Sam Rosen

SamRosen90@gmail.com ❖ (443) 758-0586 ❖ Durham, NC ❖ SamGRosen.github.io

WORK EXPERIENCE

Cruise **Summer 2022**
Machine Learning Acceleration Intern San Francisco, CA
- Implemented robust and nonparametric standard error calculations to internal simulation metrics.
- Added Kernel Density Estimation/Bandwidth Selection to automatically analyze performance degradation.

Genentech **Summer 2021**
Research and Development Intern San Francisco, CA
- Lead development of *epiviz.gl*, a JS framework for visualizing genomic data with WebWorkers and WebGL.
- Developed data selection, rendering, navigation with a pseudo grammar-of-graphics implementation.

Datadog **Sept. 2020 – Dec. 2020**
Software Engineering Intern (Cloud Integrations Team) New York, NY
- Optimized and bolstered Azure crawlers responsible for crawling millions of data points an hour.
- Debugged and implemented fixes for issues found by customers in production for crawled metrics.

DraftKings **Summer 2019**
Software Engineering Intern (DevOps Team) Boston, MA
- Created a scalable application for live tracking of release branches to production using AWS Lambda.
- Designed serverless architecture scalable to arbitrary codebase size with complete up-to-date release data.
- Designed DynamoDB schema and frontend with React for a responsive, efficient API and user interface.

Johns Hopkins University: Applied Physics Lab **Summer 2017, 2018**
Software Engineering Intern (Large-Scale Analytics Group) Laurel, MD
- Programmed low-memory implementations of machine learning algorithms for training on arbitrarily large data.
- Created analytics for graph multi-edge merging, time-series, and data fusion using Java and MapReduce.
- Developed a random forest algorithm on a distributed data system for classifying attributes on graph vertices.

SELECTED PROJECTS

OKRidge: Scalable Optimal k-Sparse Ridge Regression | arxiv.org/abs/2304.06686 **April 2023**
- Second author on Lui et al. for certifiably optimal sparse ridge regression via branch-and-bound.

JS Package: *epiviz.gl* | github.com/epiviz/epiviz.gl **May 2021 – Aug. 2021**
- Developed for Genentech to visualize genomic data seamlessly via declarative specifications and WebGL.
- Designed to visualize millions of data points and entire chromosomes at 60 FPS with high precision.

Python Package: *Diary* | github.com/SamGRosen/diary **Nov. 2016 – Present**
- Created a no-dependency package to make asynchronous logging easy with a highly customizable API.

EDUCATION

Duke University **Aug. 2021 – Present**
PhD Student, Statistics Durham, NC
- Bass Connections Fellowship (Fall 2022)

University of Massachusetts: Amherst **May 2021**
BS, Computer Science; BS, Mathematics (3.9/4.0 GPA) Amherst, MA
- Recipient of UMass Chancellor's Award – Four year academic scholarship

SKILLS & INTERESTS

- **Programming Languages:** Python***, Java***, JavaScript***, R***, C/C++**, Scala*, Matlab*
- **Research Interests:** Statistical Computing, Stochastic Optimization, Bayesian Statistics, High-Performance Computing, Simulation Optimization
- **Teaching Experience:** Calculus I/II, Differential Equations, Graduate Statistical Computing, Data Visualization, Intro to Health Data Science, Master's Programming Orientation Instructor