA*

Secondary Research

Learning to Rank with Top-k Feedback. *Sep. 2017 - Aug. 2018*

How to rank items in a streaming fashion with partial feedback. *University of Michigan*

- Advised by Prof. Ambuj Tewari at the Department of Statistics.
- Worked on online learning to rank with top-k feedback.
- Derived minimax regret rates for several widely used performance measures in learning to rank settings.
- A manuscript of this work was produced.

Independent Study in High-Dimensional Statistics. *Sep. 2017 - Dec. 2017*

What to do when the number of features is larger than the sample size. *University of Michigan*

- Advised by Prof. Gongjun Xu at the Department of Statistics.
- Learned and compared different variable selection and statistical inference methods commonly used in high-dimensional statistics by reading selected papers.
- Ran simulations in R.

Implementation and Testing of Subspace Clustering Algorithms. *May 2017 — Jun. 2017*

Part of the SPIDER project: Subspace Primitives that are Interpretable and DivERse. *University of Michigan*

- Advised by Prof. Laura Balzano and Prof. Jason Corso at EECS Department.
- Implemented K-Subspaces and Ensemble K-Subspaces packages in Python.
- Tested the implementations on Extended Yale Face Database B and Hopkins 155 Dataset.

Ordinal Embedding with a Latent Factor Model. *May 2016 — Aug. 2016*

Improve convergence of a non-metric multidimensional scaling algorithm. *University of Michigan*

- Advised by Prof. Laura Balzano at EECS Department.
- Improved convergence of Aura, a non-metric multidimensional scaling algorithm.
- Wrote Matlab code to implement the improved algorithm and to run large-scale simulations on Flux (a high-performance computing cluster).
- Produced a manuscript describing the improved algorithm and the results.

Comparison of Subspace Estimation Algorithms. *May 2015 — Aug. 2015*

Compare different subspace estimation and tracking algorithms. *University of Michigan*

- Advised by Prof. Laura Balzano at EECS Department.
- Studied GROUSE (a subspace estimation and tracking algorithm).
- Compared GROUSE with ISVD and PIMC (two other subspace estimation and tracking algorithms) in terms of convergence rate and computational complexity by running the algorithms on both synthetic and real data in Matlab.

Brain Network and Consciousness. *Sep. 2014 — Aug. 2015*

Explore better ways of displaying results from brain network simulations. *University of Michigan*

- Advised by Prof. Uncheol Lee at the Department of Anesthesiology.
- Wrote Matlab code to run simulations of brain network to study consciousness.
- Processed experimental data and made figures to display results using Pajek, Gephi and Inkscape.