



5<sup>th</sup> March 2004

- Q1. What is CART?
- Q2. What makes Salford Systems' CART the only "true" CART?
- Q3. What is a decision tree?
- Q4. What makes CART so easy to interpret?
- Q5. How are decision trees grown?
- Q6. Why is CART unique among decision-tree tools?
- Q7. What tree-growing, or "splitting", criteria can CART provide?
- Q8. What are "adjustable misclassification penalties"?
- Q9. What are "intelligent surrogates for missing values"?
- Q10. What are CART's "automatic self-validation procedures"?
- Q11. What is a "multiple-tree, committee-of-expert method", or "bootstrap aggregation"?
- Q12. When can CART be used to advantage as a standalone package?
- Q13. How can CART complement other data-mining packages and/or suites?
- Q14. How quickly can CART generate results?

### **Q1. What is CART?**

A1. CART is an acronym for Classification and Regression Trees, a decision-tree procedure introduced in 1984 by world-renowned UC Berkeley and Stanford statisticians, Leo Breiman, Jerome Friedman, Richard Olshen, and Charles Stone. Their landmark work created the modern field of sophisticated, mathematically- and theoretically-founded decision trees. The CART methodology solves a number of performance, accuracy, and operational problems that still plague many current decision-tree methods. CART's innovations include:

- solving the "how big to grow the tree" problem;
- using strictly two-way (binary) splitting;
- incorporating automatic testing and tree validation, and;
- providing a completely new method for handling missing values.

### **Q2. What makes Salford Systems' CART the only "true" CART?**

A2. Salford Systems' CART is the only decision tree based on the original code of Breiman, Friedman, Olshen, and Stone. Since the code is proprietary, CART is the only true implementation of this classification-and-regression-tree methodology. In addition, the procedure has been substantially enhanced with new features and capabilities in exclusive collaboration with CART's creators. While some other decision-tree products claim to implement selected features of this technology, they are unable to reproduce genuine CART trees and lack key performance and accuracy components. Further, CART's creators continue to collaborate with Salford Systems to refine CART and to develop the next generation of data-mining tools.

### **Q3. What is a decision tree?**

A3. A decision tree is a flow chart or diagram representing a classification system or predictive model. The tree is structured as a sequence of simple questions, and the answers to these questions trace a path down the tree. The end point reached determines the classification or prediction made by the model, which can be a qualitative judgment (e.g., these are responders) or a numerical forecast (e.g., sales will increase 15 percent).

**Q4. What makes CART so easy to interpret?**

A4. As illustrated above, the results of a decision-tree data-mining project are displayed as a tree-shaped visual diagram. Discovered relationships and patterns in the data - even in massively complex datasets with hundreds of variables - are presented as a flow chart. Compare this to complex parameter coefficients in a logistic regression output or a stream of numbers in a neural-net output, and the appeal of decision trees is readily apparent.

The visual display enables users to see the hierarchical interaction of the variables; and it often confirms previous knowledge about important data relationships, which adds confidence in the reliability and utility of the CART model. Further, because simple if-then rules can be read right off the tree, models are easy to grasp and easy to apply to new data.

**Q5. How are decision trees grown?**

A5. There are a number of different ways to grow decision trees, but CART uses strictly binary, or two-way, splits that divide each parent node into exactly two child nodes by posing questions with yes/no answers at each decision node. CART searches for questions that split nodes into relatively homogenous child nodes, such as a group consisting largely of responders, or high credit risks, or people who bought sport-utility vehicles. As the tree evolves, the nodes become increasingly more homogenous, identifying important segments. Other methods, such as CHAID, favor multi-way splits that can paint visually appealing trees but that can bog models down with less accurate splits.

**Q6. Why is CART unique among decision-tree tools?**

A6. CART is based on a decade of research assuring stable performance and reliable results. CART's proven methodology is characterized by:

Reliable pruning strategy - CART's developers determined definitively that no stopping rule could be relied on to discover the optimal tree, so they introduced the notion of overgrowing trees and then pruning back; this idea, fundamental to CART, ensures that important structure is not overlooked by stopping too soon. Other decision-tree techniques use problematic stopping rules.

Powerful binary-split search approach - CART's binary decision trees are more sparing with data and detect more structure before too little data is left for learning. Other decision-tree approaches use multi-way splits that fragment the data rapidly, making it difficult to detect rules that require broad ranges of data to discover.

Automatic self-validation procedures - in the search for patterns in databases it is essential to avoid the trap of "overfitting," or finding patterns that apply only to the training data. CART's embedded test disciplines ensure that the patterns found will hold up when applied to new data. Further, the testing and selection of the optimal tree are

an integral part of the CART algorithm. Testing in other decision-tree techniques is conducted after-the-fact and tree selection is left up to the user.

In addition, CART accommodates many different types of real-world modeling problems by providing a unique combination of automated solutions:

- surrogate splitters intelligently handle missing values;
- adjustable misclassification penalties help avoid the most costly errors;
- multiple-tree, committee-of-expert methods increase the precision of results, and;
- alternative splitting criteria make progress when other criteria fail.

**Q7. What tree-growing, or "splitting", criteria can CART provide?**

A7. CART includes seven single-variable splitting criteria - Gini, Symgini, twoing, ordered twoing and class probability for classification trees, and least squares and least absolute deviation for regression trees - and one multi-variable splitting criteria, the linear combinations method. The default Gini method typically performs best, but, given specific circumstances, other methods can generate more accurate models. CART's unique twoing procedure, for example, is tuned for classification problems with many classes, such as modeling which of 170 products would be chosen by a given consumer.

