Ensemble Methods

Bernhard Huang

Correlaid Muenchen

17 Apr, 2022

Overview

High Level Overview

- What is the goal of this workshop?
  - Covers the bare minimum for improving your classifiers/regressors with ensemble techniques (say, to win Kaggle competitions)

Great Non-Wikipedia Resources

- K. P. Murphy, Machine Learning: A Probabilistic Perspective
- C. M. Bishop, Pattern Recognition and Machine Learning
- S. Shalev-Shwartz, S. Ben-David, Understanding Machine Learning
- M. Hardt and B. Recht, Patterns, Predictions, and Actions

Paradigm

- Ensemble methods are one way to help control overfitting
- As noted before in this workshop series, there is always a bias-variance tradeoff
- There are two commonly used ensemble methods:
  - Boosting: ensemble of high bias, low variance predictors built sequentially
  - Bagging: ensemble of low bias, high variance predictors built in parallel
- Can any of you give me any pros or cons to any of the approaches mentioned above?

Boosting

- Basic recipe (for gradient boosting)
  - Calculate the scores of each class given T trees
  - Convert theses squares to probabilities
  - In each iteration, minimize a regularized objective function (with the aid of the gradient)

Bagging

- General recipe (for random forests)
  - Build T trees independently; for each tree, sample N points with replacement, grow it while only looking at a random subset of n features
  - At inference time: average (regression), or majority vote (classification)
  - Key (hyper)parameters: number of trees T, feature size set n, tree depth
  - Choose best parameters from validation dataset

Exercises

Since this is not a math workshop, let's try out two exercises:

1. Gradient boosting example
2. Random forest example

Questions?