YI GU

Evanston, IL
(713)-380-7338 ◇ Yi.Gu@u.northwestern.edu

EDUCATION

Northwestern University
Ph.D in Mathematics
Thesis Title: Generalized t-SNE and Beyond: Probabilistic Methods for Dimensionality Reduction, Combinatorial Optimization, and Machine Learning

Rice University
B.S. in Mathematics
Major GPA: 4.1/4.0, President Honor Roll

PUBLICATIONS

(Preprint) **Yi Gu**. *Equilibrium Distribution for t-Distributed Stochastic Neighbor Embedding with Generalized Kernels.* arXiv:2505.24311

Yi Gu, Nan Jiang *et al*. *Learning Combinatorial Structures via Markov Random Fields with Sampling through Lovász Local Lemma.* **AAAI 2023**

Zichao Wang, **Yi Gu** *et al*. *Iterative Imputating Variational Auto-Encoders for Partially Observed Data.* **EDM 2020**

(Book) **Yi Gu**. *Introduction to Random Graphs and Models.* Research Survey in Random Graph, 2019. (PDF link available)

EXPERIENCE

Google Sept 2024 - Present
Data Scientist - Research *Chicago, IL*

- Independently designed and implemented a novel, end-to-end LLM sentiment analysis pipeline from scratch, integrating LLM-as-judge, SFT, and RAG to classify hardware data; achieved 99.5% accuracy, directly informing critical, company-wide hardware KPI targets and RMA analysis.
- Developed a high-efficiency anomaly detection system to monitor thousands of metrics across trillion-row Pixel/Android data warehouses, using bespoke dimensionality reduction and clustering algorithms to successfully identify critical data integrity issues.
- Hired and mentored an intern during first year of employment, defining the project scope and providing direct technical guidance, leading to a successful project outcome.

Google June 2023 - Sept 2023
Data Scientist Intern (Ph.D) *Mountain View, CA*

- Designed a novel latency-tracking algorithm with millisecond-level accuracy and established a new OKR, driving a 50% latency reduction for over 90% of Pixel-related datasets and enhancing data precision for all dependent teams.
- Independently built an E2E failure prediction pipeline with 40+ high-impact features (90%+ recall, less than 10% false positive), reducing system failure rates by over 50% and saving maintenance costs.

Google June 2022 - Sept 2022
Data Scientist Intern (Ph.D) *Mountain View, CA*

- Led a time-sensitive search quality study, analyzing user surveys and query data to generate insights that resulted in the launch of a new Knowledge Panel feature.
- Researched and implemented homomorphic encryption algorithms in collaboration with Google Ads, enhancing user privacy protection for personalized advertising and ensuring compliance with new regulations.

Invesco June 2021 - August 2021
Quantitative Research Intern *Remote, IL*

- Developed quantitative models using stochastic analysis to predict fund behavior, building an automated data infrastructure to aggregate data, analyze performance, and manage strategies daily.

Northwestern University 2017-2023
Course Instructor *Evanston, IL*

- Instructed a graduate-level course on Random Matrix Theory (30+ students), a responsibility typically reserved for faculty.
- Served as Teaching Assistant for 10+ foundational courses in Mathematics and Statistics, including Real Analysis and Probability Theory at graduate level.

TECHNICAL SKILLS

Programming	Python, SQL, MATLAB, Mathematica
ML/DL Frameworks	PyTorch, TensorFlow, JAX, Scikit-learn, Pandas, NumPy
Platforms/Tools	Google Cloud (BigQuery, Vertex AI), Apache Beam (Google Flume), Kubernetes (Google Borg), Docker
Expertise	Large Language Models (LLMs), Anomaly Detection, Dimensionality Reduction and Data Visualization, Statistical Modeling, Time Series Analysis, Stochastic Analysis, Pattern Recognition, Combinatorial Optimization, Supervised Fine-Tuning (SFT), Retrieval-Augmented Generation (RAG)