ADVANCED MACHINE LEARNING: BOOSTING

Austin SIGKDD
8/3/2016

Executive Summary

- Boosting
  - Ensemble method
- Basic concept
  - Multiple weak learners combined algorithmically to form a strong model
- Primary implementation is decision trees
  - Inherits best qualities, fixes the worst
- Best off the shelf algorithm for transactional data
  - Popular in competitions & real-world modeling
- Variance in implementations (GBM)
  - R, Sci-kit, H2O, XGBoost, Salford (MART)

Background

Lead Data Scientist, H2O

BS, computer science

Additional roles: data warehousing, BI, analytics

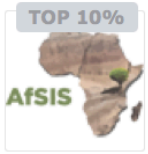
3rd/260

5th/321

6th/381

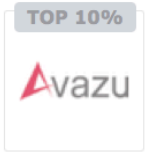
10th/340

16th/1323



15th/1233

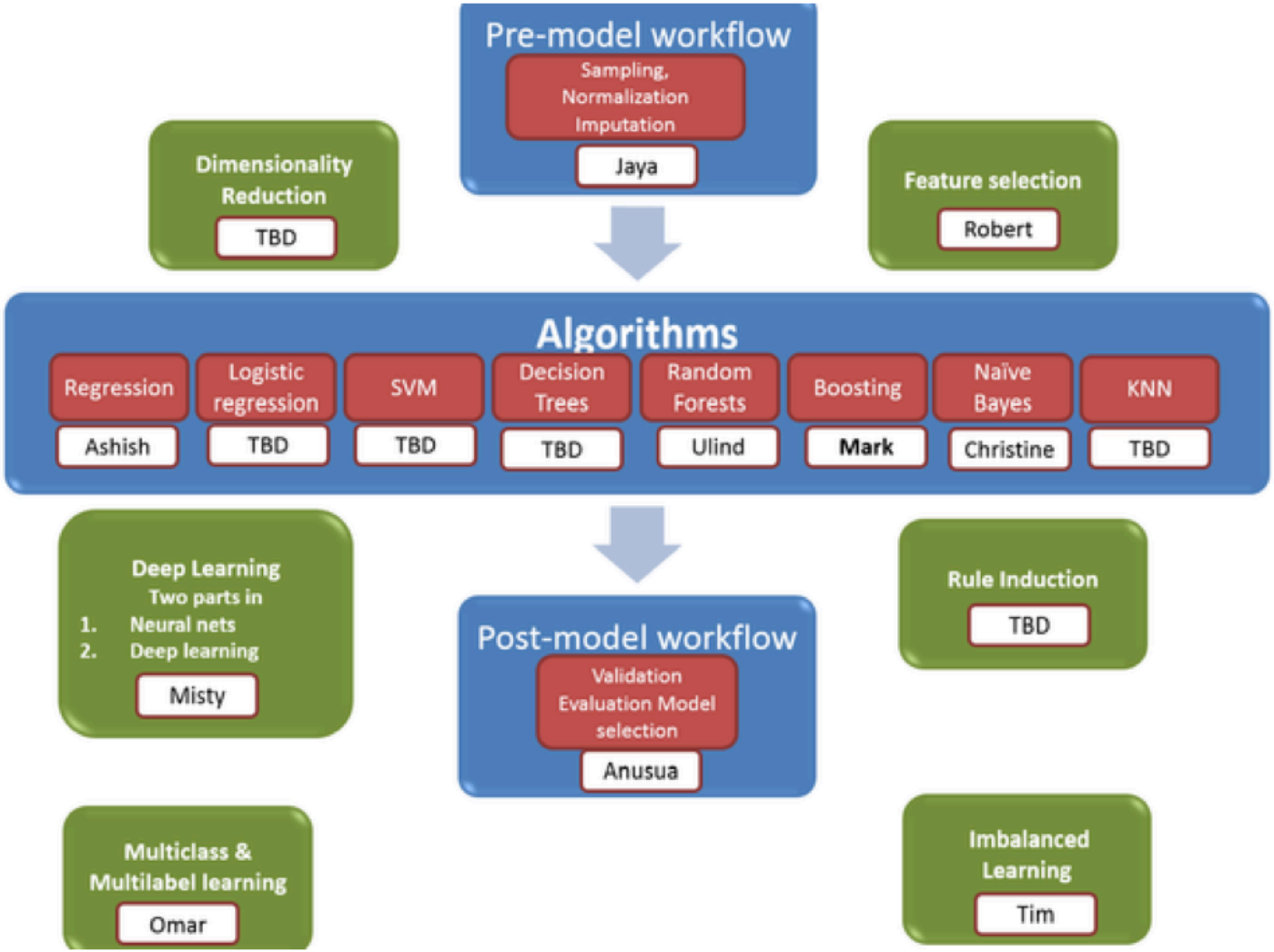
44th/3303



26th/1604

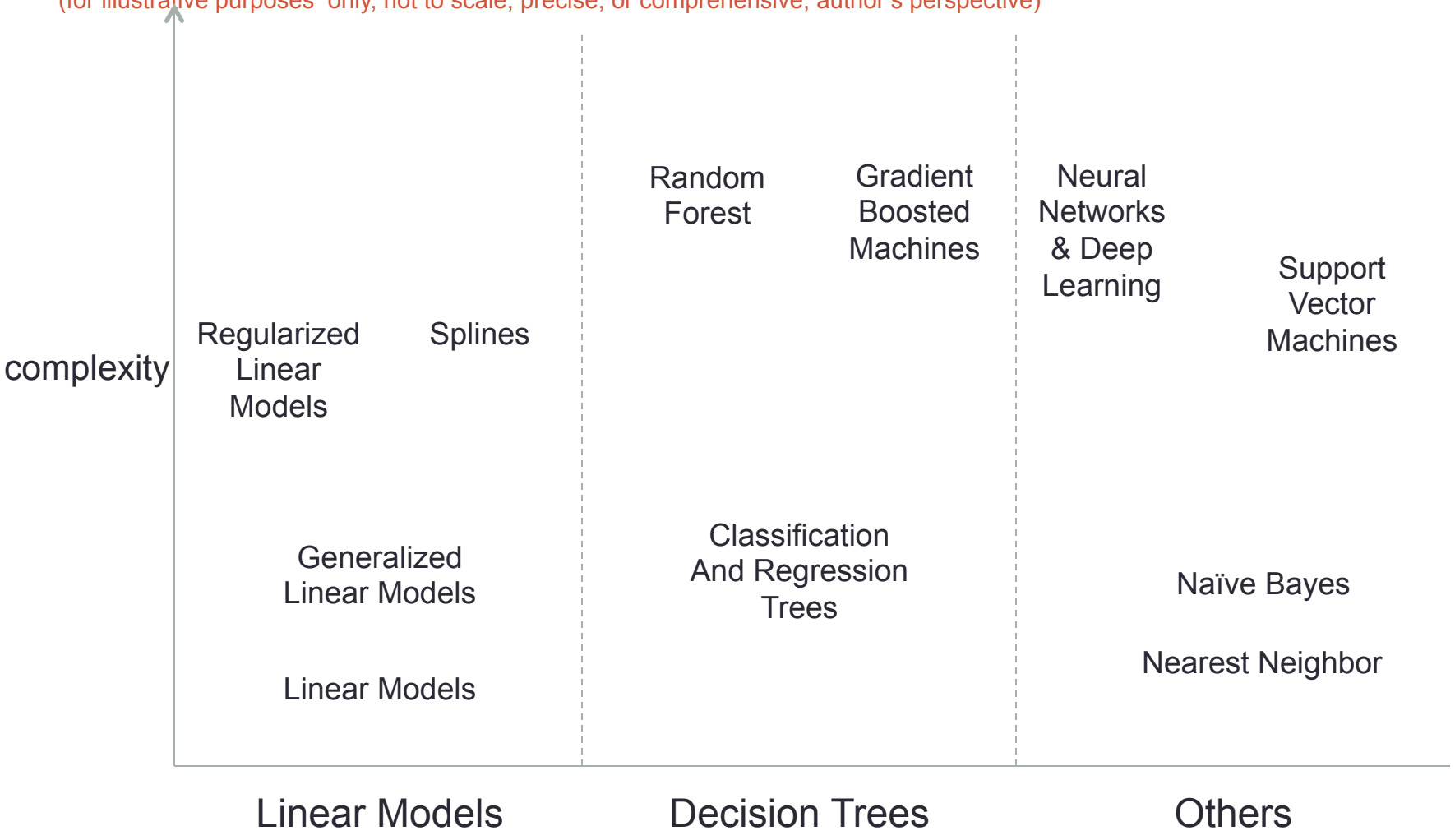
