

Guangwei Weng

425 13th Ave SE, Apt 604
Minneapolis, MN, 55414

612-245-9088
wengx076@umn.edu

Education

- University of Minnesota, Twin Cities** Minneapolis, MN
• *Ph.D., Statistics, Research Fields: Nonparametric Statistics, Machine Learning* Sep. 2015 - Present
- Fudan University** Shanghai, China
• *B.Sc. Statistics* Sep. 2010 - Jun. 2015

Skills

- R, Python, C, SQL, Fortran, Deep Learning, Linux (Debian, Ubuntu), MATLAB, SAS.

Work Experience

- School of Statistics, University of Minnesota** Minneapolis, MN
• *Teaching Assistant (Intro to Statistical Analysis & Theory of Statistics)* Jan. 2016 - Present
- Statistical Consulting Center, University of Minnesota** Minneapolis, MN
• *Summer Statistical Consultant* Jun. 2017 - Aug. 2017
 - Helped the client to analyze experimental plant gene data with 2,500 genes.
 - Used R to implement generalized linear models and Bayesian GLM to detect treatment effects.
- KPMG China** Shanghai, China
• *Audit Intern* Jul. 2012 - Feb. 2013
 - Worked on different teams to perform year-end auditing for three public companies.
 - Main tasks focused on aging test, searching for unrecorded liabilities, sending and receiving confirmation letters and organizing data in eAudit System.

Research Experience

- Bandwidth Selection for Kernel Density Estimator of Level Sets**
• *School of Statistics, University of Minnesota, Advisor: Charles Doss* Sep. 2016 - Present
 - Derive an asymptotic expansion of symmetric risk for level set estimation and propose a corresponding algorithm to select optimal bandwidth.
- On the Nonparametric Dynamic Quantile Model**
• *Department of Statistics, Fudan University, Advisor: Shusong Jin* Oct. 2012 - May. 2013
 - Proposed a boosting model for nonparametric dynamic conditional quantile estimation and developed a functional gradient descent algorithm for model estimation.

School Projects

- Leaf Classification**
• *CSCI 5525 Machine Learning* Oct. 2017 - Dec. 2017
 - Used Python with scikit-learn, TensorFlow and Keras to apply popular machine learning classifiers (LDA&QDA, Logistic Regression, SVM, Random Forest and Deep Neural Network) to identify 99 plant species with binary leaf images. Dataset contains 1584 labeled leaf images and 594 unlabeled images.
- R package for Lasso Penalized Logistic Regression**
• *STAT 8054 Advanced Statistical Computing* Mar. 2017 - May. 2017
 - Developed a C based R package implementing cyclical coordinate descent algorithm to solve Lasso Penalized Logistic Regression.
- Paraphrase Identification for Quora Questions**
• *STAT 8933 Statistical Learning and Data Mining* Apr. 2017 - May. 2017
 - Performed data cleaning and feature extraction for 20,000 pairs of Quora questions.
 - Compared the performances on paraphrase identification of different Machine Learning classifiers including Random Forest, SVM, Neural Network, AdaBoosting and penalized Logistic Regression.

Honors

- Summer Research Fellowship**
• *School of Statistics, University of Minnesota* Summer 2017
- SAS Data Mining Champion, Second Place**
• *Organized by SAS Institute Inc, Sponsored by HSBC Bank (China) Co., Ltd.* Nov. 2014
 - Worked with teammates to perform data cleansing, modeling and predicting with HSBC business data.