LINEAR MODELS

Dr. Hilman F. Pardede

Research Center for Informatics
Indonesian Institute of Sciences

The materials are compiled from the following resources:

- https://github.com/joaquinvanschoren/ML-course
- https://www.cse.iitk.ac.in/users/piyush/courses/ml_autumn16/ML.html
- http://sli.ics.uci.edu/Classes/2015W-273a

ENSEMBLE METHOD: THE BASICS

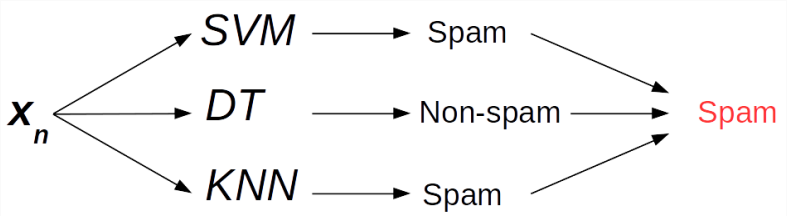
ENSEMBLE METHOD: THE BASICS

Ensemble: combine many predictors

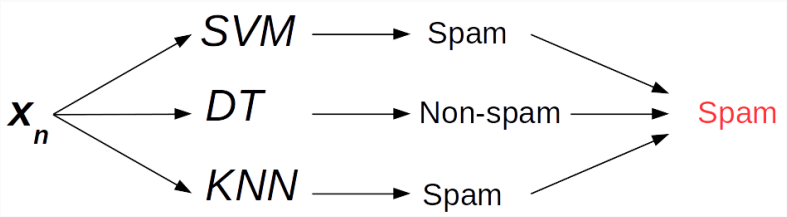
- (Weighted) combinations of predictors
- May be same type of learner or different
- Ex. Gradient boosting and random forests

SOME SIMPLE ENSEMBLES

- Voting or Averaging of predictions of multiple pre-trained models



- "Stacking": Use predictions of multiple models as "features" to train a new model and use the new model to make predictions on test data



- Another approach: Instead of training different models on same data, train same model multiple times on different data sets, and "combine" these "different" models
  - We can use some simple/weak model as the base model
  - How do we get multiple training data sets (in practice, we only have one data set at training time)?

BAGGING: BOOTSTRAP AGGREGATION

- Learn many classifiers, each with only part of the data
- Combine through model averaging

