Chipman, Hugh A.; George, Edward I.; McCulloch, Robert E.

BART: Bayesian additive regression trees. (English) [Zbl 1189.62066]

Summary: We develop a Bayesian “sum-of-trees” model where each tree is constrained by a regularization prior to be a weak learner, and fitting and inference are accomplished via an iterative Bayesian backfitting MCMC algorithm that generates samples from a posterior. Effectively, BART is a nonparametric Bayesian regression approach which uses dimensionally adaptive random basis elements. Motivated by ensemble methods in general, and boosting algorithms in particular, BART is defined by a statistical model: a prior and a likelihood. This approach enables full posterior inference including point and interval estimates of the unknown regression function as well as the marginal effects of potential predictors. By keeping track of predictor inclusion frequencies, BART can also be used for model-free variable selection. BART’s many features are illustrated with a bake-off against competing methods on 42 different data sets, with a simulation experiment and on a drug discovery classification problem.

MSC:

- 62G08 Nonparametric regression and quantile regression
- 62F15 Bayesian inference
- 65C60 Computational problems in statistics (MSC2010)
- 62H30 Classification and discrimination; cluster analysis (statistical aspects)

Keywords:
Bayesian backfitting; boosting; CART; classification; ensemble; MCMC; nonparametric regression; probit model; random basis; regularizatio; sum-of-trees model; variable selection; weak learner

Software:
AdaBoost.MH; MASS (R); e1071; gbm; R; BayesTree; LIBSVM; ROCR

Full Text: DOI

References:


Cited in 2 Reviews
Cited in 82 Documents


This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.