Alice Wu Lim

PhD, Data Scientist

Los Angeles, CA 90048
Remote, Any Time Zone
862.485.3465
limwualice@gmail.com
http://www.alicewulim.com

Education

Syracuse University/ PhD in Mathematics
August 2015 - May 2021

- Thesis in Differential Geometry
- Received Kibbey Prize for excellence in PhD research.

Syracuse University / MS in Mathematics
August 2015 - May 2018

- Relevant courses include: Probability and Statistics (4 graduate courses), Enumeration, Design, and Matroids, Graph Theory, Functions in Complex Variables, Calculus on Manifolds

University of California, Los Angeles / BS in Mathematics
August 2011 - May 2015

- Relevant courses include: Fourier Analysis, Introduction to C++

Experience

Hack for LA / Data Scientist (Volunteer)
January 2023 - Current

- Designing and implementing a data cleaning pipeline to transform raw input data from the Socrata API related to LA city parking and update the Google Cloud Platform database with geospatial data
- Developing a list of data cleaning steps to ensure the data meets the client's requirements
- Creating code to extract, transform, and load data into the database
- Conducting multiple tests to ensure data is correctly formatted and cleaned
- Deploying the pipeline to GCP for real-time data processing and updates
- Managing the end-to-end project, from initial requirements gathering to final deployment, ensuring timelines and deliverables are met
- Visualized data using Looker Studio
- Collaborated with team members to troubleshoot and resolve any issues that arose during the project

City of Hope / Data Scientist (Research fellowship)
April 2022 - October 2022

- Used an unsupervised learning machine learning predictive model in Python and SQL, statistical model, and advanced data analysis to discover statistically significant correlation between breast cancer genomic clusters and tnm-staging as well as breast cancer biomarker sets.
- Learned about breast cancer and machine learning techniques, while adapting projects to doctors' needs. Completed technical portion of data science project, manuscript is now in preparation.

How this was accomplished:

- Goal: use unsupervised learning model to cluster large datasets of patient genomic data, find statistical correlation between engineered patient features and the detected clusters
- Collected big data: Queried terabytes of data from over 30 satellite hospital databases; transformed and combined data into cleansed dataframes with hundreds of features
- Clustered data: Constructed k-modes clustering model and detected genomic clusters of patient data in order to analyze genomic features
- Engineered features: Used domain knowledge to engineer dozens of new features from patient data to test for correlation with genomic clusters
- Performed statistical analyses using analytical tools,: Performed non-parametric significance testing which revealed statistically significant correlations between several patient features and the genomic clusters
- Communicated findings: Presented the findings to business partners, collaborators, and non-experts

Syracuse University / PhD Researcher
August 2015 - May 2021

- Wrote 2 single-author differential geometry research papers which were accepted for publication in prestigious journals.
- Used Python to compute explicit solutions to N-quasi Einstein equation under certain constraints.
- Gave 12 presentations at conferences and invited talks where I communicated results to various sized audiences.
- Received Kibbey award for demonstrating excellence in PhD research.
- Studied graduate level statistics and probability courses
- Studied Calculus of variations and Euclidean and non-Euclidean geometry, which have applications in mathematical optimization and satellite access geometry, respectively

Research Papers

1. Clustering whole-exome sequences of breast cancer reveals association with staging and molecular subtype, in preparation
2. The Splitting Theorem and Topology of Noncompact Spaces with Nonnegative N-Bakry Emery Ricci Curvature; Proceedings of the AMS, May 2021; https://doi.org/10.1090/proc/15240

3. Locally Homogeneous Non-gradient Quasi Einstein 3-Manifolds; Advances in Geometry, January 2022; https://doi.org/10.1515/advgeom-2021-0036

Presentations

- Gave 3 talks at City of Hope about my data science/breast cancer research.
- Gave 12 invited math talks at: American Mathematical Society, University of Oregon, 2020 Virtual Workshop on Ricci and Scalar curvature, Dartmouth College, Union College, Lehigh University, Institute for Advanced Studies Joint Mathematics Meeting, Syracuse University (3 talks). Presentations were to various sized audiences from small departments to large conferences.
- My goal with giving presentations is to incorporate as many visual aids as I can to get my point across. I make sure to start by stating the goal of my project and spend the rest of the talk exploring how I met my goal.
- See here for a link to some of my presentations: https://www.alicewulim.com/researchtalks.html