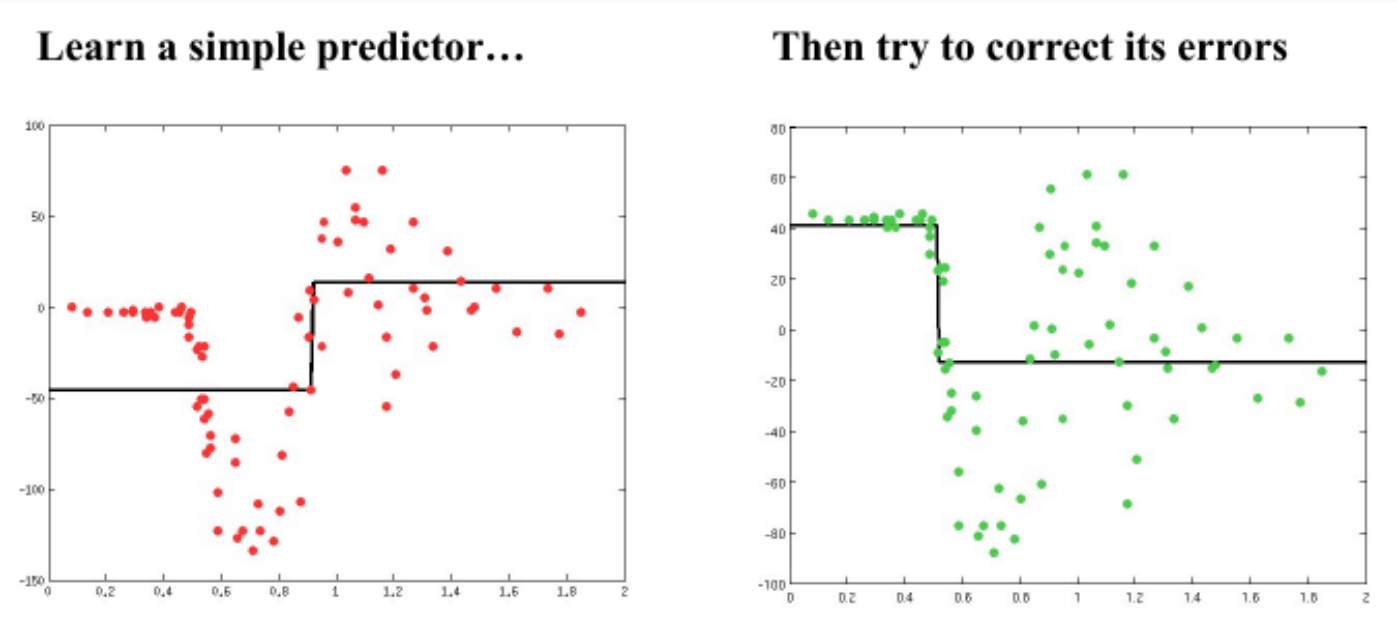
BAGGING

- Bootstrap
  - Create a random subset of data by sampling
  - Draw m' of the m samples, with replacement (sometimes w/o)
- Bagging
  - Repeat K times
    - Create a training set of m' $\leq$ m examples
    - Train a classifier on the random training set
  - To test, run each trained classifier
    - Each classifier votes on the output, take majority
    - For regression: each regressor predicts, take average

RANDOM FOREST

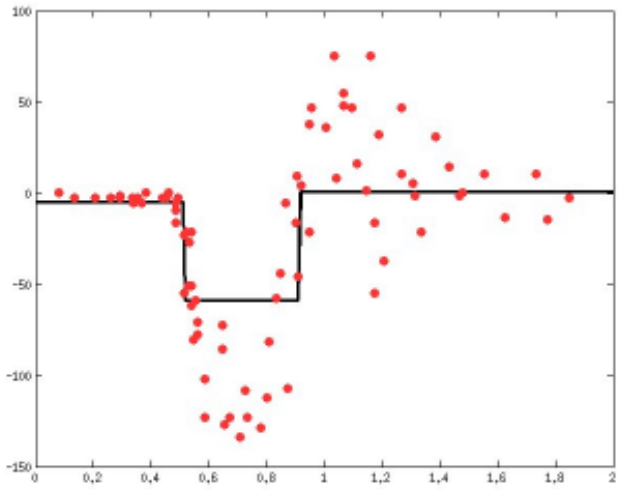
- Bagging applied to decision trees
- Problem
  - With lots of data, we usually learn the same classifier
  - Averaging over these doesn't help!
- Introduce extra variation in learner
  - Take a bootstrap sample of your data
  - At each step of training, only allow a subset of features
  - Enforces diversity ("best" feature not available)
  - Average over these learners (majority vote)

- Learn a regression predictor
- Compute the error residual
- Learn to predict the residual

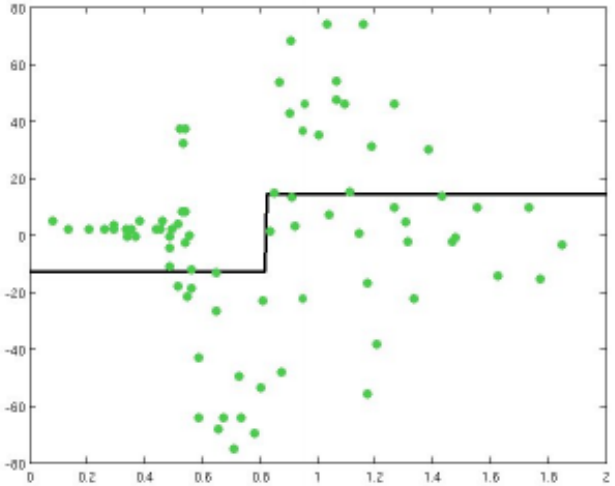


- Learn a regression predictor
- Compute the error residual
- Learn to predict the residual

