GUOYU KANG

1088 Xueyuan Road, Shenzhen, China, 518055
12032855@mail.sustech.edu.cn

EDUCATION

Southern University of Science and Technology *Shenzhen, China*
BS, Statistics and Data Science, GPA: 3.72/4.0 (Major 3.85/4.0) *Sep 2016 - Jun 2020*

University of California, Berkeley *Berkeley, CA*
Semester Exchange Program, first class scholarship (top 5%) *Jan 2020 - May 2020*

Southern University of Science and Technology *Shenzhen, China*
MS, Statistics and Data Science *Sep 2020 - Present*

RESEARCH INTERESTS

High-dimensional Data Analysis, Dimension Reduction, Variable Selection, Statistical Computing

AWARDS AND HONORS

- Honor graduate student of Southern University of Science and Technology. (top 2%) *July 2020*
- 1st Prize, National Mathematical Contest in Modeling (CUMCM) (top 5%) *Nov 2018*
- 1st and 2rd Prize, Outstanding Students Scholarship in SUSTech (top 5%) *Sep 2016 - Sep 2019*
- 2rd Prize, China Adolescents Science & Technology Innovation Contest (CASTIC) *Sep 2015*

PROJECTS

Early Diagnosis of Coronary Heart Disease with ML Methods *Mar 2021 - Apr 2021*

- Set up statistical model with logistic regression and machine learning methods (random forest (RF), SVM, K-NN, multi-layer sensor(MLP)) based on physical examination data attached with CAD condition.
- Forecasted whether a subject would potentially suffer from CAD and picked out the best prediction performance method from the above.

Functional Data Analysis Fruit Flies Data *Mar 2021 - Jun 2021*

- Established a functional linear model to fit the eggs laid during a certain lifespan of fruit flies under Fourier basis, b-spline basis functions.
- Applied dynamic function PCA to extract the prominent characteristics and predicted the confidence interval of lifespan with fertility.

Social Network Analysis on Open-Hub Data *Nov 2020 - Dec 2020*

- Explored social network models among contributors on open source platform (Black Duck Open Hub) from node and group levels and derived the most influential users and clusters for the entire network.
- Implemented data cleaning by SQL, and computed analysis and visualization by NetDraw and R.

Factor Analysis on Economic Development in Guangdong, China *May 2019 - Jun 2019*

- Evaluated the comprehensive social development among medical, education, infrastructure of 21 cities in the Guangdong province with orthogonal factor model based on data in the 2017 Statistics Almanac.
- Inferred various cities with various attributes and offered suggestions for further economic development.

Forecasting the Breakdown of Stock Market with LPPL Model
Computational Finance — Supervisor: Jingzhi Li *Oct 2018 - Nov 2018*

- Manipulated LPPL Model successfully (Log-Periodic Power Law) with MATLAB, 1stOpt software
- Examined the financial data on Shanghai composite index and predicted the bubble burst accurately.

RESEARCH EXPERIENCE

Feature Screening for High dimensional Statistics
Southern University of Science and Technology, Supervisor: Xin Chen *July 2020 - July 2021*

- Proposed an extended feature screening procedure and ranking index based on MV-SIS to achieve ultra-high dimensional variable selection with categorical response data.
- Evaluated the best predictive performance of the proposed method among typical screening methods.
- Hold the merits of mild condition required for model assumption and being robust dealing with heavy-tailed contribution or with outliers.

Graduate Research Program
Southern University of Science and Technology, Supervisor: Xin Chen *July 2021 - Present*

- Proposed a novel approach under high-dimension setting for applying an explicit form of inverse of covariance matrix (from debiassed lasso) into sliced inverse regression (SIR) .

MANUSCRIPT

- **Guoyu Kang**, Shanghui Yin, Xin Chen. Feature Screening For Ultrahigh-dimensional Categorical Data. [Abstract]

TEACHING ASSISTANT EXPERIENCE

- MA309 - Time Series Analysis *Sep 2019 - Jan 2020*
- MAT7100 - Statistical Deep Learning *Sep 2020 - Jan 2021*
- MA329 - Statistical Linear Model *Sep 2020 - Jan 2021*

SKILLS

Programming skills: R(proficient), SPSS, SAS, SQL, Matlab, Python, LaTex(proficient)
Extracurriculars: Photography, Boxing, Latin Dance
Languages: Chinese(native), English(fluent), Spanish(basic).