

# Binary recursive partitioning: Use in psychology

Edgar C. Merkle  
Wichita State University  
IMPS 2008

## ① Background

## ② Example

## ③ Simulations

## ④ Conclusions

# Binary Recursive Partitioning

Background

Example

Simulations

Conclusions

- Binary recursive partitioning (BRP, AKA classification and regression trees; Breiman et al., 1984): Building a decision tree to predict a response variable.
  - Partitioning: Data are partitioned into homogeneous subgroups based on values of predictor variables.
  - Binary: Data are always partitioned into two groups.
  - Recursive: After data have been partitioned once, more partitions are created within the original partitions.

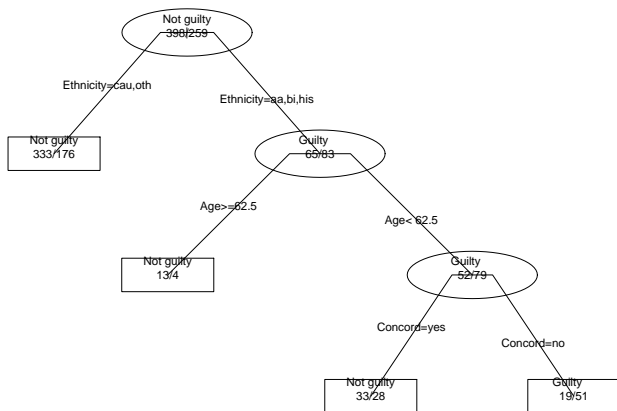
# BRP Example

Background

Example

Simulations

Conclusions



# Tree Construction

Background

Example

Simulations

Conclusions

- Fundamental issues in BRP:
  - Finding optimal partitions of predictors
  - Determining the optimal tree size

# Optimal Partitions

- Partitions are obtained via a greedy search; at each step, all possible partitions are searched to find the “best.”
- “Best” defined via an impurity function, measuring extent to which observations within a node are heterogeneous (small impurities  $\rightarrow$  good).
- Impurity for node  $t$ :

- Categorical response:

$$i(t) = \sum_{j \neq i} p(j | t)p(i | t)$$

- Continuous response:

$$i(t) = \sum_{i=1}^{n_t} (y_i - \bar{y}_t)^2$$

# Optimal Tree Size

Background

Example

Simulations

Conclusions

- Tree is initially constructed to be very big, which leads to overfitting of the observed data.
- Tree is then *pruned*: unimportant branches are removed to yield a smaller tree.
- Tree “goodness” is measured by sequentially pruning unimportant branches and examining predictive accuracy via cross-validation.

# Psychology Applications

Background

Example

Simulations

Conclusions

- While BRP may have negative connotations as *data mining*, it has the potential to be useful in many psychology applications.

## Example

Background

Example

Simulations

Conclusions

- Examining mock jurors' verdicts in a medical malpractice case (Arkes, Shaffer, & Medow, 2008).
  - Three experimental predictors, twelve demographic predictors;  $N = 657$
  - Response: Liable/not liable
  - To select predictor variables and deal with multicollinearity, researchers originally conducted a 3-step procedure with contingency tables & logistic regression.
- BRP yields similar results to the researchers' original findings, but in one step instead of three.

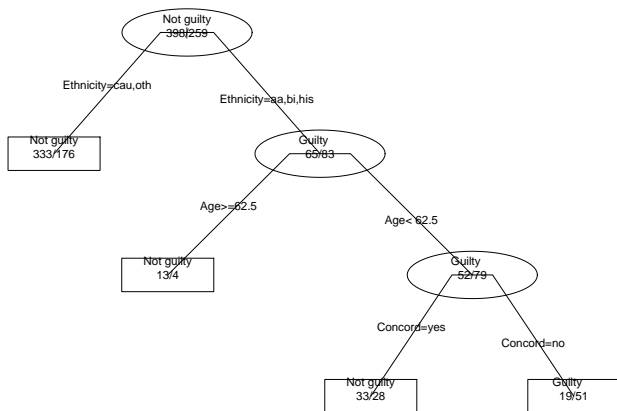
# Example

Background

Example

Simulations

Conclusions



# Simulations

- There are few applications of BRP to data with small-moderate sample sizes/small numbers of predictor variables.
  - Simulations: How does BRP compare to regression in these situations?
  - Strategy: Generate 1,000 samples of data from a known model, examine cross-validated  $R^2$  for regression and BRP.

