

Yue Li

yuel5@andrew.cmu.edu

EDUCATION	Ph.D. in Statistics , Carnegie Mellon University	2017–(exp.) 2022
	Advisors: Yuting Wei, Kathryn Roeder	
	M.S. in Machine Learning , Carnegie Mellon University	2017–(exp.) 2020
	B.S. in Statistics and Probability , Peking University	2013–2017
	B.A. in Economics (minor), Peking University	2014–2017
SKILLS	Languages Chinese (<i>native</i>), English (<i>working proficiency</i>)	
	Programming R, Python (PyTorch, TensorFlow, Jupyter Notebook), $\LaTeX$, Matlab	
RESEARCH	Random projection test for high-dimensional linear regression	Oct 2019–Mar 2020
	Supervised by Yuting Wei	
	- Proposed a random-projection based F-test for the global null in high-dimensional linear model. - Proved that the randomized procedure has a non-random signal strength and is minimax optimal among all possible tests when the model possesses a low intrinsic dimension. - Accepted by <i>NeurIPS 2020</i>; check out our preprint here.	
	Distribution of the sketched Lasso solution: a new tool for fixed-design Lasso analysis	Mar 2020–
	Supervised by Yuting Wei	
	- Derived the distribution of the sample-wise sketched Lasso solution using Gordon’s min-max theorem, which leads to new findings about the fixed-design Lasso solution. - When sparsity/feature dimension is approximately constant, both problems remain open for years.	
	Transfer learning and integration of large single-cell gene expression datasets	Mar–Sep 2020
	Supervised by Kathryn Roeder and Yuting Wei	
	- Designed a transfer learning mechanism based on a shared low-dimensional factor space, which helps us to integrate and transfer information between single-cell gene expression datasets. - Improved clustering results significantly by integrating data from different technologies, labs and species. - Submitted to <i>Genome Research</i>; check out our preprint here. Further theoretical analysis in progress.	
	Modeling and imputation of single-cell gene expression datasets	2015–2017
SELECTED COURSE PROJECTS	Supervised by Hao Ge and Hongyu Zhao	
	- Constructed a hierarchical Bayesian model to account for the variance components in single-cell data. - Applied this model for dropout imputation and batch effect correction. - Presented in <i>NeurIPS 2019 Workshop on Machine Learning in Computational Biology</i>.	
	Motion-Object-Background disentangled representation from video	Spring 2019
	Course project of 10-708 <i>Probabilistic Graphical Models</i> ; see our final report here .	
	Based on the SOTA content-motion disentangled learning architecture for video prediction task, we proposed a Motion-Object-Background disentanglement, obtaining better performances in certain scenarios.	
	Speaker recognition using deep neural network	Spring 2018
	Course project of 11-785 <i>Introduction to Deep Learning</i> ; see our poster here .	
	We designed and evaluated three different architectures on a standard speech dataset: (1) CNN + classification model; (2) attention-based model and (3) adversarial autoencoders.	
	Stochastic neighbor topic model	Fall 2017
	Course project of 10-715 <i>Advanced Introduction to Machine Learning</i> ; see presentation slides here .	
AWARDS	Inspired by t-SNE, we introduced a stochastic neighbor based loss in the LDA objective function, which can be interpreted as a regularized prior on topic distribution. The model is fitted with ADMM.	
	Silver Medal , Chinese Mathematical Olympiad	2013
	5 · 4 Youth Scholarship , Peking University	2014, 2015
	Outstanding Graduates , Peking University	2017