

# Yang Feng

## Curriculum Vitae

November 2022

Address: Department of Biostatistics  
School of Global Public Health  
New York University  
New York, NY 10003, USA  
Phone: +1 (212) 992-3810  
Email: yang.feng@nyu.edu  
Website: <https://yangfeng.hosting.nyu.edu/>

### EDUCATION

2006 B.S. University of Science and Technology of China Mathematics  
2010 Ph.D. Princeton University Operations Research  
Thesis: High-dimensional Statistical Learning and Nonparametric Modeling  
Advisor: Prof. Jianqing Fan

### PROFESSIONAL EXPERIENCE

2019–Present Associate Professor, Department of Biostatistics, School of Global Public Health,  
New York University  
2022–Present Affiliate Faculty, Center for Data Science, New York University  
2021–Present Affiliate Faculty, Center for Practice and Research at the Intersection of  
Information, Society, and Methodology (PRIISM), New York University  
2016–2019 Associate Professor, Department of Statistics, Columbia University  
2014–2016 Howard Levene Assistant Professor, Department of Statistics, Columbia University  
2010–2014 Assistant Professor, Department of Statistics, Columbia University

### PROFESSIONAL MEMBERSHIPS

- American Statistical Association (Fellow and Lifetime Member)
- Institute of Mathematical Statistics (Lifetime Member)
- International Statistical Institute (Elected Member)
- International Chinese Statistical Association (Lifetime Member)

### RESEARCH INTERESTS

- Theory, Methods, and Algorithms: Machine Learning, High-dimensional Statistics, Social Networks, Neyman-Pearson Classification, Nonparametric and Semi-parametric Statistics.
- Applications: Dementia, Bioinformatics, Cancer Diagnosis, Electronic Health Care Data, Epidemiology, Computer Vision.

### EDITORIAL ACTIVITIES

- Editorial Board
  - 2021 – Present Associate Editor, Journal of the American Statistical Association: Theory and Methods
  - 2020 – Present Associate Editor, Stat
  - 2018 – Present Associate Editor, Journal of Business and Economic Statistics
  - 2014 – Present Associate Editor, Statistica Sinica
  - 2013 – 2022 Associate Editor, Statistical Analysis and Data Mining, The ASA Data Science Journal
  - 2015 – 2018 Associate Editor, Computational Statistics and Data Analysis

## HONORS & AWARDS

- 2022 Fellow, American Statistical Association
- 2022 Finalist, NYU School of Global Public Health Teaching Excellence Award
- 2021 NYU University Research Challenge Fund
- 2020 NYU Curriculum Development Challenge Fund Award
- 2017 Elected Member, International Statistical Institute
- 2016 NSF CAREER Award
- 2015 Lenfest Junior Faculty Development Award
- 2012 New World Mathematics Award (Silver Prize)
- 2010 Wallace Memorial Honorific Fellowship (the highest award for a Princeton graduate student)
- 2009 Laha Award from the Institute of Mathematical Statistics (IMS)
- 2007 The Gordon B. and Nancy R. Stewart, Jr. Fellowship, Princeton University
- 2005 Samsung Scholarship (11 out of about 7200 undergraduate USTC students)

## RESEARCH GRANTS

2022–2026	National Institutes of Health (NIH) 1R01NS122987-01A1 Investigator (PI: Matija Snuderl) “Inducing neural maturation in medulloblastoma by targeting EZH2”	\$175,830
2022–2027	National Institutes of Health (NIH) 1R01CA268932-01A1 Investigator (PI: Jennifer Cantrell) “Using Multiphase Optimization Strategy (MOST) to Optimize a Cost-effective, Sustainable and Scalable Smoking Cessation Package for Smokers in HIV Clinical Care”	\$148,438
2021–2023	National Institutes of Health (NIH) 1R21AG074205-01 Principal Investigator “Multiclass classification under prioritized error control and specific error costs with applications to dementia classification”	\$438,473
2022–2023	National Institutes of Health (NIH) 1R56NS122987-01 Investigator (PI: Matija Snuderl) “Inducing neural maturation in medulloblastoma by targeting EZH2”	\$53,300
2021–2022	NYU University Research Challenge Fund Principal Investigator “Prioritized Multiclass Classification with Applications to Brain Tumor Diagnosis”	\$12,000
2020–2021	NYU Curriculum Development Challenge Fund Principal Investigator “Interactive Teaching and Learning of Statistical Programming and Machine Learning in R”	\$4,500
2020-2021	National Science Foundation (NSF) Grant DEB-2034022 Co-Principal Investigator (PI: Joshua Epstein) “RAPID: Behavioral Epidemic Modeling For COVID-19 Containment”	\$131,954
2016–2021	National Science Foundation (NSF) <a href="#">DMS-1554804</a> Principal Investigator “CAREER: Statistical inference of network and relational data”	\$400,000
2013–2016	National Science Foundation (NSF) <a href="#">DMS-1308566</a> Principal Investigator “Nonparametric classification, tuning parameter selection, and asymptotic stability for high-dimensional data”	\$129,980

