

# Rule Ensembles in R: The “RuleFit Batch” Library

**Giovanni Seni, Ph.D.**

Intuit, Applied Data Sciences Group  
Senior Scientist

Santa Clara University  
Adjunct Faculty, Computer Engineering

**Bay Area useR Group – December 12, 2011**

# Overview

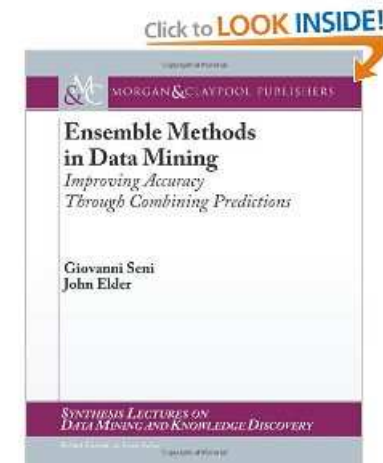
---

- Ensemble Methods in a Nutshell
- Rule Ensembles
  - Overview
  - Available Libraries
- Demo

# Ensemble Methods In a Nutshell

---

- Model:  $F(\mathbf{x}) = a_0 + \sum_{m=1}^M a_m f_m(\mathbf{x})$
- $\{f_m(\mathbf{x})\}_1^M$  : “basis” functions (or “base learners”)
  - Derived predictors capture non-linearities and interactions
- Various fitting procedures:
  - Bagging, Random Forest, Boosting, etc.
- Modern 2-stage process:
  - I. generate basis functions
  - II. Post fit to the data via regularized regression

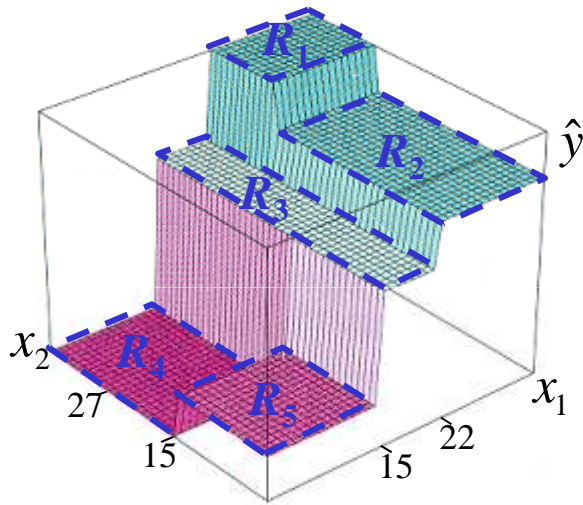


# Rule Ensembles

## Overview

---

- Trees as collection of conjunctive rules:  $T_m(\mathbf{x}) = \sum_{j=1}^J \hat{c}_{jm} I(\mathbf{x} \in \hat{R}_{jm})$



$$R_1 \Rightarrow r_1(\mathbf{x}) = I(x_1 > 22) \cdot I(x_2 > 27)$$

$$R_2 \Rightarrow r_2(\mathbf{x}) = I(x_1 > 22) \cdot I(0 \leq x_2 \leq 27)$$

$$R_3 \Rightarrow r_3(\mathbf{x}) = I(15 < x_1 \leq 22) \cdot I(0 \leq x_2)$$

$$R_4 \Rightarrow r_4(\mathbf{x}) = I(0 \leq x_1 \leq 15) \cdot I(x_2 > 15)$$

$$R_5 \Rightarrow r_5(\mathbf{x}) = I(0 \leq x_1 \leq 15) \cdot I(0 \leq x_2 \leq 15)$$

- These simple rules,  $r_m(\mathbf{x}) \in \{0,1\}$ , can be used as base learners
- Main motivation is *interpretability*

# Rule Ensembles

## Overview (2)

---

- Rule-based model:  $F(\mathbf{x}) = a_0 + \sum_m a_m r_m(\mathbf{x})$ 
  - Still a piecewise constant model
    - Linear targets can still be problematic...
  - Complement the non-linear rules with purely linear terms:

$$F(\mathbf{x}) = a_0 + \sum_m a_m r_m(\mathbf{x}) + \sum_j b_j x_j$$

- Rule generation:
    - Take advantage of a decision tree ensemble
    - E.g., one rule for each (terminal) node in each tree  $T_m(\mathbf{x})$
-

# Rule Ensembles

## Overview (3)

---

- Rule fitting
  - Linear regularized procedure

$$\begin{aligned} (\{\hat{a}_k\}, \{\hat{b}_j\}) = \arg \min_{\{a_k\}, \{b_j\}} & \sum_{i=1}^N L\left(y_i, a_0 + \sum_{k=1}^K a_k r_k(\mathbf{x}_i) + \sum_{j=1}^p b_j x_{ij}\right) \\ & + \lambda \left( \sum_{k=1}^K |a_k| + \sum_{j=1}^p |b_j| \right) \end{aligned}$$

- $K$ : total number of rules
  - $p \leq n$  total number of linear terms
- Tree size controls rule “complexity”

# Rule Ensembles

## Available Libraries

---

- RuleFit
  - [www-stat.stanford.edu/~jhf/R-RuleFit.html](http://www-stat.stanford.edu/~jhf/R-RuleFit.html)
  - Fortran engine with R front-end
- TMVA - Toolkit for Multivariate Data Analysis
  - [root.cern.ch](http://root.cern.ch)
  - C++ engine
- “RuleFit Batch”
  - Collection of R-based utilities on top of Friedman’s engine