

Education

- **University of Minnesota, Twin Cities** Minneapolis, MN
Ph.D., Statistics, Research Fields: Nonparametric Statistics Sep. 2015 - 2020 (expected)
- **University of Minnesota, Twin Cities** Minneapolis, MN
M.S., Statistics Sep. 2015 - Jun. 2018
- **Fudan University** Shanghai, China
B.Sc. Statistics Sep. 2010 - Jun. 2015

Skills

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

Work Experience

- **Apple Inc.** Cupertino, CA
Data Scientist Intern, Apple News May. 2019 - present
 - Analyze and predict Apple News user behavior with statistical and machine learning models.
- **Liberty Mutual Insurance** Boston, MA
Data Scientist Intern, Talent Analytics, Research and Modeling Jun. 2018 - Aug. 2018
 - Built a job recommender system using historical job movement data to recommend internal open positions to employees.
 - Conducted text mining to extract features, e.g., education level, skill set, qualifications, from employee resumes and job descriptions as input for recommender system.
 - Implemented classical methods (collaborative filtering and content based recommendation) as baseline methods. Used machine learning models to improve upon classical methods and increased the recommendation accuracy by 10%.
- **School of Statistics, University of Minnesota** Minneapolis, MN
Graduate Teaching Assistant, Graduate Instructor, Statistical Consultant Jan. 2016 - present

Research Experience

- Molstad, A. J., **Weng, G.**, Doss, C. R. and Rothman, A. J. *An explicit mean-covariance parameterization for multivariate response linear regression*. Under revision of *Biometrics*. (<https://arxiv.org/abs/1808.10558>)
 - Developed a new method to fit penalized multivariate response linear regression model that exploits a parametric link between the regression coefficient matrix and the error covariance matrix.
 - Proposed a novel non-convex loss function with an accelerated proximal gradient descent algorithm to estimate the model.
- Doss, C. R. and **Weng, G.** *Bandwidth selection for kernel density estimator of multivariate level sets and highest density regions*. *The Electronic Journal of Statistics*. (<https://arxiv.org/abs/1806.00731>)
 - Derived an asymptotic expansion of the symmetric risk for level set estimation and proposed a corresponding algorithm to select optimal bandwidth.
 - Applied the method to estimate the 90% density highest density region to perform outlier detection for Wisconsin Breast Cancer data.
 - Write an R package `lsbs` (<https://cran.r-project.org/web/packages/lsbs/index.html>) to implement the method.
- Jin, S. and **Weng, G.** *On the Nonparametric Dynamic Quantile Model*.
 - Proposed a boosting model to estimate conditional Value at Risk (VaR) for financial time series data.
 - A functional gradient descent algorithm was developed for model estimation.

Honors

- **Lynn Y.S. Lin Fellowship in Statistical Consulting**
School of Statistics, University of Minnesota *Spring 2019*
- **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.