Yuanxin Yao

Tallahassee, FL *yy18b@my.fsu.edu* *yaoyuanxin2017@gmail.com*

EDUCATION

Florida State University **Tallahassee, FL,** August 2018-Present
PhD in Statistics, 3.9/4.0

University of California, San Diego **La Jolla, CA,** *August 2014-Dec 2017*
Bachelor of Science in Probability and Statistics, 3.78/4.0

SKILLS

- **Programming Skills**: R, Python (Pandas, NumPy, MlPy, MatPlotlib, SciPy, TensorFlow, Scikit-learn, Apache Spark, Pytorch), SQL, SAS, Matlab, C++, Java
- **Quantitative Skills:**

  Machine Learning: Reinforcement Learning, Deep Learning, Convolutional Neural Network (CNN)

  Classic Statistics: Generalized Linear Model, Logistic Regression, Bayesian Analysis, Survival Analysis, Gibbs Sampling

  Mathematics: Real Analysis, Advanced Calculus, Black-Scholes Model

SELECT PROJECTS AND PUBLICATIONS

Iterative Coreset Subsampling Model (Dissertation Chapter, Submission in Progress)
R, Python *November 2020–Present*

- Developed an empirical Bayesian model that allows to work with high dimensional data in informative subsample (coresets)
- Proved the theoretical bounding guarantees in the iterative Hilbert coreset constructions
- Implemented the said model in a Gibbs Sampler with Markov Chain Monte Carlo on simulated probit regression and linear regression models
- Achieved a ~30% decrease in CPU run time and smaller mean square prediction error compared to existing Generalized Linear Model (GLM) methods

Single Value Decomposition on High Dimensional Matrices **Python** *October 2021–October 2021*

- Organized a Python library that allows one to solve the single value deposition (SVD) of high dimensional (~8TB, 10^6 by 10^6) matrices efficiently, using Random Projection, Principal Component Analysis and Randomized SVD
- Simulated both dense and sparce large matrices to verify the computational efficiency of the aforementioned methods

Bayesian Frailty Model for Repeated Survival Analysis **SAS, WinBugs** *September 2020–November 2020*

- Implemented cluster survival analysis with both traditional Cox model and the Bayesian Frailty Model
- Performed Bayesian analysis and Weibull Regression on DRS treatment data
- Compared the frailty model with an empirical MCMC posterior and verified the accuracy of the Bayesian frailty model

Generalized Linear Model on Housing Data **Python, R** *September 2017–November 2017*

- Utilized Ridge Regression on the Ames housing price data (2930observations and 79 parameters)
- Performed data cleaning and feature selection of the data using XGBoost
- Compared the model result with testing data and obtained a 91.6% prediction accuracy

PROFESSIONAL EXPERIENCE

Graduate Researcher — **Tallahassee, FL**
Florida State University — *August 2018–Present*

- Conduct simulations and data analysis, review and write papers towards dissertation
- Collaborate with advisors, PhD students and post-doc students across the department statistics, scientific computing and physics

Solo Instructor — **Tallahassee, FL**
Florida State University — *August 2018–Present*

- Hold 50-minute lectures three times a week for Elementary Applied Statistics.
- Coordinate lectures and write walk-through tutorials in R Markdown and SAS to simulate concepts such as Central Limit Theorem and Contingency Table

Simultaneous Interpreter — **Tallahassee, FL**
Shilei Interpretating and Translation — *Nov 2019–Mar 2020*

- Translated in court hearing settings, between English and Mandarin
- Facilitated communications between Mandarin-speaking clients and their English-speaking attorney

Undergraduate Researcher — **La Jolla, CA**
University of California, San Diego — *May 2016–Dec 2017*

- Read and applied Isotonic Regression in experiments using R
- Applied non-machine-learning techniques, specifically, Heat flow and Spline algorithms in **C++** to achieve a precision rate of 89.3% on the recognition of 2000 noisy images

Undergraduate Teaching Assistant — **La Jolla, CA**
University of California, San Diego — *Oct 2016–Dec 2017*

- Mentored students through office hours and one-on-one communication.
- Held weekly discussions for Math 10B, Math 11, Math20B, course works in Calculus and Calculus-Based Statistics
- Facilitated in proctoring exams and grading exams.