

BART: A Comparison with Non-bayesian Tree Methods

Vincenzo Dorrello Giulio Frey Guido Rossetti Giovanni Scarpato

November 24, 2024

Abstract

Your abstract here.

1 Introduction

Bayesian Regression Additive Trees (BART) were first introduced by Chipman, George, and McCulloch 2010

2 Non-Bayesian Tree Based Methods

Tree based method are more usefull compared to the linear regression model when the relation to be model is complex or non-linear. Indeed, the standard linear model is defined as:

$$f(X) = \beta_0 + \sum_{j=1}^p X_j \beta_j \quad (1)$$

Whereas the decision tree model as:

$$f(X) = \sum_{m=1}^M c_m \cdot 1(X \in R_m) \quad (2)$$

DESCRIBE BETTER THIS EQUATIONS

We take most of our arguments for this section from the James et al. 2021, Chapter 8.

2.1 Regression Trees

Tree based methods that we cover in this paper are all based on decision trees. Decision trees can be applied both to regression or classification problems, and are thus named accordingly to their application. Regression trees in particular are build in two steps.

First, the prediction space (the set of all possible values of the vector of observables) is split into J regions R_j that are distinct and non-overlapping. Secondly, a prediction for the outcome variable is made for each observation in each R_j regions, as the mean of the training observations in that same regions.

For a predictor space of P elements the regions are p dimensional boxes, that are build by minimizing the Residual Sum Squared, as the equation 3 shows.

$$R_1, \dots, R_j = \underset{R_j}{\operatorname{argmin}} \sum_{j=1}^J \sum_{i \in R_j} (y_i - \hat{y}_{r_j})^2 \quad (3)$$

Instead of considering all the possible partitions of the predictor space, a top-down approach is used. First, all observations are considered such that they belong to the same region. Then, this process chooses the best split minimizing the RSS irrespective of future outcomes. This is then repeated for a number of times, until a stopping criterion is reached (usually a minimum number of observations in all the regions).

As the regions visualizations are complex when more than 2 regressors are present, a tree based graph is used.

The process described above will likely tend to overfit the data. There are numerous solutions to this, the most simple one being increasing the number of AND GENERALLY THE ENSAMBLING MODELS observations that rise the stopping rule. A valid alternative is to grow a very large tree, and then reduce it to obtain a smaller subtree. This is known as tree pruning.

ADD COST COMPLEXITY PRUNING

One of the main disadvantages of this setting, is that trees are non-robust. In fact a small change in the data can lead to a large change in the estimation. This follows from the fact that Regression Trees suffer from high variance. This is indeed quite problematic and the main reason that simple regression trees are not used by practitioners. Ensemble methods aim at solving this issue. Those methods in the context of trees combine simpler models (called weak learners), whichy may lead to mediocre predictions, in a single more powerful and stable model. QUelli di cui sotto sopra abuiao detto sono tutti ensable models.

2.2 Bagging

Bagging is a generic method that aims at reducing the variance of a statistical learning method. Bagging is based on the fact that if we have independent observations (Z_1, \dots, Z_n) with same variance σ , then the vairance of the sample mean of the observation will be lower, as it is $V(\bar{Z}) = \frac{\sigma}{n}$.

Bagging uses a bootstrap sampling procedure to obtain different samples from the data, and then fits a tree for each of the bootstrapped samples. The single estimates for each sample are then averaged, as equation 4 shows.

$$\hat{f}_{\text{bag}}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}^{*b}(x). \quad (4)$$

References

- Chipman, Hugh A., Edward I. George, and Robert E. McCulloch (Mar. 1, 2010). “BART: Bayesian Additive Regression Trees”. In: *The Annals of Applied Statistics* 4.1. ISSN: 1932-6157. DOI: 10 . 1214 / 09 - A0AS285. arXiv: 0806 . 3286 [stat]. URL: <http://arxiv.org/abs/0806.3286> (visited on 11/23/2024).
- James, Gareth et al. (2021). *An Introduction to Statistical Learning: With Applications in R*. Second edition. Springer Texts in Statistics. New York: Springer. 1 p. ISBN: 978-1-07-161418-1.