

Jingyi Kenneth Tay

Curriculum Vitae

✉ kjytay@gmail.com

📄 kjytay.github.io/

🌐 [kjytay](#)

📱 [kjytay](#)

Education

- 2016–2021 **Ph.D. Candidate in Statistics**, *Stanford University*, Stanford, CA.
Advised by Robert Tibshirani. Jerome H. Friedman Applied Statistics Dissertation Award.
- 2016–2019 **M.S. in Statistics**, *Stanford University*, Stanford, CA.
- 2006–2010 **A.B. in Mathematics**, *Princeton University*, Princeton, NJ.
Summa Cum Laude. Certificates in Program of Applied & Computational Mathematics,
Program of Finance.
- 2004–2005 **GCE 'A' Levels**, *Anglo-Chinese Junior College*, Singapore.
4 As, 2 Special Paper Distinctions

Research

Publications

1. **J. K. Tay**, N. Aghaeepour, T. Hastie, and R. Tibshirani. (2021). Feature-weighted elastic net: using “features of features” for better prediction. *Statistica Sinica*, 2021. URL <https://doi.org/10.5705/ss.202020.0226>. R package `fwelnet`.
2. D. Shung, J. Huang, E. Castro, **J. K. Tay**, M. Simonov, L. Laine, R. Batra and S. Krishnaswamy. (2021). Neural network predicts need for red blood cell transfusion for patients with acute gastrointestinal bleeding admitted to the intensive care unit. *Scientific Reports*, 2021, 11:8827.
3. **J. K. Tay**, J. Friedman, and R. Tibshirani. (2021). Principal component-guided sparse regression. *Canadian Journal of Statistics*, 2021. URL <https://doi.org/10.1002/cjs.11617>. R package `pcLasso`.
4. D. Shung, C. Tsay, L. Laine, D. Chang, F. Li, P. Thomas, C. Partridge, M. Simonov, A. Hsiao, **J. K. Tay**, and A. Taylor. (2021). Early identification of patients with acute gastrointestinal bleeding using natural language processing and decision rules. *Journal of Gastroenterology and Hepatology*, 2021, 36(6):1590-7.
5. **J. K. Tay**, and R. Tibshirani. (2020). Reluctant generalized additive modeling. *International Statistical Review*, 2020, 88(S1):S205-S224. R package `relgam`.
6. D. L. Shung, B. Au, R. A. Taylor, **J. K. Tay**, S. B. Laursen, A. J. Stanley, H. R. Dalton, J. Ngu, M. Schultz, and L. Laine. (2020). Validation of a machine learning model that outperforms clinical risk scoring systems for upper gastrointestinal bleeding. *Gastroenterology*, 2020, 158(1):160-7.

Software

1. Author of `cvwrapr` R package. Tools for performing cross-validation.
2. Contributor to `glmnet` R package. v4.0: Extended `glmnet` to efficiently fit any generalized linear model with the elastic net penalty. v4.1: Added ability to fit stratified Cox models and Cox models for start-stop data, opening the way to fit a wide array of regularized Cox models (e.g. time-dependent covariates, left truncation, multiple events per subject).

Preprints and Theses

1. E. Tuzhilina, T. J. Hastie, D. J. McDonald, **J. K. Tay**, and R. Tibshirani. (2022). Smooth multi-period forecasting with application to prediction of COVID-19 cases. *arXiv:2202.09723 [stat.ME]*, 2022. URL <https://arxiv.org/abs/2202.09723>.
2. **J. K. Tay**. (2021). Extending the reach of the lasso and elastic net penalties: Methodology and practice. PhD Thesis. Advisor: Robert Tibshirani.
3. **J. K. Tay**, B. Narasimhan, and T. Hastie. (2021). Elastic net regularization paths for all generalized linear models. *arXiv:2103.03475 [stat.CO]*, 2021. URL <https://arxiv.org/abs/2103.03475>. R package `glmnet`.
4. **J. K. Tay**, and R. Tibshirani. (2018). A latent factor approach for prediction from multiple assays. *arXiv:1807.05675 [stat.ME]*, 2018. URL <https://arxiv.org/abs/1807.05675>.
5. **J. K. Tay**. (2010). Maximizing expected logarithmic utility in a regime-switching model with inside information. Senior Thesis. Advisor: Ramon van Handel.
6. **J. K. Tay**. (2009). Construction of space-time block codes from a decoding point of view. Junior Independent Work. Advisor: Robert Calderbank.

