# LINGFENG HUO

+1(612) 448-4122 ◇ Minneapolis, MN
huo00030@umn.edu ◇ https://github.com/LingfengHuo9756

## EDUCATION

**UNIVERSITY OF MINNESOTA, Minneapolis, MN** Sep 2021 - May 2023
M.S. in Statistics *GPA*: 4.0
*Honor: Patrick Whitcomb & Patty Napier Fellowship in Statistics*

**CHINA UNIVERSITY OF MINING AND TECHNOLOGY, Xuzhou, China** Sep 2015 - June 2019
B.S. in Applied Mathematics *GPA*: 3.5
*Honor: Second Prize Scholarship*
*Advanced Courses: Real Analysis (A), Functional Analysis (A), Complex Analysis (A), Applied Stochastic Process (A)*

## PUBLICATIONS

*Manuscripts in Preparation*:

- **Huo, L.**, Hou, J., (2022). An efficient computation algorithm for large-scale competing risks data with time-dependent covariates. (Intend to submit to *Journal of Computational and Graphical Statistics*).

## SOFTWARE

- *Demonstration of Efficiency of our New R Package* `cmprskHD` *on Time-dependent or Time-independent Competing Risks Data via Simulations*: URL https://github.com/LingfengHuo9756/Simulation-study-for-cmprskHD.git.
- *Conjugate Newton Method: A New Efficient Optimization Algorithm for BP neural network training*:
  URL https://github.com/LingfengHuo9756/Conjugate-Newton-Method.git.
- *Coding Machine Learning Algorithms in Python without Libraries*:
  URL https://github.com/LingfengHuo9756/Coding-ML-Algorithms-in-Python-Without-Libraries-.git.

## RESEARCH

**An Efficient Computation Algorithm for Large-Scale Competing Risks Data with Time-Dependent Covariates | Fortran, R** June 2022-Present
*Supervisor: Jue Hou, Ph.D.* *Division of Biostatistics, University of Minnesota*

- Did literature review on competing risk regression models, multi-state model, martingales, and computational and variable selection problem for high-dimensional Fine-Gray model.
- Employed Coordinate Descent Algorithm (CDA) for parameter estimation of high-dimensional fine-gray model penalized by elastic net.
- Proposed a new algorithm which accelerates CDA by linearly updating the derivatives of the log partial likelihood and extended it to time-dependent scenario
- Did simulations with parallel computing to show `cmprskHD` is highly efficient to handle **competing risks data with time-dependent covariates** and **10 times** more efficient than the current fastest package `fastcmprsk` for time-independent covariates
- Used `cmprskHD` to study the relationship between the mortality of colorectal cancer and the approvals of different targeted therapies by U.S.FDA based on the EHR data at Mass General Brigham healthcare

**An Efficient Optimization Method for BP Neural Network Training | MATLAB** Dec 2018-June 2019
*Supervisor: Hu Shao, Ph.D.* *School of Mathematics, China University of Mining and Technology*
*Outstanding Undergraduate Thesis*

- Developed an efficient hybrid optimization algorithm with global convergence properties called Conjugate Newton's Method (CNM) by combining Conjugate Direction Method and Newton's Method.

- Adapted CNM to more general loss function by approximating the Hessian in CNM's direction with a positive definite matrix that only requires information from the Jacobian.
- Developed an accelerated golden section search with gradually increasing precision to further increase CNM's efficiency.
- Conducted simulation study to show CNM is twice more efficient than FR conjugate gradient method and steepest gradient method, and can achieve the lowest loss for a four-layer fully-connected network with various activation.

## PROJECTS

**Variable Selection for Detecting Claim Fraud in Auto insurance** | **R, PyTorch** Oct 2021-Jan 2022
*School of Statistics, University of Minnesota*
*Supervisor: Yuhong Yang, Ph.D.* codes: **https://github.com/LingfengHuo9756/Claim-Fraud-Detection.git**

- Cleaned data by imputation and variable recoding, and pretrained data by autoencoder via PyTorch.
- Chose the best variable selection method as LASSO among stepwise selection and penalization methods based on the estimated measures like Variable Selection Deviation, F-measure and G-measure.
- Used Sparsity Oriented Importance learning to double verify the trustworthiness of the 14 variables selected.
- Compared 5 machine learning methods like Random Forest by 4 metrics like AUROC with 10-fold Cross-Validation.
- Chose logistic regression with the selected variables as the final model which won the Kaggle competition held by **Travelers** with the best F-score 0.37

## WORK EXPERIENCE

**Industrial Operation Analyst**
*JD Technology (A Unit of JD.com), Guangzhou, China* June 2020 - June 2021
*Supervisor: Mingxia Li, M.S., Bo Wang, Ph.D.*

- Analyzed the sales data of the JD stores in the industrial parks of South China region to provide advice for the next quarter's budget planning, advertising timeline and product selection

**Financial Analyst**
*Dupont Hongji Films Foshan Co., Ltd, Foshan, China* Sep 2019 - May 2020
*Supervisor: Shiping Feng, B.S.*

- Data analytic for the profitability of each product and the potential of customers; Built models for forecasting the market prospect of the 5 main film products of the company via comprehensive analytics of their corresponding users

## TEACHING EXPERIENCE

**Graduate Teaching Assistant**
*Department of Statistics & Department of Mathematics, University of Minnesota* Summer 2022

- **RUN THE WORLD: Machine Learning Summer Camp**

  Taught traditional machine learning models and help students in **Python** coding

  Instructed students on their final data analytics project and commented on their presentation

## PROGRAMMING SKILLS

- R, Python, PyTorch, Fortran, MATLAB, C++, SPSS, LaTeX