Baode GAO

100 Haven Ave, New York, NY 10032
(+1)917-470-7259 bg2715@cumc.columbia.edu

EDUCATION

Mailman School of Public Health, Columbia University, New York, US Sept. 2021 – May 2023

- Master's in Biostatistics (Theory and Methods Track)
- 7 years of programming experience: solid in R lme4, tidyverse, caret, NetworkToolbox, etc., and Python Pandas, NumPy, PyTorch & Keras frameworks, Linux.
- Core Courses: Statistical Inference / Analysis of Longitudinal Data / Data Science Ⅰ & Ⅱ / Topics in Statistical Learning and Data Mining / Deep Learning / Statistical Methods for Causal Inference / Design of Medical Experiments

Department of Mathematical Science, Xi'an Jiaotong-Liverpool University (XJTLU), China Sept. 2016 – July. 2020

- Bachelor's in Applied Mathematics
- Contributed an article (as the first author) to ICMLC
- Core Courses: Multivariable Calculus / Advanced Linear Algebra / Abstract Algebra / Statistical Distribution Theory / Optimization Theory / Analysis Ⅰ & Ⅱ / Complex Analysis / Numerical Analysis / Partial Differential Equations

RESEARCH EXPERIENCE

LASSO improves the true detection rate of connectome-based predictive modeling June 2022 – Nov. 2022

Columbia University | School of Nursing | Student Research Worker

Advisor: Yihong Zhao, Professor of Data Sciences at Columbia University Medical Center, Columbia University

- Discovered the low true detection rate deficiency of the Connectome-based Prediction Model by the simulation study.
- Established penalized LASSO by features in networks to select the independent variable.
- Achieved higher true detection rate and prediction accuracy by variable selection methods in simulation and real data.

Covid-19 Infection Percentage Estimation by Transfer Learning Mar. 2022 – May 2022

Columbia University | The Fu Foundation School of Engineering and Applied Science | Codalab Competition

Advisor: Iddo Drori, Adjunct Associate Professor at the Department of Computer Science, Columbia University

- Split CT scans by subjects to avoid data leakage and enhanced model robustness by data augmentation – horizontal flip, rotation, random cropping, etc.
- Employed 4-fold cross-validation to adapt the uneven positive and negative covid images and small dataset.
- Reduced bias by halving the learning rate for every five epochs.
- Improved training speed and accuracy of results (MAE: 6.938, #36) by establishing the pre-trained ResNet50 and freezing parameters of the feature extraction part during training.

Subgroup Analysis Under Cure Rate Model June 2019 – Sept. 2019

XJTLU | Department of Mathematical Science | Research Assistant

Advisor: Xiaojun Zhu, Assistant Professor at the Department of Mathematical Science, XJTLU

- Deduced the expectations of latent variables, the approximation of the marginal likelihood, and the estimation of the baseline survival function of the subgroup cure rate model.
- Employed cox proportional-hazards model and logistic model to obtain the initial value of parameters in the survival function, cure rate function, and grouping function.
- Calculated expectations of hidden variables and estimated baseline survival function in E-step and estimated parameters by maximum likelihood estimation of the marginal likelihood in M-step.
- Compared our approach (AIC 1083.046) with Peng's Cure Rate Model (AIC 1090.103) (Likelihood Ratio Test, ratio 13.057; critical value 7.81; p-value 0.0045).

AWARDS AND ACHIEVEMENTS

- Honorable title of XJTLU Outstanding Student (2017-18, top 2.5%)
- Conference Travel Fellowship (2019, XJTLU)
- Student Research Worker at Irving Medical Center, Columbia University (2022, Paid)

PUBLICATIONS

- **Gao, B.**, Zhan, G., Wang, H., Wang, Y., & Zhu, S. (2019, February). Learning with Linear Mixed Model for Group Recommendation Systems. In *Proceedings of the 2019 11th International Conference on Machine Learning and Computing* (pp. 81-85). ACM.
- **Gao, B.**, Zhu, X. Subgroup Analysis Under Cure Rate Model. (To be submitted)