



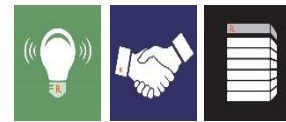
# Ensemble Methods

## Gradient Boosted Trees



# Learning Objectives

1. Understand Principles of Gradient Boosting
2. Facility with R package gbm
  - A. Application to different problem types
  - B. Parameters for controlling training
  - C. Interpreting Results
    - i. Out of Sample Performance
    - ii. Influence Ranking of Attributes
    - iii. Pictures of Model Output for Different Attribute Values



# What are Ensemble Methods?

- General Purpose Predictive Algorithms
- Combine Results from Numerous Algorithms – “Base Learners”
- For Example: Train an SVM, Logistic Regression and BDT then Vote
- That Generates 3 Different Models – How to Generate Thousands?



# Bagging

- Use BDT as Base Learner.
- Instead of Training on Full Training Data Set, Train on 50% of Rows Randomly Selected
- That Gives One Model
- Resample Random 50% and Repeat
- R Script 1 - Bagging



# Gradient Boosting

- Developed by Prof Jerry Friedman (see refs)
- For a Regression Problem
  - Targets  $Y$  and Attributes  $X$
  - Train a Tree using  $X$  to Predict  $Y$
  - Call the Predictions  $T_1(X)$  (the predict function)
  - Pick a Step Size Parameter  $\epsilon$  (say 0.01)



# Gradient Boosting - Iteration

- Iteration
  - Train  $T_2$  to predict  $Y - \epsilon T_1(X)$
  - Train  $T_3$  to predict  $Y - \epsilon T_1(X) - \epsilon T_2(X)$
  - Train the  $n$ th tree to predict  $Y - \epsilon \sum T_i(X)$
- R Script 2 – Gradient Boosting



# R package gbm

- Main Functions
- Training – `gbm()`, `gbm.fit()`
- Visualize Training Progress – `gbm.perf()`
- Understanding variable usage and importance – `plot()`, `summary ()`



# gbm Training Functions

- Simplest specification for regression

dataSet holds attributes and labels (col name “y”)

```
myGbmObj <- gbm(y~., data = dataSet,          distribution="gaussian",  
                n.trees = 100,  
                interaction.depth = 1  
                shrinkage = 0.001,  
                cv.folds=10)
```





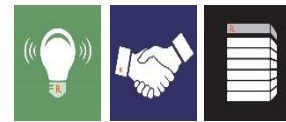
# Training gbm on Concrete Data

- Starting Point
  - Regression Problem -> gaussian or laplace
  - n.trees -> start with 2000 to 3000
  - cv.folds -> 10 make smaller if 10x repetitions takes too long
  - shrinkage -> 0.001 tune
  - interaction.depth -> start with 1 and work up
- R script Gradient Boosting 3



## gbm on Classification Prob

- Classification setup is similar. Only a Few Differences
  - distribution = “bernoulli”
  - Targets must be 0,1
- R Script Gradient Boosting 4



# Understanding gbm Models

- gbm has several functions to help understand what driving the predictions from the trained model.
- `summary()` - ranks attributes by importance
- `plot()` – plots effect on prediction of change in a single variable.



# Wrap Up

- Review Learning Objectives
- Outline of Next Section



## References

- J.H. Friedman (2001). “Greedy Function Approximation: A Gradient Boosting Machine,” *Annals of Statistics* 29(5):1189-1232.
- J.H. Friedman (2002). “Stochastic Gradient Boosting,” *Computational Statistics and Data Analysis* 38(4):367-378.