Zhenfeng Lin

Tel: +1 443 808 3242, Email: zflin@stat.tamu.edu
Linkedin: https://www.linkedin.com/in/zhenfeng-lin-3141a39b/
1600 Southwest Pkwy, Apt. 220, College Station TX, 77840

SUMMARY

- Expected graduation with Statistics Ph.D. degree in May 2019
- Looking for a Data Science internship in 2018 Summer
- Research focus: Machine learning in classification and prediction; Gaussian processes; Big and high-dimensional data analysis
- Proficient in R, Python, C/C++, SQL, MATLAB, and SAS

EDUCATION

06/2015~06/2019 (expected) **Texas A&M University, USA**
Degree: Ph.D. Candidate in Statistics
08/2017~08/2017 **La Serena Winter School for Data Science, Chile**
Visiting scholar program: Applied Tools for Data-driven Sciences
09/2011~07/2013 **Sun Yat-Sen University, China**
Degree: M.S. in Probability & Mathematical Statistics
09/2007~07/2011 **Sun Yat-Sen University, China**
Degree: B.S. in Mathematics & Applied Mathematics

PROFESSIONAL EXPERIENCE

05/2016~08/2016 **Data Analyst Intern at Chevron, Houston, USA**
Designed a work-flow to predict production of shale oil wells using R. The work-flow involves standard machine learning techniques:

- Data cleaning that deals with outliers and missing values
- Feature engineering that creates and selects important predictors
- Prediction modeling that involves random forest, gradient boosting, neural network, SVM, multiple general linear regression, and LASSO
- Parameter tuning that uses grid search to optimize model performance
- Cross validation that compares performance of different models
- Data visualization that smartly shows data's natural properties and prediction performance

07/2013~05/2015 **Statistics Researcher at OriginLab CO., Guangzhou, China**
Led the research and development of statistics algorithms for a popular scientific graphical software (OriginLab®) using C/C++/R/MATLAB. Developed built-in algorithms including Bivariate Gaussian Kernel Density Estimation, Bootstrapping, Distribution Fitting, and Unbalanced Repeated Measure ANOVA

06/2012~08/2012 **Intern at Yuexiu Group - Yuexiu Industrial Investment CO., Guangzhou, China**

- Developed algorithm to recognize macroeconomic cycles for asset allocation: identified turning points of the selected indicators and similar historical scenarios, using current leading indicators for judgment
- Researched how cash flows can be matched for personal finance plans: with reference to ING retirement Plan, simulated cash flows under various investment style situations
- Designed evaluation standards for bank financial products: performed comprehensive analysis based on trading, selecting and performance

RESEARCH & ACADEMIC EXPERIENCE

04/2016~04/2016 Development of an Optimal Credit Card Transaction Fraud Prevention Strategy – 2016 Capital One Student Modeling National Competition, 2nd place

Role: Team Leader

Finalist to present at Capital One Financial Corporation, VA

- Analyzed Big (~10 millions observations with hundreds of features) and Dirty (many outliers, missing) data
- Extracted credit card profiles from real transactions and created Recency, Frequency, Monetary (RFM) features to predict fraudulent transaction
- Performed reasonable data segmentation and built Ensemble eXtreme Gradient Boosting (EXGB) models

Our work resulted in a near perfect ROC curve with low insample/outsample misclassification rate.

06/2015~04/2016 CryoEM 3D Nano-scale Particle Structure Reconstruction Problem, TAMU

Role: Research Assistant

Challenge: highly noisy raw data, heavy computational load, as well as unknown projection angles and unknown 3D reference particles

- Reconstructed nano-scale particle structures by creating 3D density maps based on over 100,000 CryoEM 2D projection images
- Applied parallel programming and machine learning techniques like EM algorithm, regularization, and bootstrapping methods

09/2011~06/2013 M.S. thesis: Robust Control of Contracting Discrete-time Markov Decision Processes (DTMDPs) with First Passage Expectation Criteria

Role: Student Researcher

- Integrated the first passage expectation criteria into DTMDPs to optimize the system performance
- Controlled the transition law in a fuzzy set with uncertainty

10/2012~10/2012 Gene Recognition Algorithm -The 9th National Graduate Mathematical Contest in Modeling

Role: Team Member

- Established indicator sequences from training DNA sequence (mitochondrial gene of human) and obtained DNA power spectrum sequence by using fast discrete fourier transform (FDFT)
- Created ROC to estimate a threshold value for human species and predicted exon regions for the target DNA sequences

10/2011~10/2011 Analysis and Evaluation of Lodging Resistance in Wheat - The 8th National Graduate Mathematical Contest in Modeling

Role: Team Leader

- Identified the most important predictors of wheat lodging resistance by correlation analysis
- Established multiple linear regression of lodging index using R

PUBLICATIONS & PRESENTATIONS

1. **Zhenfeng Lin** and James P. Long. Mixture Proportion Estimation via Dimension Reduction with Classifier. *Rice 2017 Data Science Conference.* Poster presentation. Houston, TX, October 9-10, 2017.
2. **Zhenfeng Lin** and James P. Long. Mixture Proportion Estimation for Positive-Unlabeled Learning via Classifier Dimension Reduction. Submitted. *American Statistical Association (ASA) 2018 Joint Statistical Meetings (JSM) Student Paper Competition*.
3. **Zhenfeng Lin**. Fitting Multi-band Gaussian Processes Mira Model with RSTAN (HMC). Course project presentation. College Station, TX, November 20, 2017.

4. **Zhenfeng Lin**. Learning From Noisy Labels via Modified Logistic Regression. *Southeast Texas Chapter of the American Statistical Association (SETCASA) Poster Session*. Poster presentation. College Station, TX, April 21, 2017.

AWARDS& RECOGNITIONS

2017	3rd Prize, Southeast Texas Chapter of the American Statistical Association Poster Competition
2016	2nd Place, Capital One Student Modeling National Competition
2015	College of Science Lechner Fellowship, TAMU
2015	OGAPS Dean's Doctoral Fellowship, TAMU
2011-2013	Excellent Graduate Scholarship, SYSU
2011	1st Prize, 8th National Graduate Mathematical Contest in Modeling
2008-2010	1st Class Excellent Student Scholarship, SYSU