Teaching

Course Instructor

Autumn 2019-20	STATS 32, Introduction to R for Undergraduates	Stanford University
Autumn 2018-19	STATS 32, Introduction to R for Undergraduates	Stanford University
Summer 2017-18	STATS 302, Qualifying Exams Workshop (Applied Statistics).	Stanford University
Autumn 2017-18	STATS 32, Introduction to R for Undergraduates	Stanford University

Teaching Assistant

Winter 2020-21	CS 229M/STATS 214, Machine Learning Theory	Stanford University
Autumn 2020-21	STATS 200, Introduction to Statistical Inference	Stanford University
Spring 2019-20	STATS 315B, Modern Applied Statistics: Data Mining	Stanford University
Winter 2019-20	STATS 216, Introduction to Statistical Learning	Stanford University
Winter 2018-19	STATS 191, Introduction to Applied Statistics	Stanford University
Summer 2017-18	STATS 216V, Introduction to Statistical Learning	Stanford University
Spring 2017-18	STATS 305C, Methods for Applied Statistics II: Applied Multivariate Statistics	Stanford University
Winter 2017-18	STATS 216, Introduction to Statistical Learning	Stanford University
Summer 2016-17	STATS 116, Theory of Probability	Stanford University
Summer 2016-17	STATS 290, Paradigms for Computing with Data	Stanford University
Autumn 2016-17	STATS 116, Theory of Probability	Stanford University

Work Experience

- 11/2022– Present **Senior Data Scientist (Optimization, Data & Artificial Intelligence Foundations)**, *Linkedin*, Mountain View, CA.
- 09/2021– 11/2022 **Senior Data Scientist (Experimentation Science, Data Science & Research Productivity)**, *Linkedin*, Mountain View, CA.
- 06/2020– 09/2020 **Data Scientist Intern (Payments Data Science)**, *Google*, Sunnyvale, CA.
 - o Developed novel method and R package for computing variance for post-stratified estimator in potential outcomes setting. In one application, confidence interval width was reduced by 11%.
 - o Developed new algorithm that reports an experiment's heterogeneous treatment effect concisely while ensuring statistical validity. This work enables analysts to quickly understand how the treatment varies along dimensions of interest.
- 06/2019– 09/2019 **Applied Scientist Intern (Search Relevance)**, *A9.com, Amazon Search*, Palo Alto, CA.
 - o Conceptualized and constructed data pipelines for new, granular metrics for Amazon search relevance models. Processed ~1B queries and ~20B item responses to obtain dataset for predictive modeling.
 - o Built a model based on these pre-experiment metrics to predict performance on live customer traffic, so that experimental bandwidth can be allocated more efficiently. Model improved test performance metric by 20% over baseline.

- 10/2015–08/2016 **Data Scientist (Data Science Division)**, *Infocomm Development Authority of Singapore*, Singapore.
- Spearheaded engagements with a wide array of government agencies (economic, transport, social) to analyze their data to support public policymaking. Responsibilities included project scoping, data cleaning, visualization, statistical analysis and presentation of results.
 - Systematized and tested recruitment framework and materials for all roles in the division, including data scientist, quantitative strategist and front-end developer.
 - Developed division's operating policy for data management and statistical disclosure control.
- 04/2014–03/2015 **Staff Well-Being Committee Chair**, *Ministry of the Environment & Water Resources*, Singapore.
- Led team of 10 officers in conceptualizing and executing activities to improve staff welfare and morale.
 - Managed and accounted for budget (20K+) for staff welfare.
- 09/2013–09/2015 **Assistant Director (Environmental Policy Division)**, *Ministry of the Environment & Water Resources*, Singapore.
- Drove progressive policies to ensure sustainability and efficiency of Singapore's waste management system.
 - Facilitated Ministry's strategic planning by conceptualizing and organizing Policy Wing retreat.
 - Evaluated usefulness of movement data in predicting spread of *Chikungunya* virus in Singapore.
- 04/2012–08/2013 **InfoComm Technologies Engineer**, *Ministry of Defense*, Singapore.
- Evaluated operational performance of critical communications systems.
 - Made operational effectiveness more transparent by developing and implementing a new reporting dashboard for senior management.
 - Strengthened in-house user adoption of systems through crafting and delivering technical presentations.

Awards and Honors

- 2021 Jerome H. Friedman Applied Statistics Dissertation Award.
- 2019, 2020 Honorable Mention, American Statistical Association's Statistical Learning and Data Science Student Paper Competition.
- 2017, 2018 Departmental Teaching Assistant Award.
- 2017 Two Sigma Graduate Fellowship in Statistics.
- 2010 Election to Sigma Xi Honor Society.
- 2009 Early Induction to Phi Beta Kappa Honor Society (top 1% of cohort).
- 2007, 2008 Shapiro Prize for Academic Excellence, Princeton University.
- 2007 Second Prize, International Mathematics Competition.
- 2006, 2008 Honorable Mention, William Lowell Putnam Competition.

- 2006-2010 Public Service Commission Overseas Merit Scholarship (Open) (full-ride college scholarship), Public Service Commission, Singapore.
- 2006 Lee Kuan Yew Award for Mathematics & Science, Ministry of Education, Singapore.
- 2005 Gold Award (top score), Asian-Pacific Mathematical Olympiad.
- 2004, 2005 Silver Medal, International Mathematical Olympiad.
- 2003 Honorable Mention, International Mathematical Olympiad.