TreeNet Tree Ensembles and CART Decision Trees: A Winning Combination

October 2012

Mikhail Golovnya
Salford Systems

Course Outline

- CART decision tree pros/cons
- TreeNet stochastic gradient boosting: a promising way to overcome the shortcomings of a single tree
- Introducing TreeNet, a powerful modern ensemble of boosted trees
  - Methodology
  - Reporting
  - Interpretability
  - Post-processing
  - Interaction detection
- Advantages of using both CART and TreeNet
  - Contribution from CART
  - Contribution from TreeNet

Demonstration Dataset

108,376 bank customers (commercial and individual) with 6,564 in bad standing over the past two years

<u>Goal:</u> identify customers in bad standing using the following predictors

Revolving utilization of credit

Age of the primary account holder

Debt ratio of the primary account holder

Monthly income

Number of open credit lines

Number of mortgages

Number of dependents

CART Advantages

1. Relatively fast
2. All types of variables
    1. Numeric, binary, categorical, missing values
3. Invariant under monotone transformations
    1. Variable scales are irrelevant
    2. Immunity to outliers
    3. Most variables can be used "as is"
4. Resistance to many irrelevant variables
5. Few tunable parameters "off-the-shelf" procedure
6. Interpretable model representation

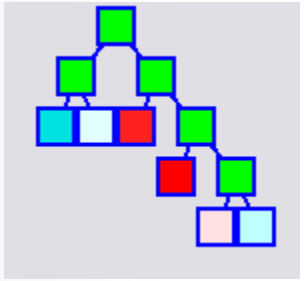
CART Disadvantages

1. Trade-off: accuracy vs. interpretability
2. Piecewise-constant model
    1. Big errors near region boundaries
    2. Impossible to detect fine differences within the segment
3. Instability => high variance
    1. Small data change => big model change (especially for large trees)
4. Data fragmentation – splitting
5. High interaction order model, unreasonably complicated way to represent simple additive dependencies

TreeNet Tree Ensembles

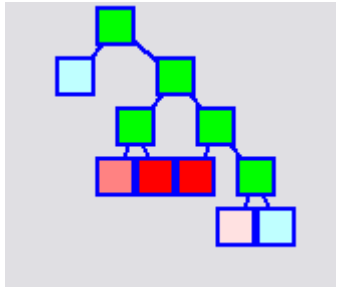
- Complements CART advantages, while dramatically increasing accuracy

Tree 1



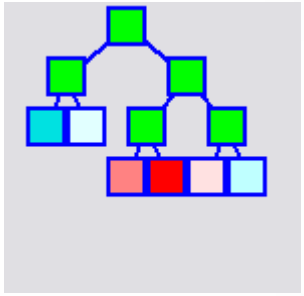
+

Tree 2



+

Tree 3



First tree grown on original target. Intentionally "weak" model

2nd tree grown on residuals from first. Predictions made to improve first tree

3rd tree grown on residuals from model consisting of first two trees

TreeNet Overcomes CART's Shortcomings

Piecewise-Constant Model	CART Big errors near region boundaries, coarse predictions	TreeNet Fine predictions, nearly emulating smooth continuous response surface
Instability and Variance	CART Small data changes induce big model changes (especially for large trees)	TreeNet Stable models due to averaging of individual tree responses
Data Fragmentation	CART Relatively few predictors make it into the model	TreeNet Each tree works with the entire data – many opportunities for variables to enter
High Interaction Order Model	CART Always enforced	TreeNet Allows precise control over the interactions

TreeNet and CART
A Winning Combination