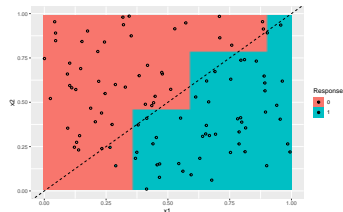


Introduction to Machine Learning

CART

Advantages & Disadvantages

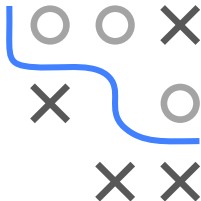


Learning goals

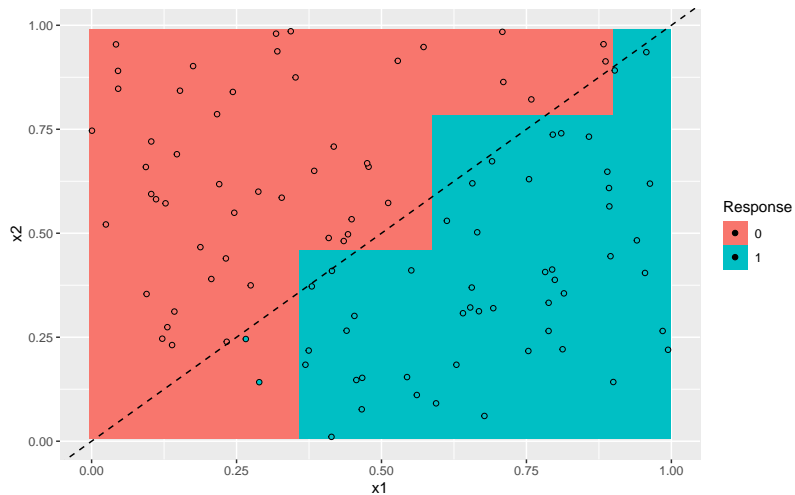
- Understand the advantages and disadvantages of CART
- Know when and where CART are applied

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.

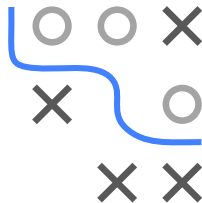
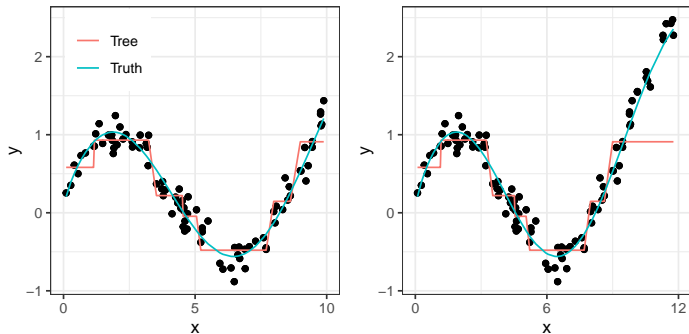


DISADVANTAGE: LINEAR DEPENDENCIES



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.

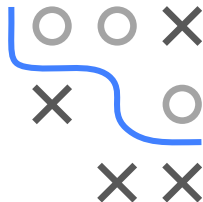
DISADVANTAGE: INSTABILITY

- High instability (variance) of the trees
- Small changes in the data may lead to very different splits/trees
- This leads to a) less trust in interpretability b) is a reason why the prediction error of trees is usually comparably high

Consider the Wisconsin Breast Cancer data set with 699 observations on 9 features and binary target (“benign” vs. “malignant”). We fit two trees: (A) with the full data set and (B) where we eliminated a single observation. Results in label flip for 17 observations of the training data:

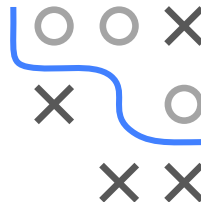
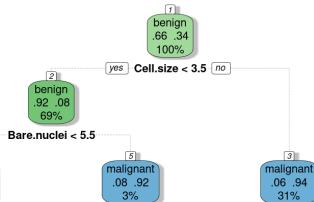
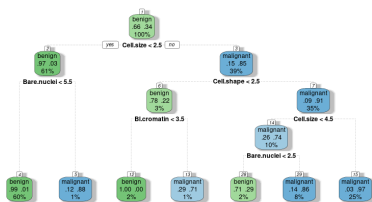
	benign	malignant
benign	445	6
malignant	11	236

Rows: Predictions of (A), columns: Predictions of (B)



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, 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 et al., 2006: Improves the variable selection algorithm

