Houjie Wang

Seattle, Washington 98195 • (979) 402-6690 • whj666@uw.edu • https://github.com/HoujieWang

Education

University of Washington-Seattle **Seattle, WA**
Master of Science in Statistics (GPA: 3.93) *09/2020-03/2022(estimated)*

Texas A&M University **College Station, TX**
Bachelor of Science in Statistics (GPA:3.86) 08/2017 – 08/2020

Selected Coursework: Graduate level: Stochastic Process, Machine Learning, Statistical Inference I&II, Machine Learning, Experimental Design, Real Analysis (Honors), Probability Theory (Honors), Bayesian Statistics, Optimization, Numerical Analysis

Standard Tests:
GRE Score: Verbal 160 (86%), Quantitative 169 (95%), Analytical Writing 4.0 (57%) 11/2019

Programming Skills

R; Python; Latex; C++; MATLAB; Maple

Honors and Scholarships: Bill & Rita Stout Academic Excellence Scholarship, 19 Summer Research Funding ($1500), Dean's Honor (5 semesters)

Academic Research & Projects

Literature Review of "Constructing Priors that Penalize the Complexity of Gaussian Random Fields"
Supervised by Dr. Vincent Roulet 02/2021 – 03/2021

- Detail the proof of Theorem 2.1, showing the exponential conditional prior of Matern parameters.
- Discuss the benefits of using PC prior as a weakly-informative prior with respect to prediction and inference.
- Reproduce the simulations and apply the PC prior to other real-world data.
- Apply PC prior to several covariance functions and discuss possible issues with respect to spherical covariance.

Tropical Data Analysis in R 10/2020 – present
Supervised by Prof. Xiaoxian Tang & Prof. Ruriko Yoshida
GitHub Site: https://github.com/HoujieWang/Rtropical

- Ready to publish a R package to CRAN (Rtropical: data analysis in tropical projective spaces)
- Provide efficient algorithmic solution in the language of R to realize previous work
- Write all function documentations as user guide
- Provide all programming support to formulate a complete package in a user-friendly way
- Participate in paper writing for submission purpose

Tropical Linear Discriminant Analysis and its Applications to Phylogenomics 04/2020 – 09/2020
Supervised by Prof. Xiaoxian Tang & Prof. Ruriko Yoshida

- Defined a novel linear discriminant in tropical semiring and converted into a series of discrete optimization problems
- Addressed feasibility issues of the optimization problem
- Proposed an iterative and converging method to address the serious combinatorial computational issue
- Successfully separated two type of phylogenetic point clouds in projection with better performance than classical LDA
- Achieved visualization of high dimensional point cloud on 2-d graph

Accounting for Difference in Voter Behavior; Results of Analysis from 2012-2016 01/2020 – 05/2020
Supervised by Prof. Huiyan Sang

- Proposed the method of combining spatially weighted regression models and LASSO to address the bias of conventional regression models
- Analyzed county-level election data of "swing states" and made prediction to voting behavior in 2020 presidential election
- Employed a variety of machine learning techniques and obtained results with practical and realistic meanings.

Tropical Support Vector Machine and its Applications to Phylogenomics *(Submitted)* 03/2019 – 04/2020
Preprint link: https://arxiv.org/abs/2003.00677; GitHub site: https://github.com/HoujieWang/Tropical-SVM
Supervised by Prof. Xiaoxian Tang & Prof. Ruriko Yoshida

- Defined a novel support vector machine in tropical geometry space for phylogenetic tree classification
- Completed theorems for SVM hard margin and proved fast and computable $O(n)$ formulas for hyperplane
- Convert combinatorically and computationally hard problem into computable linear programming problems
- Yielded 20% accuracy improvement for highly mixed phylogenetic data compared to classical SVM

R Package Development for Vecchia Approximation 06/2019 – 08/2019
Supervised by Prof. Matthias Katzfuss

- Contributed work of coding to R package, *GPvecchia*.
- Engaged in intensive academic writing with latex for introductory context of fast Gaussian Process computation as part of R vignette (e.g. Vecchia approximation, multi-resolution approximation)
- Developed R scripts as a part of R vignette that realize fast Gaussian-related computation with Vecchia approximation and compare the performance between Vecchia approximation and exact calculation

Data Analysis with Persistent Homology 04/2019 – 05/2019

- Sampled data from multi-dimensional geometric objects
- Developed scripts summarizing topological features based on persistent homology
- Analyzed topological features and noise based on barcode

Simulation of Hierarchical Clustering 03/2019 – 04/2019

- Developed scripts that simulate the functionality of agglomerative hierarchical clustering in various distance measures
- Compared performance across different distance measures and self-developed and existing algorithm

Text Analysis with R 11/2018 – 12/2018

- Developed R scripts displaying the functionality of the package "tm"
- Utilized algorithm-based method to summarize frequent terms appeared in selected paragraphs
- Applied clustering techniques to 10 most frequent terms and examined the interrelationship among them

Activities

Teaching Assistant of Multivariate Calculus 08/2019 – 12/2019

- Assisted Prof. Phillp Yasskin in MATH221 to organize weekly supplementary practice sessions, and gave reviewing lectures before the final exams
- Graded the homework and quizzes and proctored the exams

Math Grader of Linear Algebra 01/2019 – 05/2019

- Graded the homework and quizzes on a weekly basis

Volunteer in Badminton Club 09/2017 – Present

- Interacted with other members and volunteered to teach tricky techniques of badminton

Hobbies

- Badminton, Driving, Movies