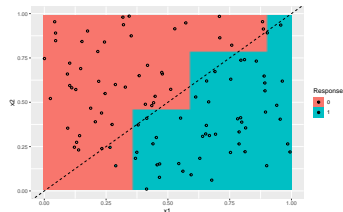


Introduction to Machine Learning

CART

Advantages & Disadvantages

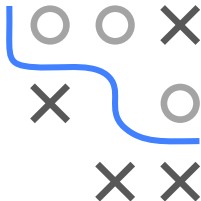


Learning goals

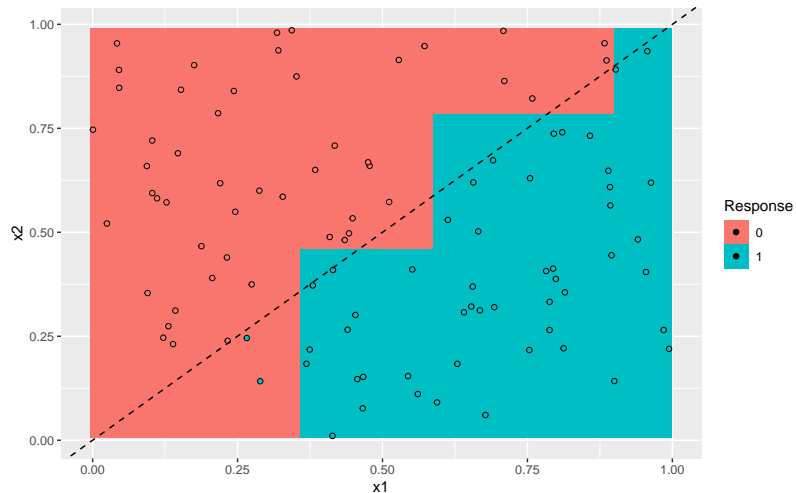
- Understand the advantages and disadvantages of CART
- Know when and where CART are applied

ADVANTAGES

- Fairly easy to understand, interpret and visualize.
- Not much preprocessing required:
 - Automatic handling of non-numerical features
 - Automatic handling of missing values via surrogate splits
 - No problems with outliers in features
 - Monotone transformations do not affect the model fit: scaling becomes irrelevant
- Interaction effects between features are easily possible
- Can model discontinuities and non-linearities
- Performs automatic feature selection
- Relatively fast, scales well with larger data
- Flexibility through the definition of custom split criteria or leaf-node prediction rules: clustering trees, semi-supervised trees, density estimation, etc.

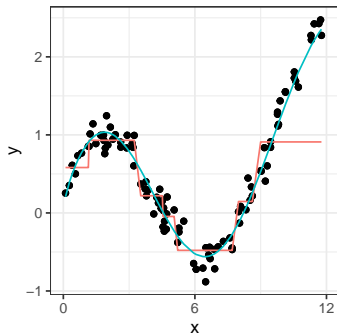
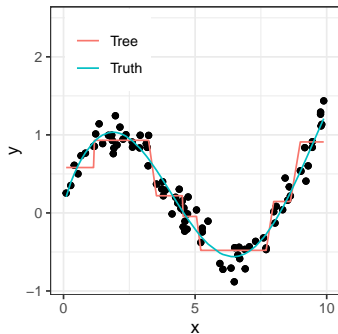


DISADVANTAGE: LINEAR DEPENDENCIES



Linear dependencies must be modeled over several splits. Logistic regression would model this easily with fewer parameters.

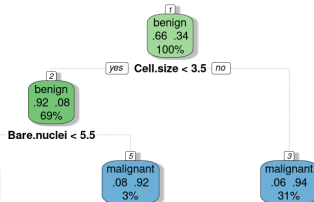
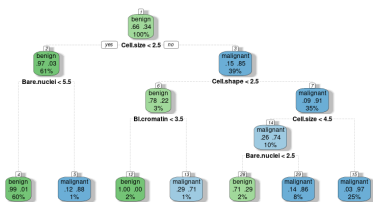
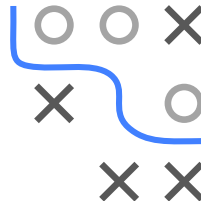
DISADVANTAGES: SMOOTH FUNCTIONS AND EXTRAPOLATION



Prediction functions of trees are never smooth as they are always step functions and do not extrapolate well beyond the training observations.

DISADVANTAGE: INSTABILITY

The resulting decision trees look very different:



CART IN PRACTICE

- Compared to other learners CART has suboptimal predictive performance, mainly because of the problems previously shown.
- However, most disadvantages can be overcome when trained in ensembles: bagging or random forests.
- Furthermore, trees are attractive tools if an interpretable model is desired or legally required.



FURTHER TREE METHODOLOGIES

This lecture is mainly focused on Classification and Regression Trees (CART) ▶ Breiman et al. 1984

However, there are noteworthy other tree-based approaches as well:

- Automatic Interaction Detection (AID) ▶ Sonquist and Morgan 1964 and Chi-squared Automatic Interaction Detection (CHAID) ▶ Kass 1980 :
Creates all possible cross-tabulations for each categorical predictor until the best outcome is achieved and no further splitting can be performed
- C4.5 ▶ Quinlan 1993 : Not limited to binary splitting
- Unbiased Recursive Partitioning ▶ Hothorn, Hornik, and Zeileis 2006 :
Improves the variable selection algorithm

