Machine Learning (ML1)

CART

By Dorina Grossu

Summary

- What is CART?
- What makes CART easy to interpret?
- How are CART decision tree grown?
- Why is CART unique among decision tree tools?
- What splitting criteria does CART provide?
- What are adjustable misclassification penalties?
- What are intelligent surrogates for missing values?
- What are CART's test procedures?
- What is a committee-of-experts or bootstrap aggregation?
- How can CART complement other data mining packages and/or suites?

Learning Outcomes

- Methodology used by CART
- Where and when CART is used?
- CART Benefits

What is CART?

- **C**lassification **A**nd **R**egression **T**rees is a decision-tree procedure introduced in 1984 by world-renowned 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.

Cont.’

- 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 with the answers tracing a path down the tree.
- The end point reached determines the **classification** or **prediction made by the model**, which can be a **qualitative judgment** or a **numerical forecast**.
- The CART methodology solves:
  - performance,
  - accuracy,
  - operational problems.

What makes CART so easy to interpret?

- The results of a decision tree are displayed as a **tree diagram** using a simple set of if-then rules.
- Discovered relationships and patterns - even in massively complex datasets with thousands of variables - are presented as a **flow chart**.
- Compared to complex parameter coefficients in a logistic regression, or a stream of numbers in neural-nets, the visual display enables users to see the **hierarchical interaction of the variables**.

Cont.’

- Compared to complex parameter coefficients in a logistic regression, or a stream of numbers in neural-nets, the visual display enables users to see the **hierarchical interaction of the variables**.

How are CART decision trees grown?

- CART uses an exhaustive, recursive partitioning routine to generate binary splits that divide each parent node into two child nodes by posing a series of yes-no questions.
- CART **searches for questions** that split nodes into relatively homogenous child nodes.
- As the tree evolves, the nodes become increasingly more homogenous, identifying segments.

Why is CART unique among decision tree tools?

- Reliable pruning strategy - CART's developers determined that a no-stopping rule would discover the optimal tree, so they introduced the strategy of over-growing trees and then pruning them back. This fundamental idea ensures that important structure is not overlooked by stopping too soon.
- Powerful binary-split search approach - CART's binary decision tree is sparing with data and detects structures before too little data are left for further splitting.
- Automatic self-validation procedures - CART's test methods ensure that the tree structure will retain its predictive power when applied to new data. The testing and selection of the optimal tree are integral parts of the CART algorithm.

What splitting criteria does CART provide?

- CART includes several single-variable splitting criteria for classification.
  - It offers gini, symgini, twoing, ordered twoing, class probability, and class entropy.
- For regression, CART provides least squares and least absolute deviation.
- Additionally, CART offers one multi-variable splitting criteria using linear combinations.

What are adjustable misclassification penalties?

- Misclassification penalties accommodate situations in which some misclassified segments are more serious than others.
- CART users can specify an adjustable penalty for misclassifying certain segments, and the software will direct the tree away from that type of error.
- When CART cannot guarantee a correct classification, it will try to ensure that the misclassification is less costly.

What are intelligent surrogates for missing values?

- CART handles missing values by use of *surrogate splits, a splitting rule that closely mimics the action of a primary split*.
- Not only must a good surrogate split the parent node into descendant nodes similar in size and composition to the primary descendant nodes, but, to the extent possible, the surrogate must also match the primary split on the specific cases that go to the left and right child nodes.
- A surrogate is thus evaluated by its ability to match a primary split on a case-by-case basis.

What are CART's test procedures?

- CART uses two test procedures to select the optimal tree with the lowest overall misclassification cost, thus the highest accuracy.
- Both test disciplines are automated and ensure the *optimal tree will accurately classify existing data and predict results*.
- For smaller datasets, the user can employ cross-validation where ten different trees are typically grown, each built from a different ten percent of the total sample.
- When the results of the ten trees are considered, the optimal tree size is obtained. For larger datasets, the user may specify a random test sample, a separate test file, or a pre-determined set of test records.

What is a committee-of-experts or bootstrap aggregation?

- The use of multiple trees in a committee-of-experts is an effective way of combining trees.
- Prediction errors can be considerably reduced by selecting many different random samples from the training data, then, by growing a different tree on each random sample, and, finally, by allowing the different trees to “vote” on the best classification.

How can CART complement other data mining packages and/or suites?

- CART is an excellent pre-processing complement to classical statistical packages, such as SAS®.
- In the first stage of a data mining project, CART can extract the most important variables from a 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.

Cont.’

- For neural nets, CART bypasses noise and irrelevant variables, effectively selecting the best variables for input.
- In addition, 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, and
  - impute missing values.

Thank you!