## SELECTED PUBLICATIONS

Note: <sup>1</sup> represents co-first authors, \* represents the corresponding author(s), underline represents students and junior collaborators, and <sup>†</sup> represents the author list is alphabetically ordered according to the mathematics convention. Full Publication List on Google Scholar (**3300+ Citations** as of Oct 2022)

<https://scholar.google.com/citations?user=QXHb8CcAAAAJ&hl=en>.

1. Tian, Y. and **Feng, Y.\*** (2022), Transfer Learning under High-dimensional Generalized Linear Models, *Journal of American Statistical Association*, to appear.
2. Yuan, M., Liu, R., Feng, Y.\*, and Shang, Z. \* (2022), Testing Community Structures for Hypergraphs, *Annals of Statistics*, 50(1): 147-169.
3. Demirkaya, E.<sup>1</sup>, **Feng, Y.<sup>1</sup>**, Basu, P., and Lv, J. (2022), Large-scale model selection in misspecified generalized linear models, *Biometrika*, 109(1), 123-136.
4. Tian, Y. and **Feng, Y.\*** (2021). RaSE: A variable screening framework via random subspace ensembles. *Journal of the American Statistical Association*, to appear.
5. Tian, Y. and **Feng, Y.\*** (2021). RaSE: Random subspace ensemble classification, *Journal of Machine Learning Research*, 22(45): 1-93.
6. Tang, F.<sup>1</sup>, **Feng, Y.\*<sup>1</sup>**, Chiheb, H., Fan, J. (2021), The Interplay of Demographic Variables and Social Distancing Scores in Deep Prediction of U.S. COVID-19 Cases, *Journal of the American Statistical Association*, **116(534)**, 492-506.
7. Tong, X., Xia, L., Wang, J., and Feng, Y.\* (2020), Neyman-Pearson classification: parametrics and sample size requirement. *Journal of Machine Learning Research*, **21(12)**, 1-48.
8. <sup>†</sup>Fan, J., Feng, Y.\*, and Xia, L.\* (2020), A Projection Based Conditional Dependence Measure with Applications to High-dimensional Undirected Graphical Models, *Journal of Econometrics*, 218, 119-139.
9. Liu, J., Psarakis, E., Feng, Y.\*, and Stamos, I. (2019), A kronecker product model for repeated pattern detection on 2D urban images, *IEEE Transactions on Pattern Analysis and Machine Intelligence<sup>1</sup>*, **41**, 2266-2272.
10. Hao, N.<sup>1</sup>, **Feng, Y.<sup>1</sup>** and Zhang, H.H. (2018), Model selection for high dimensional quadratic regression via regularization, *Journal of the American Statistical Association*, **113**, 615-625.
11. Tong, X.<sup>1</sup>, **Feng, Y.<sup>1</sup>**, and Li, J. (2018), Neyman-Pearson classification algorithms and NP receiver operating characteristic, *Science Advances<sup>2</sup>*, Vol. 4, no. 2, eaao1659.
12. Fourati, S., Talla, A., Mahmoudian, M., Burkhart, J. G., Klen, R., Henao, R., ... & Sieberts, S. K. (2018). A crowdsourced analysis to identify ab initio molecular signatures predictive of susceptibility to viral infection. *Nature Communication*, 9(1), 1-11.
13. Ji, J., He, D., Feng, Y., He, Y., Xue, F., Xie, L. (2017), JDINAC: joint density-based non-parametric differential interaction network analysis and classification using high-dimensional sparse omics data, *Bioinformatics*, **33**, 3080-3087.
14. Zhao, A., Feng, Y., Wang, L., and Tong, X. (2016), Neyman-Pearson classification under high-dimensional settings, *Journal of Machine Learning Research*, **17**, 1-39.
15. <sup>†</sup> Fan, J., **Feng, Y.\***, Jiang, J., and Tong, X. (2016), A classification rule of feature augmentation via nonparametrics and selection (FANS) in high dimensional space, *Journal of the American Statistical Association*, **111**, 275-287

<sup>1</sup>A leading journal in computer vision. 2018 Impact Factor: 17.73.

