Liyi Zhang

2900 Broadway, New York, NY 10027 | zhang.liyi@columbia.edu | (929)732-8260

EDUCATION

Columbia University: Columbia Engineering New York, NY
M.S. in Data Science Sep 2021 – (expected) May 2022

Columbia University: Columbia College New York, NY
B.A. in Statistics and Applied Mathematics, GPA: 3.9/4.0 Sep 2017 – May 2021
Math coursework: Modern Analysis I, Modern Analysis II, Linear Algebra, Ordinary Differential Equations, Analysis and Optimization, Seminar in Applied Mathematics
Statistics coursework: Probability Theory, Multivariate Statistical Inference, Bayesian Statistics, Stochastic Processes, Linear Regression Models, Statistical Computing in R, Graphical Models
Computer Science coursework: Data Structures, Analysis of Algorithms, Natural Language Processing, Optimization for Machine Learning, Mathematics of Deep Learning

RESEARCH EXPERIENCE

Columbia Dept. of Computer Science – Professor David Blei *Researcher | May 2021 – present*
- Create a novel algorithm to meld variational inference with Markov chain Monte Carlo
- Code the algorithm and benchmark methods in Python and TensorFlow, and apply 2 families of distributions to 3 datasets, including MNIST image data with variational auto-encoders
- Write a full academic draft, recently submitted to a machine learning conference

Columbia Dept. of Computer Science – Professor Itsik Pe'er *Researcher | Jan 2020 – May 2021*
- Created a novel algorithm Variational Combinatorial Sequential Monte Carlo (VCSMC) for phylogenetic inference using Python and TensorFlow and 9 real DNA datasets
- Collaborated side-by-side with a PhD and postdoc of the lab and presented results to the lab twice
- Led the implementation process by proposing and executing new schemes for coding
- Wrote a full academic paper, accepted at Uncertainty in Artificial Intelligence 2021

Columbia Dept. of Statistics – Professor Andrew Gelman *Researcher | May 2020 – Sep 2020*
- Implemented stacking algorithm and Bayesian model averaging (BMA) for discrete parameters in phylogenetic inference from scratch in R
- Conducted experiments on synthetic data and real primate DNA data and demonstrated that stacking produces less spuriously high posteriors than BMA

IN-CLASS PROJECT EXPERIENCE

Foundations of Graphical Models – Professor David Blei *Final Project | Dec 2020*
- Implemented Bayesian Neural Networks (BNN) using TensorFlow with customized layer and training
- Wrote and executed adversarial perturbations of the MNIST image data
- Compared BNN with 16 deep learning models trained across a range of regularization levels, and demonstrated that BNN is more robust with an average margin of 15.4% of classification accuracy

PROFESSIONAL EXPERIENCE

QTG Capital Management, Shanghai, China *Quantitative Researcher Intern | Jun – Aug 2019*
- Cooperated with a full-time trader to analyze and clean financial and commodities futures datasets
- Conducted clustering analyses using the K-Means algorithm for commodities futures
- Built factor models and performed feature engineering on financial and commodities futures to develop and back-test trading strategies

SKILLS AND INTERESTS

Skills: Python (+TensorFlow), R, Stan, LaTeX, Machine Learning (certified via ColumbiaX on edX), Deep Learning (certified via Coursera)
Language: Mandarin Chinese: fluent; Latin: intermediate
Interests: piano, clarinet, chess, distance running, hiking