Lijia Zhou

zlj@uchicago.edu • zhoulijia.github.io

Education

University of Chicago

Ph.D. candidate in Statistics *2018 – Expected 2023*

- Advisor: Prof. Nathan Srebro
- William Rainey Harper Dissertation Fellowship
  - One of the highest honors given to a PhD candidate at the University of Chicago

B.S. in Applied Mathematics & Statistics *2015 – 2018*

University of California Los Angeles

Major in Mathematics, College Honors *2014 – 2015*

Research

I am broadly interested in statistical learning theory. In particular, I have been thinking about high dimensional interpolation, benign overfitting in generalized linear models, and implicit regularization.

Publications

A Non-Asymptotic Moreau Envelope Theory for High-Dimensional Generalized Linear Models
with Frederic Koehler, Pragya Sur, Danica J. Sutherland and Nathan Srebro

- Under review.

Optimistic Rates: a Unifying Theory for Interpolation Learning and Regularization in Linear Regression
with Frederic Koehler, Danica J. Sutherland and Nathan Srebro

- Under review. Available on ArXiv.

Uniform Convergence of Interpolators: Gaussian Width, Norm Bounds and Benign Overfitting
with Frederic Koehler, Danica J. Sutherland and Nathan Srebro

- published at *Conference on Neural Information Processing Systems (NeurIPS) 2021*
- Oral (top **0.6%** of 9122 submissions)

On Uniform Convergence and Low-Norm Interpolation Learning
with Danica J. Sutherland and Nathan Srebro

- published at *Conference on Neural Information Processing Systems (NeurIPS) 2020*
- Spotlight (top **2.9%** of 9454 submissions)

Academic Projects

Statistical consulting at UChicago

- *Higher-order-thinking (HOTT) talk in parent-child interaction*
  - Work with researcher in the Psychology department and propose multiple solutions for modeling HOTT utterance, such as Poisson rate model or Beta Generalized linear model with random effects
- *Medication discrepancies and blood pressure control in Botswana hypertension clinics*
  - Work with researcher in the medicine school to estimate the effect of medication discrepancy on systolic and diastolic blood pressure by fitting a bivariate multiple linear regression

Machine learning course projects

- *Paraphrase identification*
  - Train a bidirectional Long Short Term Memory (LSTM) model on a training set that contains over 1 million examples for determining whether a pair of sentences is a paraphrase
  - Experiment with different hyperparameter choices and the usage of pre-trained word embeddings, and achieve over 95% accuracy in the testing data set
- *Image generation*
  - Implement multiple deep generative models for images on MNIST, including the Variational AutoEncoder (VAE), two-stage VAE, Generative Adversial Network (GAN) and Wasserstein GAN

Undergraduate research

- *Spatial-Temporal Modeling of Ozone Monitoring Instrument (OMI) data*
  - Explore the application of Recursive Skeletonization Factorization techniques to evaluate the log-likelihood of various parametrized models in MATLAB, including rotational Gaussian processes with a powered-exponential kernel or Matérn kernel
  - Successfully fit a dataset composed of 64 orbits of OMI data, which has over 70,000 observations, in a standard laptop
- *Spectral Clustering On Ratios-of-Eigenvectors (SCORE)*
  - Write an R program that implements the SCORE algorithm for communities detection, the Mixed-SCORE algorithm for undirected mixed membership network estimation, and a Singular Value Decomposition method for topic estimation in text

Work Experience

Citadel New York, NY

Quantitative Researcher Internship *Summer 2022*

- Conduct research and statistical analyses in equity index options and bond future options
- Work with large data sets to predict and test statistical market patterns
- Back-test and implement options trading models and signals

Talks

1. A Non-asymptotic Generalization Theory for Over-parameterized Generalized Linear Models, *Student seminar (UChicago Statistics)*, March 2022
2. Uniform Convergence of Interpolators: Gaussian Width, Norm Bounds and Benign Overfitting, *Conference on Neural Information Processing Systems (NeurIPS 2021)*, December 2021
3. On Uniform Convergence and Low-Norm Interpolation Learning, *Collaborations on the Mathematical and Scientific Foundations of Deep Learning (MoDL)*, March 2021
4. On Uniform Convergence and Low-Norm Interpolation Learning, *Conference on Neural Information Processing Systems (NeurIPS 2020)*, December 2020
5. Uniform Convergence of Low-Norm Interpolators in Overparametrized Linear Regression, *Student seminar (UChicago Statistics)*, May 2020

Teaching

Serve as the teaching assistant for

- *Convex Optimization* (Winter 2020, Winter 2022)
- *Statistical Theory and Methods* (Autumn 2020, Autumn 2019, Winter 2019 & Autumn 2021)
- *Introduction to Random Matrices* (Winter 2021)
- *Optimization* (Spring 2019)

and the course reader/grader for

- Mathematical Methods for Social Sciences (Winter 2017)
- Linear Algebra (Autumn 2016)

Programming Language

- Python (PyTorch, scikit-learn, CVXOPT), R, SQL
- MATLAB, Mathematica, LaTeX