Yimeng Shang

☎ +1 (646) 704-5390 ✉ ys3298@cumc.columbia.edu https://ys3298.github.io/

EDUCATION

Columbia University, Mailman School of Public Health — New York City, NY, USA
- Master of Science in Biostatistics (GPA: 4.0/ 4.0) — 08/2019-Expected in 05/2021

East China Normal University — Shanghai, China
- Bachelor of Science in Mathematics and Applied Mathematics — 09/2015-06/2019

University of California, Berkeley — Berkeley, CA, USA
- Berkeley International Study Program — 08/2017-05/2018

RESEARCH/ PROJECTS EXPERIENCE

Variable selection and Prediction using EM algorithm base on Multiple label-noise response — 05/2020-09/2020

- Research assistant, supervised by Professor Hua Shen — *University of Calgary*

- Combined EM algorithm and Adaptive LASSO to estimate sensitivity and specificity of each surrogate, did feature selection and prediction based on this algorithm.
- Constructed multiple simulation studies under different settings, including low dimensional data and high dimensional data under different assumptions, different penalty terms and different tuning parameter selection criteria
- Achieved improvement in parameter estimation, feature selection and prediction compared to the Naïve method and Ad Hoc method
- A manuscript is prepared

Statistical analysis on Autism Metabolomics data — 05/2020-08/2020

- Research assistant, supervised by Professor Xiaoyu Che — *Columbia University*

- Constructed both logistic regression and Cox hazard model to estimate the effect size; did power analysis to select the model and tested interaction term; After adjusted multiple comparison, found significant association between pseudo uridine and autism
- Applied Bayesian generalized linear models to calculate credible intervals and select analytes with Bayesian factor greater than 3
- Used Adaptive LASSO, Random Forest and XGBoosting to select features from high dimensional metabolomics dataset and fit a final model with selected feature to do prediction
- Constructed conditional logistic regression, feature selection and prediction base on one to one matched dataset
- A manuscript is prepared

COVID-19 Analysis via Logistic Curve and Clustering — 04/2020

- Course: The Advanced Topics in Statistical Computing — *Columbia University*

- Estimated parameters in logistic curve using the combination of Ordinary Differential Equation method and Gradient Descent model. First used ODE to get the initial guess and plug into gradient descent to estimate the parameters (No package used).
- Implemented both Gaussian Mixture Model with EM algorithm and Kmeans algorithm to cluster the estimated parameters for each country (No package used)
- Used the clustering results drawing world to visualize the distribution of clusters to see the geometric association

Implementation and optimization of algorithms on breast cancer diagnosis dataset 03/2020

- Course: The Advanced Topics in Statistical Computing *Columbia University*

- Built a predictive model based on logistic regression to facilitate cancer diagnosis
- Implemented and trained logistic regression models with Newton Raphson, Gradient Decent algorithms and a logistic-LASSO regression model with Pathwise Coordinate Descent from scratch in R
- The misclassification rate of logistic-LASSO model is the lowest by comparing models' performance by Cross Validation in R

TEACHINGS

Graduate Teaching Assistant 09/2020– 12/2020

- Course: BISTP8130 Biostatistical Methods I (95 students) *Department of Biostatistics, Columbia University*

- Prepared homework and exam problem set and solutions; Graded homework and exams; Held weekly office hours; Attended lectures and answered questions.

INTERNSHIPS

Eli Lilly China Shanghai, China, 09/2018– 06/2019

- Data Science& Solution Intern

- Supported data management work in clinical trials including data cleaning and missing data query under supervision of China DSS team and participated in the discussion of statistical methods applied for process design of clinical trials.
- Constructed quantitative analysis of possible interference risks during clinical trials and organization operation, developed an automatic web page for reproducible monthly analysis to improve efficiency

SKILLS & RELEVANT COURSES

- **Computer Skills**: R, SAS, Python, MATLAB, Microsoft Office Software.