Competitions

Advanced ML Topics



Predictive Modeling Landscape: General Purpose Algorithms

(for illustrative purposes only, not to scale, precise, or comprehensive; author's perspective)



More Comprehensive List: http://caret.r-forge.r-project.org/modelList.html

Ensembles in Sci-Kit Learn

`sklearn.ensemble`: Ensemble Methods

The `sklearn.ensemble` module includes ensemble-based methods for classification and regression.

User guide: See the Ensemble methods section for further details.

`ensemble.AdaBoostClassifier` ([...])	An AdaBoost classifier.
`ensemble.AdaBoostRegressor` ([base_estimator, ...])	An AdaBoost regressor.
`ensemble.BaggingClassifier` ([base_estimator, ...])	A Bagging classifier.
`ensemble.BaggingRegressor` ([base_estimator, ...])	A Bagging regressor.
`ensemble.ExtraTreesClassifier` ([...])	An extra-trees classifier.
`ensemble.ExtraTreesRegressor` ([n_estimators, ...])	An extra-trees regressor.
`ensemble.GradientBoostingClassifier` ([loss, ...])	Gradient Boosting for classification.
`ensemble.GradientBoostingRegressor` ([loss, ...])	Gradient Boosting for regression.
`ensemble.RandomForestClassifier` ([...])	A random forest classifier.
`ensemble.RandomTreesEmbedding` ([...])	An ensemble of totally random trees.
`ensemble.RandomForestRegressor` ([...])	A random forest regressor.
`ensemble.VotingClassifier` (estimators[, ...])	Soft Voting/Majority Rule classifier for unfitted estimators.

Ensembles in Sci-Kit Learn

`sklearn.ensemble`: Ensemble Methods

The `sklearn.ensemble` module includes ensemble-based methods for classification and regression.

User guide: See the Ensemble methods section for further details.

