Jinjin Tian

Email: jinjint@andrew.cmu.edu | Phone: (412)-616-6460 | Github: JINJINT

Education Background

PhD in Statistics and Data Science, Carnegie Mellon University — ***GPA 4.08, 2018-(exp) May 2023***

Bachelor in Statistics, University of Science and Technology of China — ***GPA 3.86, July 2018***

Qualification

- **5+ Years research experience in statistical modeling and methodology**
- **5+ Years of programming experience, familiar with Python (Tensorflow, PyTorch), R, MATLAB, know of C, SQL**

Work Experience

- **Online experimentation in E-commerce with revenue and time constraints** (package) *June– Aug. 2020.*
  Applied Scientist Intern, AWS, Amazon. Supervisor: Lenon Minorics, Professor Guido Imbens.
  - Constructed a new online experimentation algorithm that allows an efficient trade-off between revenue constrain and time constraints in E-commerce utilizing recent advances in reinforcement learning and any-time valid inference.
  - Proposed new algorithm to deal with unknown abrupt changes over time using Thompson Sampling and sequential change detector, which only cost O(1) for each update, and has much lower regret comparing with other state-of-arts.

Research Experience

My research interests lie in developing methodology and theory for high dimensional, nonparametric data analysis with special focus on **feature interaction** and **unlabeled data**.

- **Detecting non-linear non-monotone relationship by sample specific dependency** *Feb. 2020 –Present*
  Under supervision of Professor Jing Lei & Kathryn Roeder.
  - Developed a density based dependency measure, together with a consistent, tuning-free, nonparametric estimator. Proved its ability to detect non-linear, non-monotone relationship and robustness to outliners. When applied as an independence test, it is competitive among state-of-arts like taustar, distance correlation, and HSIC.
- **Signals recovery in noisy high-dim mixture via local structure learning** *Sept. 2020 –Present*
  Under supervision of Professor Jing Lei & Kathryn Roeder.
  - Proposed a statistics to capture local structure between a pair of features. Proved exact recovery of signals based on this statistics in a high-dim mixture model, under even impossible scenarios for canonical methods like sparse PCA.
- **Self-supervised learning with and without contrastive pairs** *Feb. 2021 –Present*
  Under supervision of Professor Andrej Risteski.
  - Investigated the role of augmentation, adversarial samples and network structure in self-supervised learning (SSL) of image with and without contrastive pairs, through downstream classification task, feature visualization and texture-shape bias analysis. Working on a sparse coding model to theoretically compare the feature learning process of SSL with and without contrastive pairs.
- **Single-cell gene expression data modeling incorporating gene interaction** (Bioinfomatics 2021, package)
  Supervisors: Professor Kathryn Roeder & Jiebiao Wang. *July. 2019 –May. 2020*
  - Modeled the single-cell gene expression data to not only characterize various types of heterogeneity through a hierarchical model but also depict higher-order gene interaction using a copula model. Wrapped up the model as a simulation software which also allows mimicking real data via fitting the model using a moment matching method.
- **Large-scale post-hoc model selection and inference under dependence**(arXiv) *Sept. 2019 – May. 2020*
  Supervisors: Professor Aaditya Ramdas & Eugene Katsevich & Jelle Geoman.
  - Developed a class of simultaneous inference and model selection methods under arbitrary dependence with only linear time computation. Derived tighter calibration and the corresponding asymptotic power under a Gaussian dependence model.
- **Adaptive algorithms for online error control** *Jan – Sept. 2019*
  Supervisor: Professor Aaditya Ramdas.
  - Constructed new algorithm for online false discovery rate control via adapting to both signal proportion and noise level, which outperforms current state of arts in terms of applicable range as well as power. (NeurIPS 2019, code)
  - Extended the adaptive idea for familywise error rate control. Formally proved substantial gains of power for the new methods, and derived closed-form optima of hyperparameters in a Gaussian sequence model. (SMMR 2021, code)

Coursework

- **Machine Learning:** Introduction to Advanced ML (A+), Statistical ML (A+), Convex Optimization, Deep Reinforcement Learning, Adavanced Deep Learning (A+)
- **Statistics:** Measure Theory, Intermediate Statistics, Advanced Probability Theory, Reproduciblility, Regression Analysis, Advanced Data Analysis, Stochastic Process, Time Series
- **Math:** Functional Analysis, Real Analysis, Complex Analysis, Mathematical Analysis, Linear Algebra
- **Programming:** Statistical Computing, Data structure and Data base, C and Algorithm.