

Tree-Based Machine Learning Methods

Prediction, Inference, and Variable Selection with the randomForestSRC Ecosystem

Half Day

4 Hours

Virtual Only

COURSE OVERVIEW

This half-day virtual workshop gives a practical, code-centered introduction to tree-based machine learning methods using R. Tree-based methods are useful for nonlinear signals, mixed data types, robustness, and scalable prediction. The workshop focuses on random forest ensembles and centers on the R-package randomForestSRC, which implements random forests for regression, classification, and survival. Topics include out-of-bag inference, test-data prediction, and variable selection using permutation VIMP, minimal depth, and rule-based variable priority with varPro. Advanced topics cover class imbalance, imputation and test time OOD, random hazard forests (RHF), and super greedy trees (SGT). Related packages in the randomForestSRC ecosystem will also be discussed.

WORKSHOP OUTLINE

1 Training

- Regression, classification, survival
- Examples and R code
- Worked interpretations

2 Inference and Prediction

- Out-of-bag inference
- Test-data prediction
- Restore mode and partial plots

3 Variable Selection

- Permutation VIMP
- Minimal depth
- Variable priority with varPro

4 Advanced Topics

- Class imbalance
- Imputation and test time OOD
- Random Hazard Forests
- Super Greedy Trees

SOFTWARE ECOSYSTEM



randomForestSRC

Unified forests for regression, classification, survival, and competing risks.



varPro

Rule-based variable priority for model-independent variable selection.



randomForestRHF (RHF)

Random hazard forests for time-varying covariates in survival analysis.



randomForestSGT (SGT)

Super greedy trees and flexible split geometry.

Designed for applied statisticians and data scientists who want hands-on R workflows for prediction, inference, and variable selection.

Instructors

Workshop leaders for Tree-Based Machine Learning Methods



Hemant Ishwaran

Professor of Public Health Sciences
Graduate Program Director; Director of Statistical Methodology
Division of Biostatistics, University of Miami

Hemant Ishwaran develops machine learning methods for complex biomedical and time-to-event data and turns them into practical open-source tools for investigators. He created Random Survival Forests and the R package randomForestSRC. His work has been applied in cardiovascular disease, heart transplantation, cancer, and genomics.

 <https://ishwaran.org>



Min Lu

Research Associate Professor
Division of Biostatistics, University of Miami

Min Lu works on random forests and trees, causal inference, variable selection, infectious disease modeling, statistical genomics, and meta-analysis. She develops methods and software in the randomForestSRC and varPro ecosystem and collaborates on practical applications in medicine and public health.

 luminwin.net

Workshop software ecosystem



randomForestSRC



varPro



RHF



SGT

randomForestSRC • varPro • randomForestRHF (RHF) • randomForestSGT (SGT)