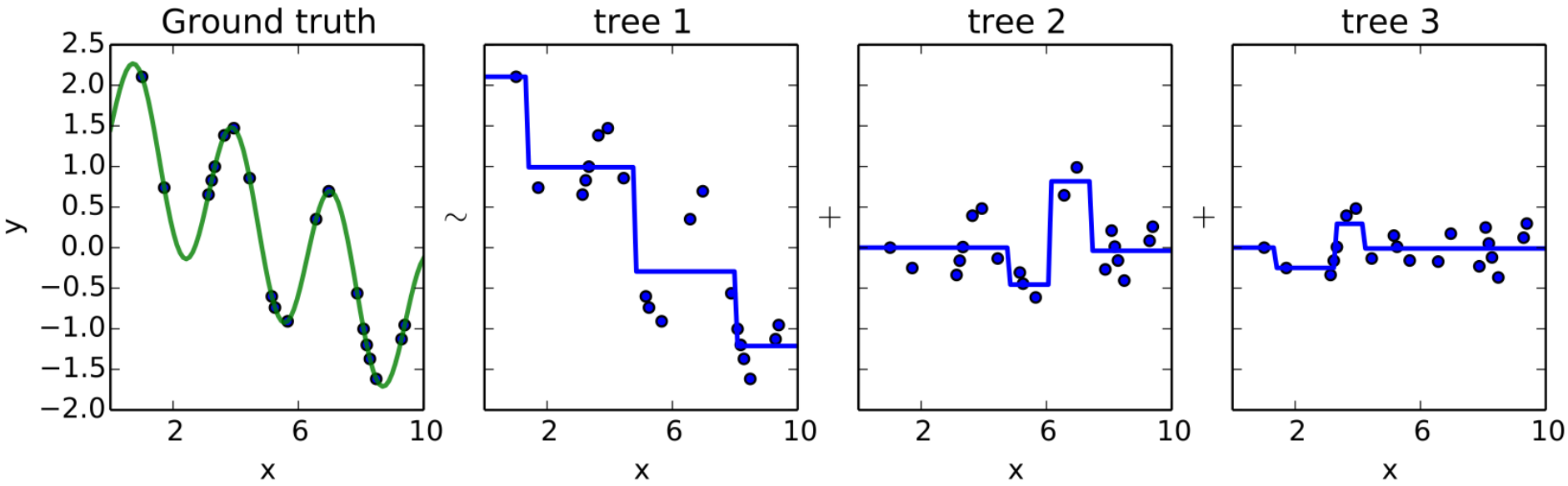
`ensemble.AdaBoostClassifier` ([...])	An AdaBoost classifier.
`ensemble.AdaBoostRegressor` ([base_estimator, ...])	An AdaBoost regressor.
`ensemble.BaggingClassifier` ([base_estimator, ...])	A Bagging classifier.
`ensemble.BaggingRegressor` ([base_estimator, ...])	A Bagging regressor.
`ensemble.ExtraTreesClassifier` ([...])	An extra-trees classifier.
`ensemble.ExtraTreesRegressor` ([n_estimators, ...])	An extra-trees regressor.
`ensemble.GradientBoostingClassifier` ([loss, ...])	Gradient Boosting for classification.
`ensemble.GradientBoostingRegressor` ([loss, ...])	Gradient Boosting for regression.
`ensemble.RandomForestClassifier` ([...])	A random forest classifier.
`ensemble.RandomTreesEmbedding` ([...])	An ensemble of totally random trees.
`ensemble.RandomForestRegressor` ([...])	A random forest regressor.
`ensemble.VotingClassifier` (estimators[, ...])	Soft Voting/Majority Rule classifier for unfitted estimators.

Source: http://scikit-learn.org/stable/modules/classes.html#module-sklearn.ensemble

Boosting

- Common definition
  - Multiple weak learners combined algorithmically to form a strong model
- Weak learner examples
  - Single-featured predictions
  - Small decision trees
  - No definition of weak
- Combined algorithmically
  - Sequential fits focusing on the training cases with highest error
  - Adaboost: reweighting at each iteration based on misclassification error
  - GBM: new target = gradient loss of model as of prior iterations, stochastic with regularization

Revisit Sequential Tree Building



Source: http://www.slideshare.net/DataRobot/gradient-boosted-regression-trees-in-scikitlearn

