Alex Chen

(814)380-8601 | aclheexn1346@g.ucla.edu | aclheexn1346/aclheexn1346.github.io | U.S. Citizen

Education

University of California Los Angeles (UCLA)
PhD Candidate in Statistics
GPA: 4.0/4.0

Los Angeles, California
Expected Graduation Date: December 2025

Pennsylvania State University
Double Major in Data Science and Statistics
GPA: 3.83/4.0

University Park, Pennsylvania
Graduation Date: May 2020

Skills

Programming	R, Python, SQL, SAS, Docker High-Performance Cluster parallel computing, RShiny, AWS, C, C++
Notable Graduate Classes Taken	Machine Learning, Graphical Models, Monte-Carlo Optimization, Causal Inference

Experience

Machine Learning and Computational Biology Internship
Informatics and Predictive Sciences (IPS) Internship at Bristol Myers Squibb
Seattle, Washington
July 2024 - Present

- **Created and applied innovative computational methods:** Designed causal graph estimation method to ingest multiple different data sources into one consensus graph
- **Developed computational pipeline for constructing gene regulatory networks:** Used R and Github to develop R package to clean datasets and generate consensus causal graphs. Optimized methods for computational efficiency through parallelization, minimizing function calls, and implementing faster data reading functions.
- **Collaborated with IPS colleagues to identify key drivers that improve efficiency of cell therapy drugs:** Analyzed causal graphs and introduced new potential drivers of cell exhaustion. Led discussions on causal graph methods and findings, giving presentations in both team and division-wide meetings.
- **Continuing work to introduce context-dependent information into methods:** Currently collaborating, post-internship, with IPS team to incorporate other data sources and analyses to improve causal graph estimation

PhD Research
UCLA Causality Gaussian DAGs from Network Data with Dr. Qing Zhou
Los Angeles, California
June 2021 - Present

- **Causal Inference on Network Data:** NSF funded research on causal inference estimation of graph networks to do effective analysis of causal relationships between discrete variables with an EM-algorithm resulting in 20-50% improvement over current methods using High-performance cluster for parallel computing
- **RNA-seq Clustering and Causal Network Detection:** Implemented unsupervised clustering algorithms on single-cell RNA-seq data to group cells based on gene expression that resulted in significant improvement to finding GRN structure

Data Science Internship
Database and Machine Learning Internship at Genie Aerial Work Platforms
Redmond, Washington
June 2021 - Sept. 2021

- **Outlier detection and Modeling User Usage:** Utilized SQL and Python to continuously pull and merge sensor IoT data from AWS Redshift databases for modeling fault detection and analysis of user usage time on live dashboards
- **Recommendation System for client purchases:** Supported Sales team using collaborative filtering recommendation system through R for purchases from major clients

Data Science Internship
Machine Learning and IoT Internship at Lockheed Martin
Denver, Colorado
May 2019 - Aug. 2019

- **Machine learning methods for fault detection:** Used outlier estimation methods and self-taught causal graph learning methods to determine sources for faults in large machinery. Created and implemented custom time variable to assist in modeling predictive maintenance with 85% average cross-validation accuracy using R and Python.

Projects

2020	**Show Recommendation System**, Web-scraped korean drama data and used nearest-neighbor matching algorithm with BERT LLM to understand show synopses.
2024	**Segmenting Yelp reviewers using LLM Open AI with Neo4j for knowledge graph extraction**, Understanding Yelp reviews to determine trustworthiness of reviewers. Using MongoDB to store review data and Open AI with Neo4j to create knowledge graphs for context-dependent relationships