Other splitting criteria are available for inherently difficult problems in which even the best models are expected to have a relatively low accuracy. Demographics, for example are often weak predictors of attitude- and preference-based segments. Special CART tree-growing options can dramatically increase the predictive accuracy of such demographic-based models. Additional unique tree-growing criteria are available for problems involving unequal misclassification costs, ordered target variables, and continuous dependent variables. To deal more effectively with select data patterns, CART also offers splits on linear combination of continuous predictor variables. For this option, CART looks for weighted averages of predictor variables to use as splitters; these weighted averages can reveal important database structure and can uncover new critical measures.

**Q8. What are "adjustable misclassification penalties"?**

A8. Unlike many data-mining tools, CART can accommodate situations in which some misclassifications, or cases that have been incorrectly classified, are more serious than others. CART users can specify a higher penalty for misclassifying certain data, and the software will steer the tree away from that type of error. Further, when CART cannot guarantee a correct classification, it will try to ensure that the error it does make is less costly. If credit risk is classified as low, moderate, or high, for example, it would be much more costly to classify a high-risk person as low-risk than as moderate-risk. Traditional data-mining tools cannot distinguish between these errors.

**Q9. What are "intelligent surrogates for missing values"?**

A9. CART handles missing values in the database by substituting surrogate splitters, which are back-up rules that closely mimic the action of primary splitting rules. Suppose that, in a given model, CART splits data according to household income. If a value for income is not available, CART might substitute education level as a good surrogate.

The surrogate splitter contains information that is typically similar to what would be found in the primary splitter. Other products? approaches treat all records with missing values as if the records all had the same unknown value; with that approach all such ?missings? are assigned to the same bin. In CART, each record is processed using data specific to that record; this allows records with different data patterns to be handled differently, which results in a better characterization of the data.

By using surrogates to stand in for missing values, CART generates robust and reliable predictive models, even when applied to very large databases with hundreds of variables and many missing values. CART?s identification of surrogate predictor variables also provides an effective way to discover low-cost predictive mechanisms. If the best splitting criterion in a tree involves an expensive or difficult-to-obtain measure, a less-expensive surrogate can be considered instead.

**Q10. What are CART?s "automatic self-validation procedures"?**

A10. CART uses two test procedures to select the ?optimal? tree, which is the tree with the lowest overall misclassification cost, thus the highest accuracy. Both test disciplines, one for small datasets and one for large, are entirely automated, and they ensure the optimal tree model will accurately classify existing data and predict results.

For smaller datasets and cases when an analyst does not wish to set aside a portion of the data for test purposes, CART automatically employs cross-validation. This frequently occurs in medical research, but a shortage of training data can occur in the study of any rare event, such as specific types of fraud. In cross-validation, 10 different trees are typically grown, each built from a different 10 percent of the total sample. When the results of the 10 trees are put together, a highly reliable determination of the optimal tree size is obtained. For large datasets, CART automatically selects test data or uses pre-defined test records or test files to self-validate results.

**Q11. What is a "multiple-tree, committee-of-expert method", or "bootstrap aggregation"?**

A11. The use of multiple trees in a committee of experts is a relatively new technique, and one of CART?s creators has developed a dramatically effective way of combining trees in CART. Prediction errors can be reduced as much as 50 percent by directing CART to draw 50 or more different random samples from the training data; growing a different tree on each sample; and then allowing the different trees to ?vote? on the best classification. When appropriate, combining trees can yield a substantial performance edge over any other data-mining procedure. For more information, see [Committee of Experts](#).

**Q12. *When can CART be used to advantage as a standalone package?***

A12. Most data-mining projects involve classification for gaining insight into existing data and turning that knowledge into a predictive model. Typical classification projects include sifting profitable from unprofitable; detecting fraudulent claims; identifying repeat buyers; profiling high-value customers who are likely to churn; and flagging high credit-risk applications. CART is a state-of-the-art classification tool that, as a standalone package, can investigate any classification task and provide a robust, accurate predictive model. The software tackles the core data-mining challenges by accommodating classification - for categorical variables, such as responder and non-responder - and regression - for continuous variables, such as sales revenue.

In addition to delivering accuracy, CART offers three distinct advantages over other data-mining tools. First, CART is easily accessible to beginning users, and it does not require a high level of technical expertise to operate. CART's new, user-friendly GUI and reference manual guide users through a quick process, and the default settings perform so well that many highly experienced experts do not change them. Second, CART results are extremely easy to interpret; the tree-shaped flow chart easily identifies the most important predictors. Lastly, CART costs thousands of dollars less than a data-mining suite, and it comparably handles classification projects.

**Q13. How can CART complement other data-mining packages and/or suites?**

A13. CART is an excellent pre-processing complement to data-mining packages, such as SAS®. In the first stage of a data-mining project, CART can extract the most important variables from a very large list of potential predictors. Focusing on the top variables from the CART model can significantly speed up neural networks and other data-mining techniques. For neural nets in particular, CART bypasses "noise" and irrelevant variables, quickly and effectively selecting the best variables for input. The result is significant reductions in neural-net training speeds and more accurate and robust neural networks. In addition, the CART outputs, or "predicted values," can be used as inputs to the neural net.

CART can also be used to:

- establish performance benchmarks;
- detect important interactions that should be included in statistical models; and,
- impute values for variables with missing values.

**Q14. How quickly can CART generate results?**

A14. CART's efficient algorithm generates results much faster than other methods, such as neural nets. On industry-standard servers, CART models based on 300,000 records and 1,000 variables can be generated in less than an hour. More typical problems involving 100,000 records and 450 variables run in approximately 10 minutes, while 100 variables and one-million records can be run in less than 30 minutes. Exploratory analyses based on extracts from a large database can be conducted even faster; for example, a sample of 30,000 records with 100 selected input variables can be explored in less than five minutes.