# Simulation 1

Background

Example

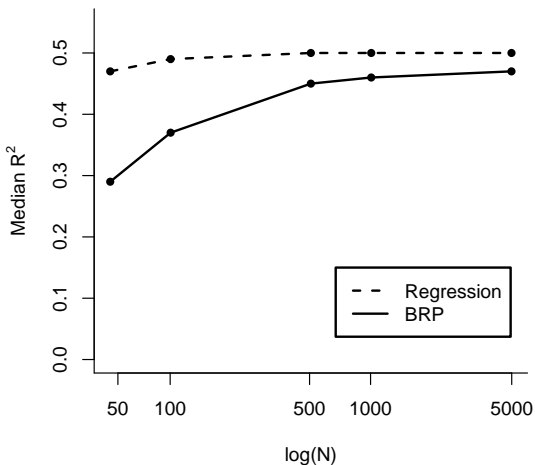
**Simulations**

Conclusions

- Simulation 1: One predictor, linear model,  $N = 50$ :  
 $Y = X + N(0, \sigma^2 = 1)$ 
  - Median  $R_{\text{reg}}^2 = .47$
  - Median  $R_{\text{brp}}^2 = .29$

# Simulation 1

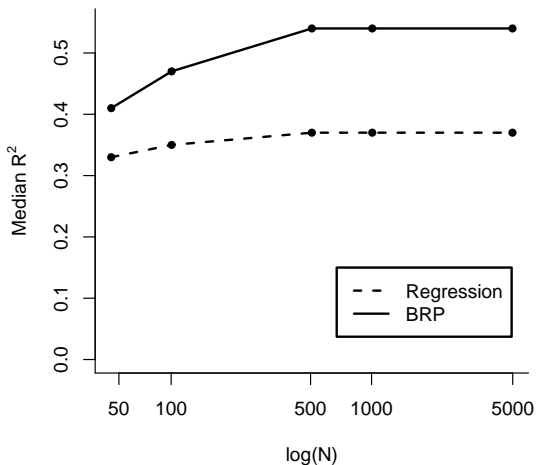
- Impact of sample size on  $R^2$ :



- Simulation 2: One predictor, exponential function,  
 $N = 50$ :  
 $Y = 5^X + N(0, \sigma^2 = 25)$ 
  - Median  $R_{\text{reg}}^2 = .33$
  - Median  $R_{\text{brp}}^2 = .41$

## Simulation 2

- Impact of sample size on  $R^2$ :



- Tentative conclusions:
  - Depending on the data, BRP predictions can be more accurate than regression at  $N = 50$  and one predictor.
  - When data come from a regression model, BRP predictions are almost as accurate as regression at large  $N$ .
  - $R_{\text{reg}}^2$  can be considerably smaller than  $R_{\text{brp}}^2$  at all sample sizes.

# Conclusions

- Binary recursive partitioning is a useful tool for social science research. It is especially suitable for:
  - Nonlinear relationships
  - Moderate to large sample sizes
  - Many variables; variable selection
- Interpretable
- Few assumptions
- No significance tests

Questions?

## Example

Background

Example

Simulations

Conclusions

- “Not guilty” base rate: 57%
- Tree accuracy: 63% within sample, 60% cross-validated

# Comparison

- Compared to linear models, BRP has:
  - Heavier focus on cross-validated predictive accuracy
  - Fewer assumptions (i.e., no normality, homoscedasticity)
  - No significance tests