

Letian Yang

Email: letianya@usc.edu | Website: <https://lettyyang01.github.io/>

EDUCATION

University of Southern California - Los Angeles, CA, U.S. Aug.2023-May.2028(*Expected*)

Ph.D. student, Department of Data Science and Operations (Statistics); Advisor: Prof. Dennis Shen

Southern University of Science and Technology (SUSTech) - Shenzhen, Guangdong, China Aug.2019-Jun.2023

B.S., Department of Statistics and Data Science, Major in Statistics; GPA: 3.78/4.00

Honors:

- Summa Cum Laude in the College of Science
- National Scholarship (0.2%, Ministry of Education, China)

University of California, Berkeley - Berkeley, CA, U.S.

Jan.2022-Jul.2022

Exchange Student, Berkeley International Study Program; GPA: 4.00/4.00

RESEARCH INTEREST

High-dimensional statistics; causal inference; machine learning; deep learning.

ACADEMIC PROJECT

Algebraic and Statistical Properties of the Partially Regularized OLS Interpolator

Enhances the understanding of overparameterized linear regression with the partially regularized OLS estimator, complementing research on *benign overfitting* and advancing bias-variance trade-off analysis and predictive modeling for large-scale datasets.

[1] L. Yang, D. Shen. Algebraic and Statistical Properties of the Partially Regularized Ordinary Least Squares Interpolator (2024) <https://doi.org/10.48550/arXiv.2411.06593>

Refining Tissue-Specific Gene-Biological Process Relations with DNN

Advisor: Prof. Haiyan Huang (University of California, Berkeley); ongoing

Develop a novel procedure to refine Gene Ontology (GO) annotations in a biological and tissue-specific context using transcriptomic data and existing GO annotations. The procedure incorporates deep neural networks (DNN), clustering, and stability testing to enhance gene function analysis across different tissues within specific biological processes.

PROFESSIONAL EXPERIENCE

Data Analysis Internship, Peking University Shenzhen Hospital

Jun.2021-Aug.2021

- TCGA and GEO database analysis (Head and neck tumors)
 - Led efforts in data mining, cleaning, and preparation to ensure high-quality datasets for analysis. Conducted survival analysis and identified key prognostic factors to support clinical decision-making.
 - Performed single-gene expression analysis, identifying 60 potential target genes for further research.
- Survey on oral health knowledge and behavior of dental patients
 - Designed, distributed, and processed patient questionnaires to gather insights on oral health practices.
 - Analyzed behavioral trends across different age groups, with a focus on children under 15, enabling doctors to develop personalized care plans.

SKILLS SUMMARY

Programming Languages: R, Python, Java, MATLAB, LaTeX.

Systems and Tools: Linux, SSH, anaconda, AWS.

Statistical Modeling: Probability theory and measure theory, Stochastic process; Generalized linear models, Survival analysis, Bayesian inference; Multiple testing, High-dimensional statistics; Dimensionality reduction, Text mining, Time series; Exploratory data analysis.

Language: English (Fluent); Mandarin (Native Speaker).