Alex Chen

Los Angeles, CA | aclheexn1346@g.ucla.edu | (814) 380-8601

https://www.linkedin.com/in/achen-ucla | aclheexn1346.github.io

Education

University of California Los Angeles, PhD in Statistics — Sept 2020 – Dec 2025
- GPA: 4.0/4.0

Pennsylvania State University, BS in Statistics, BS in Data Science, Minor in Math — Aug 2016 - May 2020
- GPA: 3.83/4.0

Experience

Graduate Student Researcher, UCLA – Los Angeles, CA — Aug 2021 - Present
- Led NSF-funded research to develop a novel algorithm for causal discovery in datasets with strong data dependencies, achieving a 20–40% improvement in accuracy compared to standard methods.
- Applied unsupervised clustering algorithms and custom data preprocessing pipelines to analyze single-cell RNA-seq datasets, enhancing causal graph (GRN) estimation accuracy on real-world data.
- Implemented state-of-the-art and novel causal inference techniques in Python and R, benchmarking performance to showcase significant improvements over existing approaches.
- Optimized causal discovery pipeline by leveraging high-performance cluster computing, running distributed computations (MapReduce-inspired framework) to process datasets in parallel and aggregate results for large-scale causal analysis reducing computation time by 50%.

Machine Learning Research Intern, Bristol-Myers Squibb – Seattle, WA — June 2024 - Aug 2024
- Designed a novel methodology to integrate information across massive raw single-cell datasets and cell types for estimating Gene Regulatory Networks (GRNs) using causal discovery techniques.
- Developed an R package to streamline and containerize the GRN estimation process, enhancing scalability, reproducibility, and usability across diverse biological datasets.
- Identified potential gene regulators through in-silico gene perturbation simulations based on the estimated GRN, collaborating with domain experts to refine models and improve gene identification accuracy.

Data Scientist Intern, Terex – Redmond, WA — June 2021 - Aug 2021
- Developed end-to-end predictive maintenance system to estimate Remaining Useful Life (RUL) for 1,000+ industrial machines, using raw sensor data integrating transformer models for sequential data and neural networks for static features using Python. Increased interaction with customer viewable dashboard by 5%
- Implemented a collaborative filtering recommendation system on sales data, profiling similar customers together for more targeted sales strategies.
- Contributed to development of scalable ETL pipelines using Apache Airflow, enabling real-time monitoring and feature engineering key machine usage metrics on AWS Quicksight dashboard.
- Migrated SQL code across database platforms, ensuring data integrity and optimizing system interoperability.

Data Scientist Intern, Lockheed Martin – Denver, Colorado — May 2019 - July 2019
- Designed and implemented a instrumental variable two-stage regression model to estimate the causal effect of pressure on tensile strength in manufactured parts from autoclaves. Included systematic sensitivity analysis by iteratively removing confounders to assess the robustness of the causal estimates.
- Worked closely with engineering teams to identify relevant confounders affecting both pressure and tensile strength. Used R to clean and align sensor information with manufactured parts data, ensuring valid IV analysis and comparing the causal effects with traditional regression.

Technologies

Languages: R, Python, PostgreSQL, Shell Scripting

ML Libraries: Pandas, Numpy, Pytorch, Scikit-learn, Matplotlib, RShiny

Technologies: GitHub, AWS, MongoDB, MapReduce, Amazon Redshift, Docker, Tableau, Spark