Introduction to Statistical Machine Learning

Dr Rafael de Andrade Moral
Associate Professor of Statistics, Maynooth University

rafael.deandrademoral@mu.ie
https://rafamoral.github.io

Outline

- Extensions to tree-based methods
- Bayesian additive regression trees (BART)
- Generalized additive models for location, scale and shape (GAMLSS)

BART

BART

- Bayesian Additive Regression Trees (BART; Chipman et al., 2010) is a Bayesian model based on a sum of trees
- Very good prediction properties
- Deals with interactions and non-linear relationships
- Based on a probability distribution so ‘easily’ extendable
- A Bayesian model, so have probabilistic uncertainty intervals for any required quantities
- A focus of research, with many extended versions already proposed in the literature

BART

$$y = \sum_{j=1}^{M} g(X, T_j, \Theta_j, \mu_j) + \varepsilon; \quad \varepsilon \sim \mathsf{N}(0, \sigma^2)$$

- y is the response and X are the covariates
- T is the tree structure (parameter)
- Θ is a set of split variables and values (parameters)
- μ is the set of terminal node values (parameters)
- The algorithm works by guessing initial values of all these (usually a stump), then proposing and accepting/rejecting new trees
- The big problem that occurs here is that the number of parameters changes when the tree structure changes
- The main reason that BART works is that we can collapse over the tree structure when we use a normally distributed prior on the μ terminal node parameters

BART

- Standard BART MCMC uses a back-fitting approach where each tree is fixed and the others updated in turn
- Requires us to propose and accept/reject a new tree at each iteration for M trees
- We put extra prior distributions on the size and shape of the trees to keep them small, and to keep the μ values from dominating the predictions
- Default BART has four moves to generate new trees, all of which are reversible:
  - *Grow*: picks a random terminal node and splits it in two by choosing a random split variable and split value
  - *Prune*: picks a pair of adjacent terminal nodes and merges them
  - *Swap*: picks two random internal nodes and swaps their split variables and values
  - *Change*: picks a random internal node and changes the split variable and value

GAMLSS

GAMLSS

- Generalized Additive Models for Location, Scale, and Shape (GAMLSS; Rigby and Stasinopoulos, 2005) are a very flexible semi-parametric modelling framework

$$Y \sim f(\mu, \sigma, \nu, \tau)$$

- Includes many distributions with up to 4 parameters
- Allows for distributional regression, i.e. all parameters can be modelled with covariates
- Spline and loess smoothing, as well as random effects allowed
- Unified framework for model diagnostics

GAMLSS

- The `gamlss` package is the main implementation
- Extra features available through companion packages
- The package `gamlss.add` includes extra features, such as the inclusion of regression trees in the linear predictor for any parameter of interest