HANBIN LIU

Homepage: `https://hanbin-liu.github.io` ⋄ Email: liuhanbin1023@gmail.com

EDUCATION

Southern University of Science and Technology, Shenzhen, China *Sep. 2019-Present*
Major in Statistics

· GPA: 3.88/4.00 (ranking: 2/43, top5%)
· Full mark courses: Advanced Linear Algebra II (H), Probability Theory, Mathematical Statistics, Computational Statistics

SKILLS

R, Python, MATLAB, Java, LaTeX

RESEARCH EXPERIENCE

Prof. Jicong Fan's Group (Clustering Algorithms) *Jul. 2021-Sep. 2021*
Research Intern *CUHK-Shenzhen, School of Data Science*

· Learning sparse subspcae clustering (SSC) and selective sampling-based scalable sparse subspace clustering (S^5C). Deriving the kernel version of S^5C for non-linear cases (KS^5C).
· Learning ADMM and cyclic coordinate descent algorithm. Solving LASSO problems in KS^5C by using the cyclic coordinate descent algorithm rather than ADMM.
· Simulation of KS^5C on datasets COIL-20, COIL-100, GTSRB, YaleBCrop025, and MNIST. The Gaussian kernel has the best performance. [code]

Prof. Alice Cheng's Group (Content Analysis) *Jul. 2021-Aug. 2021*
(Remote) Undergraduate Research program (GEARS) *NC State, Department of Communication*

· Analyzing the corpus using a quantitative content analysis method, where the corpus is Weibo posts related to the debate on COVID-19. Creating a code sheet and coding the content. [poster]

SEMINAR

Biostatistics and Computational Statistics Seminar *Fall 2021*
Supervisor: Prof. Guo-Liang Tian

· Learning biostatistics with R. Learning the EM algorithm and its derivatives, QLB algorithm, and De Pierro algorithm.

PROJECT

Statistical Analysis of Short Video with R [code] [report] [slides] *Fall 2021*

· A linear model is used to perform regression analysis on the number of likes of short videos, and nonparametric methods are used to verify, such as permutation test and bootstrapping.
· Clustering and generative models based on the number of likes, comments, and shares, as well as a classification model using type as the label.

Data Mining Applications in DC Crime [code] [report] *Spring 2021*

· Data preprocessing, exploratory data analysis, and analysis of crimes over geography and time.
· Classifying (clustering) the geography by the crime events. Models include decision tree, KNN, random forest, AdaBoost, GBDT, K-means, and DBSCAN.

Self-contained Report on Discrete Mathematics [report] *Fall 2020*

· Stating the methods of solving linear recurrence relations in the language of linear algebra.

AWARDS & HONORS

· The 12th Chinese Mathematics Competitions (Provincial Fisrt Prize) *Dec. 2020*
· China National Encouragement Scholarship *Nov. 2020, 2021*

OTHERS

· My assignments of statistics major of SUSTech: `https://github.com/Hanbin-Liu/sustech-assign`