Introduction to the bartMachine R package

Saint Louis R User Group

John Snyder

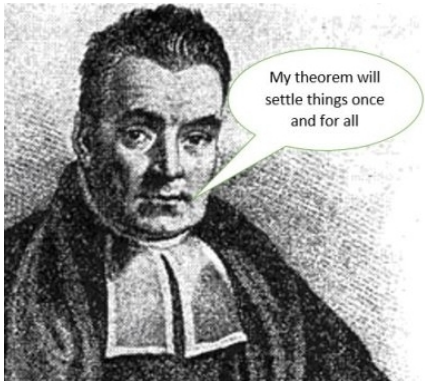
April 11, 2019

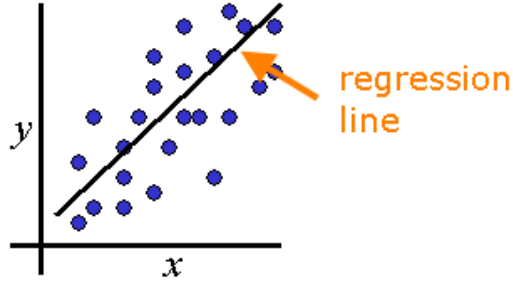
Outline

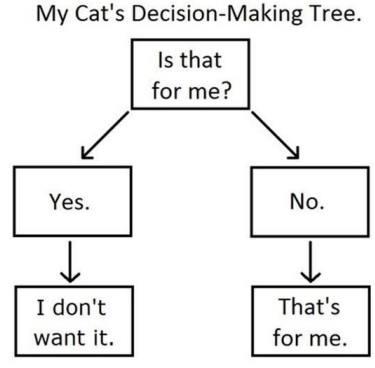
1. Brief BART overview
2. Installation and features
3. Demo
4. Further Considerations

What is BART?

Bayesian **A**dditive **R**egression **T**rees







Interpretation

- Emsamble method combining many shallow trees
- Bayesian means variation is fully quantified
  - Yay Statistics

Powerful Predictive Performance

- 100 replications of out of sample predictive performance

Function	BART	XGBoost	Random Forest	Linear Reg(lol)
Friedman	1.08	1.21	1.64	2.61
Mirsha's Bird	1.53	2.78	2.90	26.59
Weird Exp	1.04	1.05	1.07	6.08
Linear	1.025	1.032	1.034	1.004

- bartMachine is relatively unknown
  - xgboost: ~43k downloads per month
  - randomForest: ~88k downloads per month
  - bartMachine: ~2k downloads per month

Package Features:

- Functions for Cross Validation
- Model fitting:
  - Is done in parallel[1]
  - Can incorporate missing data
- Lots of fun statistical things
  - Credible iterval calculation
  - Diagnostic plots/tests
- Variable selection
- Interaction detection
- Export fit trees

[1] MCMC

Installation and loading steps

1. Google "How to install rJava on [your OS]"
2. Do that
3. Run the following

```
install.packages("bartMachine")
```

To load the package with:

- 10GB of memory
- All but one core available for compute

```
options(java.parameters = "-Xmx10g")
library(bartMachine)
numcores <- parallel::detectCores()
set_bart_machine_num_cores(numcores - 1)
```

Code Time

Coding demo

Computational Considerations

- Table with memory/time

John's Final Thought

- BART is a powerful technique which brings many advantages
  - At the expense of computational efficiency.
- Good results with removing expected variation and feeding residuals into BART.