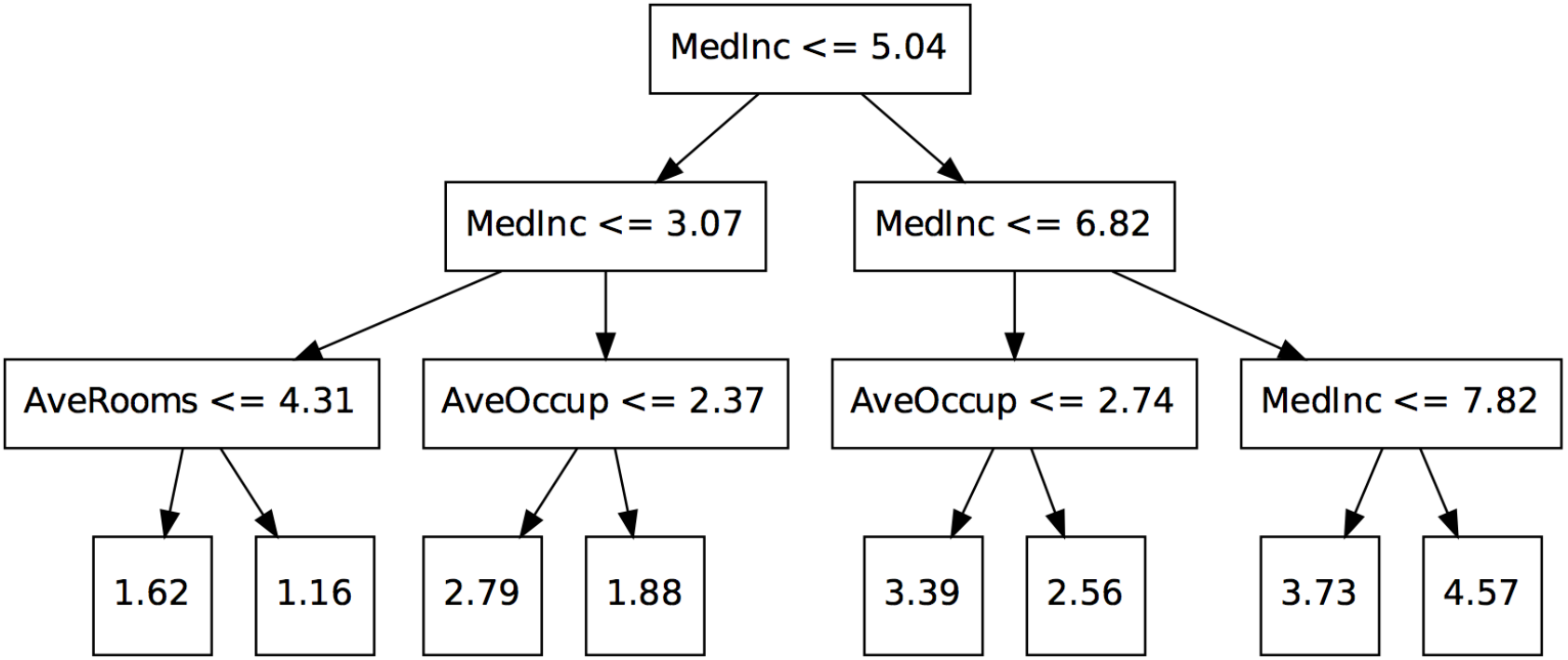
Boosting: Brief History

- 1988: original concept in Kearns/Valiant paper
- 1990: original Freund & Schapire
- 1997: Adaboost, by Freund & Schapire
- 1999: Friedman - Stochastic Gradient Boosting
- 2002: Friedman – Gradient Boosted Machines
- 2014: Extreme gradient boosting (XGBoost) T. Chen

- Interesting how few developments existed between 2002 and 2014

Decision Trees

Decision Tree Overview



Source: http://www.slideshare.net/DataRobot/gradient-boosted-regression-trees-in-scikitlearn

Decision Trees: Practical Use

Strengths

- Non linear
- Robust to correlated features
- Robust to feature distributions
- Robust to missing values
- Simple to comprehend
- Fast to train
- Fast to score

Weaknesses

- Poor accuracy
- Cannot project
- Inefficiently fits linear relationships

Boosting Decision Trees

- Strengths
  - Inherits strengths of decision trees
  - Often best possible model
    - Dominant in Kaggle
    - Increasing adoption in production use cases
  - Robust
  - Directly optimizes cost function
- Weaknesses
  - Overfits
    - Need to find proper stopping point
  - Sensitive to noise and extreme values
  - Several hyper-parameters
  - Lack of transparency

Basic GBM Parameters

- Number of iterations
- Learning rate
- Tree depth
- Minimum number of observations
- Observation sampling
- Column sampling (multiple methods)
- Minimum split gain requirement
- Regularization strength

GBM In Action