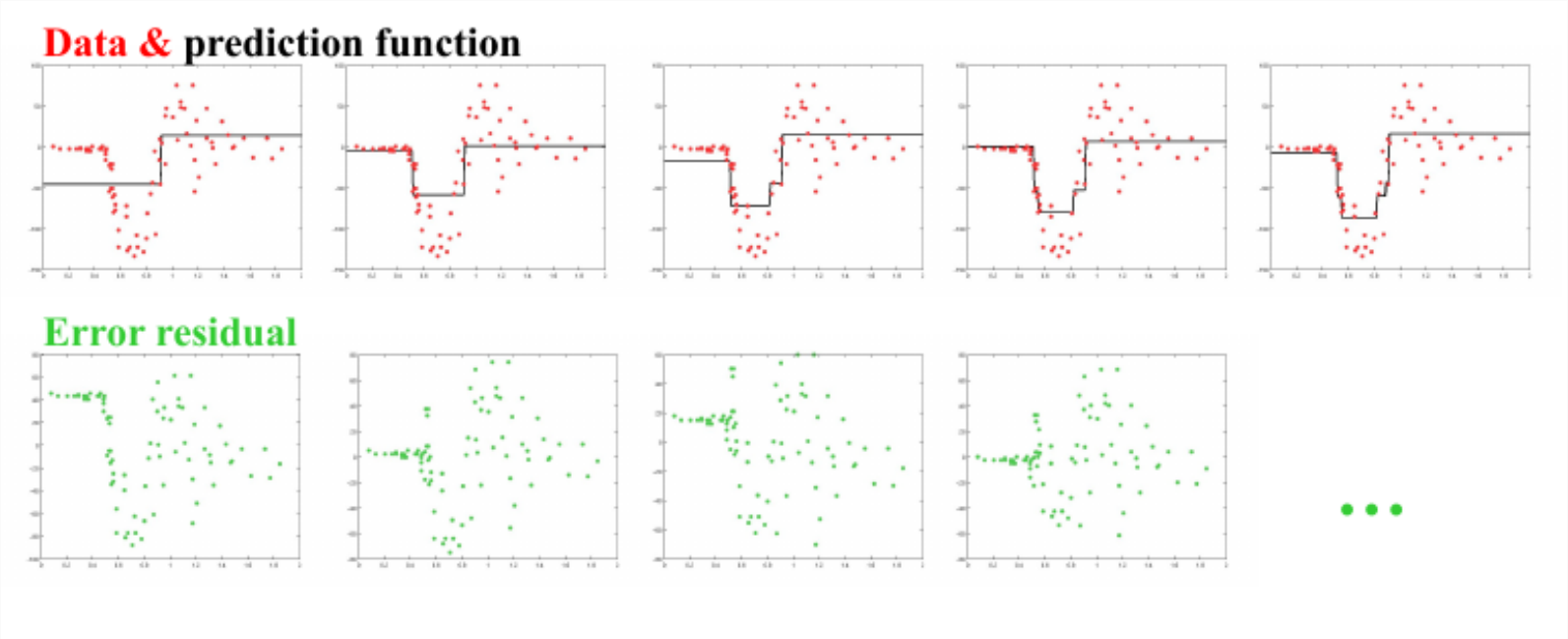
Combining gives a better predictor…



Can try to correct its errors also, & repeat



- Learn sequence of predictors
- Sum of predictions is increasingly accurate
- Predictive function is increasingly complex



SUMMARY

- Bagging / RandomForest is a variance-reduction technique
- Build many high-variance (overfitting) models
- Aggregation (soft voting or averaging) reduces variance
- Parallellizes easily
- Boosting is a bias-reduction technique
- Build many high-bias (underfitting) models
  - Typically shallow decision trees
  - Sample weights are updated to create different trees
- Aggregation (soft voting or averaging) reduces bias
- Doesn't parallelize easily
- It is also possible to build heterogeneous ensembles
- Models from different algorithms
- Are combined by letting each algorithm predict
- Often a meta-classifier/regressor is trained on the predictions: Stacking