Alex Chen

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

Education

University of California Los Angeles (UCLA)

Los Angeles, California
Expected Graduation Date: June 2025

PhD Candidate in Statistics

GPA: 4.0/4.0

Pennsylvania State University

University Park, Pennsylvania
Graduation Date: May 2020

Double Major in Data Science and Statistics
Double Minor in Math and Korean
GPA: 3.83/4.0

Experience

Data Science Internship
Redmond, Washington
Database and Machine Learning Internship at Genie Aerial Work Platforms
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 new metrics fault detection and analysis of user usage time on live dashboards
- **Recommendation System for client purchases:** Supported Sales team using recommendation system through R for repurchases from major clients
- **SQL Code Transfer:** Assisted in converting frequented information requests to new SQL queries for database repositories during large transition period. Required translating old SQL code to run properly on new databases.

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

- **Fault detection on Machine IoT:** Self-taught Causal Inference through R to model fault sources after pre-processing 1000s of variables using variable selection methods
- **Early detection for predictive maintenance:** Performed fault detection for predictive maintenance on large machinery with 85% average cross-validation accuracy using R and Python through using gradient boosted trees after multiple model testing such as deep learning, random forest, and Adaboost

Research Experiences

PhD Research
Los Angeles, California
UCLA Causality Gaussian DAGs from Network Data with Dr. Qing Zhou
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 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 real RNA-seq data to group cells based on gene signals that resulted in significant improvement

Undergraduate Statistics Researcher
University Park, Pennsylvania
Penn State Eberly College of Science with Dr. Lingzhou Xue
Jun. 2018 - May 2020

- **Chromosome Interaction Prediction Lasso Model:** Developed a constrained penalized Lasso model using a stratified minibatch sub-sampling procedure to fit the model approximately 100 times faster than traditional methods to estimate interaction intensity between chromosomes as a substitute for costly Hi-C analysis: a model for 25K by 25K intensity matrix collapsed to fit into databases
- **Amazon Review Text:** Used variational EM model to segment consumers into different groups based on reviewers heterogeneity to review different products using metrics like user fairness and product goodness as a means for a reviewer recommendation system. Used Hadoop to store and access Amazon Product-Review text and rating data

Awards

2018 **PSU DataFest,** Won 1st Overall and awarded Best Insight in annual DataFest Competition using Indeed data to make insights and recommendations to Indeed potential additional analyses to increase user usage.

Skills

Programming	R, Python, SQL, SAS, Hadoop, High-Performance Cluster parallel computing
Statistical Techniques	Causal Inference, Big data modeling, Network Modeling
Notable Graduate Classes Taken	Machine Learning, Graphical Models, Monte-Carlo Optimization