Regression tree algorithms with applications to missing data imputation, precision medicine and post-selection inference
Professor Wei-Yin Loh (University of Wisconsin, Madison)

Description

This lecture series will provide an introduction to state-of-the-art classification and regression tree and forest algorithms and their applications, with emphasis on the GUIDE method (`https://pages.stat.wisc.edu/~loh/guide.html`). Regression tree methods fit piecewise-constant or piecewise-linear models to data by recursive partitioning. The models are nonparametric and are unconstrained by high dimensionality, multi-collinearity, and missing values in predictor variables. They are relatively fast to build, their predictions can be represented as decision trees, and they yield asymptotically consistent regression function estimates. The lectures will present a mix of theory, methodology, and applications, including examples from economics, medicine, public health, and automotive engineering. It will be of interest to researchers and students wishing to learn more about this area of statistical machine learning.

Lecture 1 (May 11) Basic ideas of classification and regression trees. *Example on predicting mortality from Covid-19.*

Lecture 2 (May 13) Models for sample survey data with multiple missing value codes; pitfalls of missing-value imputation. *Example with consumer expenditure data.*

Lecture 3 (May 16) Proportional hazards tree models for randomized trials with censored response variables; identification of subgroups with differential treatment effects for precision medicine. *Example with breast cancer data.*

Lecture 4 (May 18) Models for randomized trials with multiple and longitudinal response variables. *Precision medicine application from a diabetes trial.*

Lecture 5 (May 23) Prediction models for data with cyclic or angular predictor variables. *Example with data from experiments on vehicle crash tests.*

Lecture 6 (May 25) Models for data from non-randomized studies. *Example on estimation of causal effects of right heart catheterization from observational data.*

Lecture 7 (May 30) Post-selection inference: bootstrap calibrated simultaneous confidence intervals for parameters in tree node models.

Lecture 8 (Jun 1) Wrap-up and demonstration of free GUIDE software.