<sup>2</sup>A high-profile multidisciplinary journal in the Science series published by the American Association for the Advancement of Science. 2018 JCR Impact Factor: 12.804

16. <sup>†</sup> Fan, J., **Feng, Y.**, and Tong, X. (2012), A road to classification in high dimensional space: the regularized optimal affine discriminant, *Journal of the Royal Statistical Society Series B*, **74**, 745-771.
17. <sup>†</sup> Fan, J., **Feng, Y.\*** and Song, R. (2011), Nonparametric independence screening in ultra-high dimensional additive models, *Journal of the American Statistical Association*, **106**, 544-557.
18. <sup>†</sup> Fan, J., **Feng, Y.** and Niu, Y. (2010), Nonparametric estimation of genewise variance for microarray data, *The Annals of Statistics*, **38**, 2723-2750.
19. MAQC-II Consortium (2010), MAQC-II Project: A comprehensive survey of common practices for the development and validation of microarray-based predictive models, *Nature Biotechnology*, **28**, 827-841.
20. Fan, J., Wu, Y., and **Feng, Y.** (2009), Local quasi-likelihood with a parametric guide, *The Annals of Statistics*, **37**, 4153-4183.

## SOFTWARE

The following software packages have been downloaded over **310,000** times as of Oct 2022, according to CRAN.

1. **SIS**, an R package for (Iterative) Sure Independence Screening for Generalized Linear Models and Cox Proportional Hazards Models, available at [CRAN](#).
2. **glmtrans**, an R package for implementing transfer learning under high-dimensional generalized linear models, along with the construction of confidence intervals based on a new debiasing technique, available at [CRAN](#).
3. **RaSEn**, an R package for implementing a new model-free ensemble classification framework, RaSE algorithm, for the sparse classification problem, available at [CRAN](#).
4. **nproc**, Given a sample of class 0 and class 1 and a classification method, the package generates the corresponding Neyman-Pearson classifier with a pre-specified type-I error control and Neyman-Pearson Receiver Operator Curve, available at [CRAN](#). A vignette for demonstration is available [here](#).
5. **r02pro**, a companion R package including the interactive exercises and datasets for the book "R Programming: Zero to Pro", available at [CRAN](#).
6. **RAMP**, an R package for fitting the entire solution path for high-dimensional regularized generalized linear models with interactions effects under the strong heredity constraint, available at [CRAN](#).
7. **MIRL**, an R package for multiple imputation random lasso, designed to solve the high-dimensional variable selection problem with missing data, available at [CRAN](#).
8. **fcd**, an R package for implementing the fused community detection method, available at [CRAN](#).
9. **fusedPCA**, an R package for implementing the fused principle component analysis method, available at [CRAN](#).
10. **FANS**, Matlab code for implementing the FANS (Feature Augmentation via Nonparametrics and Selection) classification method for high-dimensional data, available at [GitHub](#).
11. **CLBIC**, R code for implementing Composite Likelihood BIC for selecting the number of communities, available at [GitHub](#).
12. **pgraph**, Implements a general framework for creating dependency graphs using projection. Both lasso and sparse additive model projections are implemented. Both Pearson correlation and distance covariance can be used to generate the graph, available at [CRAN](#).
13. **apple**, an R package for calculating the Approximate Path for Penalized Likelihood Estimators for Generalized Linear Models, available at [CRAN](#)

14. **ROAD**, a Matlab package designed for the Regularized Optimal Affine Discriminant method for high-dimensional classification, available at [GitHub](#).
15. **xtab**, An R function for generating LaTeX tables from a data matrix, available at [GitHub](#).

#### REFERENCES

- Available upon request.