EXPLORING & MODELING USING INTERACTIVE DECISION TREES IN SAS® ENTERPRISE MINER™
ANALYTICS LIFECYCLE

Formulate Problem
Data Preparation
Data Exploration
Develop Models
Transform & Select
Validate Models
Deploy Model
Evaluate & Monitor Model

Develop Model
Evaluate & Monitor Model
Formulate Problem
Data Preparation
Data Exploration
Develop Models
Transform & Select
Validate Models
Deploy Model

SAS: THE POWER TO KNOW

Copyright © 2013, SAS Institute Inc. All rights reserved.
ANALYTICS LIFECYCLE

DECISION TREES CAN HELP IN VARIOUS STAGES

- Formulate Problem
- Data Preparation
- Data Exploration
- Transform & Select
- Develop Models
- Validate Models
- Deploy Model
- Evaluate & Monitor Model
- Develop Models
- Validate Models
- Evaluate & Monitor Model
- Deploy Model
WHY DECISION TREES?
DECISION TREES  ADVANTAGES

- Decision Trees are powerful predictive and explanatory modeling tools.
- They are flexible in that they are able to model targets that are:
  - Interval (regression trees)
  - Ordinal, nominal and binary (classification trees)
- Trees can accommodate nonlinearities and interactions.
- Trees are simple to understand and present.
Node = 10

if Saving Balance >= 2615.09
AND Credit Card Balance < 641.915
then
  Tree Node Identifier = 10
  Number of Observations = 981
  Predicted: INS=1 = 0.68
  Predicted: INS=0 = 0.32
DECISION TREE BACKGROUND
Decision trees are statistical models designed for supervised prediction problems. The tree is fitted to data by recursive partitioning. Partitioning refers to segmenting the data into subgroups that are as homogeneous as possible with respect to the target. Many algorithms – CHAID, CART, C4.5, C5.0
2 TYPES OF TREES

Classification tree – target is categorical
Regression tree – target is continuous
DECISION TREES

MULTI-WAY SPLITS

X10

<1

7 (96%)

1-41

X1

1 (79%)

42-51

X1

≥52

≥66

<1

X10

1 (94%)

1-65

9 (75%)

≥27

<1

1-26

7 (61%)

≥33

≥33

7 (66%)

<22

7 (84%)

23-32

1 (70%)

≥33

7 (66%)
DECISION TREES
PARTITIONED INPUT SPACE

NOX

.9

.8

.7

.6

.5

.4

.3

14

27

19

22

27

33

16

46

3 4 5 6 7 8 9

RM
DECISION TREES
MULTIVARIATE STEP FUNCTION
| Leaf | $\Pr(1|x)$ | $\Pr(7|x)$ | $\Pr(9|x)$ | Decision |
|------|------------|------------|------------|----------|
| 1    | 0.03       | 0.96       | 0.01       | 7        |
| 2    | 0.09       | 0.91       | 0.00       | 7        |
| 3    | 0.56       | 0.44       | 0.00       | 1        |
| 4    | 0.95       | 0.05       | 0.00       | 1        |
| 5    | 0.80       | 0.10       | 0.10       | 1        |
| 6    | 0.64       | 0.09       | 0.27       | 1        |
| 7    | 0.00       | 0.13       | 0.87       | 9        |
| 8    | 0.10       | 0.73       | 0.17       | 7        |
| 9    | 0.78       | 0.01       | 0.21       | 1        |
| 10   | 0.01       | 0.00       | 0.99       | 9        |
USING DECISION TREES FOR INITIAL AND EXPLORATORY DATA ANALYSIS
DETECTION TREES

INITIAL DATA ANALYSIS AND EXPLORATORY DATA ANALYSIS

- Interpretability
- No strict assumptions concerning the functional form of the model
- Resistant to the curse of dimensionality
- Robust to outliers in the input space
- No need to create dummy variables for nominal inputs
- Missing values do not need to be imputed
- Computationally fast (usually)
• Dimension Reduction
  • Input subset selection
  • Collapsing levels of nominal inputs
• Dimension Enhancement
  • Discretizing interval inputs
  • Stratified modeling
DETECTION TREES

INPUT SELECTION
DECISION TREES  COLAPSING LEVELS

one input

one level deep

multi-way split

one input
DECISION TREES INTERACTIVE TRAINING

- Force and remove inputs
- Define split values
- Manually prune branches and leaves
**TIP:** Prior to invoking interactive mode, modify the Decision Tree properties to reflect the type of tree you wish to build.
BUILDING SEGMENTATION TREES
Interactively build trees while considering more than one target.
Use Start Groups & End Groups Nodes
GRADIENT BOOSTING

- Sequential ensemble of many trees
- Extremely good predictions
- Very effective at variable selection
Random Forest

- Predictive Model called a Forest
- Creates Several Trees
- Training Data sampled without replacement
- Input variables sampled

Available in EM 13.1
The Interactive Decision Tree may not use all of your data. It uses a sample of at most 20,000 observations to prevent the excessive time and memory consumption that can occur with large data sets. You can control the size and method for creating the sample with Project Start Code.
%let EM_INTERACTIVE_TREE_MAXOBS = <max-number-of-observations-in-sample>;

%let EM_INTERACTIVE_TREE_SAMPLEMETHOD = <RANDOM | FIRSTN | STRATIFY>;

TIP  INTERACTIVE DECISION TREE
%let EM_INTERACTIVE_TREE_MAXOBS = 100000;

%let EM_INTERACTIVE_TREE_SAMPLEMETHOD = RANDOM;
LEARNING MORE

• SAS® Enterprise Miner™ In-product Help File
• Documentation:
  • http://support.sas.com/documentation/onlinedoc/miner/index.html
• Getting Started with SAS® Enterprise Miner™
  • Documentation PDF
  • Sample Data ZIP
  • Recorded Webinar: http://www.sas.com/reg/web/corp/1864003
Decision Tree Modeling

https://support.sas.com/edu/schedules.html?ctry=us&id=1463

Data Mining Techniques: Theory and Practice

https://support.sas.com/edu/schedules.html?ctry=us&id=1244
<table>
<thead>
<tr>
<th>ISBN</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>SAS® Press</strong></td>
</tr>
<tr>
<td>978-0-470-65093-6</td>
<td><em>Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, Third Edition</em></td>
</tr>
<tr>
<td>978-1-59047-567-6</td>
<td><em>Decision Trees for Business Intelligence and Data Mining: Using SAS® Enterprise Miner™</em></td>
</tr>
<tr>
<td>978-0-12-374765-5</td>
<td><em>Handbook of Statistical Analysis and Data Mining Applications</em></td>
</tr>
<tr>
<td>978-1-59047-703-8</td>
<td><em>Predictive Modeling with SAS® Enterprise Miner™: Practical Solutions for Business Applications</em></td>
</tr>
</tbody>
</table>
Decision Trees for Analytics Using SAS Enterprise Miner

By: Barry de Ville and Padraic Neville
Copyright Date: July 2013
SAS Bookstore:
https://support.sas.com/pubscat/bookdetails.jsp?catid=1&pc=63319
Table of Contents [PDF]
Free Chapter [PDF]
Example Code and Data
THANK YOU FOR USING SAS

CUSTOMER LOYALTY TEAM • Support You Can Count On