Sida (Star) Li

(510)-847-2494 ⋄ listar2000@uchicago.edu ⋄ https://listar2000.github.io/

EDUCATION

The University of Chicago Starting in Sept 2024
Ph.D., Data Science Institute

The University of Chicago Sept 2022 - June 2024
Master of Science, Statistics GPA: 3.97
Thesis Mentor: Daniel Sanz-Alonso

University of California, Berkeley August 2018 - May 2022
Bachelor of Arts, Statistics & Computer Science GPA: 3.95
Statistics Department Citation (Valedictorian) Winner

RESEARCH INTERESTS

Approximate Bayesian Inference, Probabilistic ML, AI4Science, Empirical Bayes

RESEARCH EXPERIENCE

The University of Chicago Statistics Department, UChicago, IL
Mentor: Nikolaos Ignatiadis *April 2024 - Present*
Researching into empirical Bayes mean estimation problems under semi-supervised assumptions. Working on extending the prediction-powered inference (PPI) framework to a compound decision setting and building a new empirical Bayes estimator that (1) utilizes massive unlablled data with PPI de-biasing and (2) enjoys risk guarantees comparable to the full-Bayesian oracle estimator.

UChicago Master Thesis, UChicago, IL
Mentor: Daniel Sanz-Alonso *June 2023 - April 2024*
Worked on accelerating and generalizing Langevin Monte Carlo (LMC) methods for sampling. Experimented and verified how adding a curl matrix into the Langevin SDE accelerates convergence in various statistical models. Implemented and benchmarked a new Fisher-information based LMC method that outperformed traditional counterparts in various metrics.

Autonomous Empirical Research Group, Brown University, RI
Mentor: Sebastian Musslick *March 2022 - Present*
Researching into symbolic regression (SR) - the ML problem that searches the best-fitting expression for a given dataset. Developed a hierarchical Bayesian framework for the SR problem and corresponding inference algorithm to sample from the posterior. Pioneered the design of a new SR method based on Generative Flow Networks (GFlowNets) and deep learning that achieves SOTA performance in noisy settings.

FHL Vive Center for Enhanced Realtiy, UC Berkeley, CA
Mentor: Allen Yang *March 2020 - May 2021*
Developed ROAR, an autonomous racing simulator, and implemented a set of perception, planning, and control algorithms. Applied model-based deep reinforcement learning algorithms to vehicle controllers for autonomous racing.

Sandrine Dudoit Lab, UC Berkeley, CA
Mentor: Hector Roux de Bezieux and Koen Van den Berge *January - May 2020*
Participated through the Undergraduate Research Apprentice (URAP) program. Investigated how initialization affects unsupervised dimensionality reduction methods such as UMAP and t-SNE for scRNAseq data, with an emphasis on the preservation of global structures in low dimensional space.

WORK EXPERIENCE

Software Engineer Intern, Duolingo, Pittsburgh, PA *May-August 2021*
Implemented internal tools in the ETL data pipeline that support efficient querying and computation on key metrics (e.g. daily bookings, active users); revised the A/B testing framework by enabling auto-correction in confidence intervals for ad-hoc metrics.

Data Consulting Intern, Concha Inc., Berkeley, CA *January-May 2020*
Worked on predicting customer's hearing loss curve based on response data from online testings. Applied and evaluated existing machine learning methods such as regression tree and RNN for the prediction tasks.

PAPERS & REPORTS

Sida Li, Ioana Marinescu, Sebastian Musslick. "GFN-SR: Symbolic Regression with Generative Flow Networks." **NeurIPS 2023 AI4Science Workshop**. [Link] [Poster]

Sebastian Musslick, Joshua Hewson, Ben Andrew, **Sida Li**, George Dang, John Gerrard Holland. "Evaluating Computational Discovery in the Behavioral and Brain Sciences." **AAAI 2023 Spring Symposium Series, Computational Approaches to Scientific Discovery**. [Talk Abstract]

Sida Li, Joshua Hewson, Sebastian Musslick. "Hierarchical Bayesian Symbolic Regression." **Work in Progress, 2023**. [Link]

Michael Estrada, **Sida Li**, Xiangyu Cai. "Feedback Linearization of Car Dynamics for Racing via Reinforcement Learning." **Preprint, 2021**. [Link]

THESIS

Beyond Vanilla Metropolis-Adjusted Langevin Dynamics. *Mentored by Prof. Daniel Sanz-Alonso.* [Link]

SOFTWARES

Automated Research Assistant (AutoRA) [Link]
An open-source framework for automating multiple stages of the empirical research process, including model discovery, experimental design, data collection, and documentation for open science.

ROAR Simulator [Link]
An open-source platform/API for autonomous driving simulations based on CARLA. Include pre-built algorithms in perception (computer vision), control, planning and visualizations.

AWARDS

UChicago M.S. Stat Scholarship (25% tuition remission)	FA22, FA23
UC Berkeley Statistics Department Citation	FA22
UC Berkeley Dean's Honors List (top 10% GPA)	SP19, FA19, SP20, SP21
Upsilon Pi Epsilon (top one third of CS majors)	FA19, SP20, FA20, SP21

SKILLS

Languages: English, Mandarin, Cantonese
Programming: Python, R, C++, Java, Javascript, Ruby, LaTeX
Frameworks: PyTorch, TensorFlow, NumPy, Scikit-learn

TEACHING

CS 198-097 Robot Autonomous Racing DeCal (Head Instructor)	Fall 2021
CS 198-097 Robot Autonomous Racing DeCal (Instructor)	Fall 2020
STAT 134 Probability Theory (Tutor)	Spring 2020
STAT 134 Probability Theory (Tutor)	Fall 2019
MATH 32 Precalculus (Tutor)	Summer 2019