

Yiqiu Shen

Los Angeles, CA | Email: yiqiushe@usc.edu | Mobile: 213-421-7745

EDUCATION

- **University of Southern California** Los Angeles, California
Ph.D Candidate - Business Statistics; GPA: 3.96
◦ **Field of Interest:** High Dimensional and Robust Statistics, Statistical Learning Theory
Aug 2019 - Present
- **University of Southern California** Los Angeles, California
Bachelor of Science - Applied and Computational Mathematics; GPA: 3.92; summa cum laude
Aug 2016 - May 2019

SKILLS SUMMARY

- **Programming Languages:** R, MATLAB, Python, SQL, Java
- **Softwares:** \LaTeX , Microsoft Excel, JMP, Mathematica
- **Languages:** English (Proficient), Japanese (Limited), Chinese (Native)

TEACHING EXPERIENCE

- **Introduction to Business Analytics** *Present, 8 sessions, 50 students each*
 - Providing in-class software support and instructing Excel and JMP tutorials.
 - Designing quiz questions based on real life datasets for A/B testing and classification.
- **Deep Learning and Business Application** *Spring 2020, 1 session, 20 students*
 - Instructed Python and R tutorials on implementing SGD, PCA, MLP classifier manually and creating CNN and RNN via Tensorflow and Keras.
 - Showcased various deep learning applications on subjects above.
- **Introduction to MATLAB** *Spring and Fall 2017, 4 sessions, 20 students each*
 - Provided in-and-after-class support on MATLAB learning experience and engineering projects.

APPLIED PROJECTS

- **Membership Identification in Large Scaled Sparse Networks**
 - Implemented SIMPLE in both R and Python (<https://arxiv.org/abs/1910.01734v1>) and applied to Python-scraped S&P1000 company data to **identify clusters within networks**.
 - Improved calculation efficiency by utilizing parallel computation.
- **High Dimensional Classification via Spiked Eigenvalue Theory**
 - Wrote the Python code for a novel classification method based on a combination of spiked eigenvalue theory and dimension reduction theory and applied to identify virus and bacteria given codon information separated from their genetic sequence data.
- **Pairwise Relationship Identification via Mixed Neural Networks**
 - Constructed a novel combination of CNN and RNN in Tensorflow. Incorporated NLP methods and applied to identify virus and host relationship given their genetic sequence data.

THEORETICAL PROJECTS

- **Average Case Analysis of Lasso under Adversarial Contamination and Correlated Design**
 - *Joint Work with Dr. Stanislav Minsker*
 - Showed that the prediction risk of Lasso converges at **fast rate** with high probability under a **highly correlated** (coherent) and **outlier-contaminated** design. Achieved the above results under **mildest** conditions on sparsity of coefficient vector and design matrix. Drew a novel connection between principal angle between subspace spanned by columns of the design matrix and the Euclidean space and prediction risk of Lasso.
- **Minimax Supervised Clustering in the Anisotropic Gaussian Mixture Model**
 - *Joint Work with Dr. Mohamed Ndaoud and Dr. Stanislav Minsker*
 - Derived a lower and a matching upper bound for the minimax risk of supervised clustering under the two-component **anisotropic** Gaussian mixture model in **high dimension** and **non-asymptotic** setting. Characterized precisely the risk of ℓ_2 -regularized supervised least squares classifiers. Deduced that the **interpolating solution may outperform the regularized classifier** and that **interpolation can be robust to corruption in the covariance of the noise** when the signal is aligned with the "clean" part of the covariance.
 - **Manuscript available at:** <https://arxiv.org/abs/2111.07041>

HONORS AND AWARDS

- USC Marshall Graduate Assistantship and Fellowship
- USC Dornsife Summer Undergraduate Research Fund

VOLUNTEER EXPERIENCE

- **Accreditation and Media Assistant at World Athletics Championships 2022** *July 2022*
- **Personal Statement Supporter at 826LA** *Jan 2017 - Dec 2017*