NAN XIAO

me@nanx.me
https://nanx.me
GitHub | LinkedIn

Qualifications

- Machine learning researcher with 5 years' experience and published learning methods for high-dimensional data analysis, data fusion, and translational bioinformatics.
- R developer with 8 years' of R engineering experience. Author and contributor of 20+ open source R packages and Shiny applications. Journal referee for *The R Journal*.
- Full-stack data scientist with experience building and leading data science teams; managing and guiding complex data analysis projects; coordinating data sharing and collaborations across engineering and product teams to serve the executive team.

Work Experience

GENOMIC DATA SCIENTIST, SEVEN BRIDGES GENOMICS; CAMBRIDGE, MA – NOV 2016-PRESENT

Team Lead. Led an elite data science team to:

- Apply intensive quantitative and analytical skills to internal/external data to design a new pricing model that can reduce customers' AWS instance costs significantly.
- Provide direct decision support to the Chief Strategy Officer, Business Development, Product, and Marketing teams to help shape optimized data-driven company strategy.
- Deliver Shiny web applications for internal data visualization, reporting, and consulting.
- Develop the R API client package and related software for accessing, analyzing, and democratizing petabyte-scale genomic data on cloud-based Seven Bridges Platform.

Software

- First author of 10+ open source R packages for machine learning modeling, data visualization, and reproducible research. Total downloads: 30,000+ per month.
- Linear models; feature selection; metric learning: msaenet | enpls | OHPL | RECA
- Sequence feature extraction and feature engineering: protr | Rcpi | grex
- Docker-based reproducible research and benchmarking: liftr | hdnom
- Data visualization: ggsci | Restful API clients: sevenbridges-r
- Contributor to mxnet-r (R binding for Amazon-backed deep learning framework), ashr (empirical Bayes method for large-scale hypothesis testing), and dml (metric learning).

Web Apps

- First author of 6 Shiny apps for delivering the best interactive data analysis experience.
- hdnom.io (selected as RStudio Shiny user showcase app)
- dockflow.org | imgsvd.com | targetnet.org | protr.org | p-values.org

Publications

- 9 journal papers on statistical machine learning methodology, high-dimensional sparse regression, feature selection, and recommender systems. [Google Scholar]
- 4 translated books on predictive modeling, data visualization, and R programming.

Education

Ph.D. Student (Human Genetics). The University of Chicago. Chicago, IL – 2015-2016

Ph.D. Candidate (Statistics). Central South University. Changsha, China – 2012–2018

Bachelor of Science (Statistics). Central South University. Changsha, China – 2008–2012