Stacking Algorithm for Ensemble Modelling

Frederik Schreck

Course: Numerical Introductory Course
Lecturer: Prof. Dr. Brenda López Cabrera
Humboldt–Universität zu Berlin

Motivation - The wisdom of the crowd

- The aggregation of individual guesses in groups is often superior to individual guesses - even to experts
- BUT: Only fulfilled under certain criteria
  - Variation of guesses
  - Independence of guesses
  - Decentralization
  - Algorithm

Outline

Ensemble Learning - Terminology

Machine Learning

- Part of computer science that uses statistical techniques to train models on data
- Typically used for prediction purposes

Stacking and Ensemble Learning

- Idea is to combine hypotheses of multiple learning algorithms (base learners)
- Goal is to obtain a better predictive performance than with each of the single algorithms alone
- Mainly used in supervised learning
- Very flexible method

Ensemble Learning

Which models to combine?

- Effective ensembling builds on diverse and little correlated models
- Best to use strong base learners

Similar criteria as mentioned in the Motivation!

Decision Tree

- Idea: use a set of splitting rules to recursively partition the dataset.
- Classification trees:
  - Minimize impurity within nodes
- Regression trees:
  - Minimize variance of the response variable within nodes

Decision Tree for classification

- Choice of splitting rule: maximizing information gain (IG) by decreasing node impurity (I)

$$IG_n = I_n - p_{n_1} * I(n_1) - p_{n_2} * I(n_2), \tag{1}$$

for node n with branching nodes n_1 and n_2, and p_{n_i} as the fraction of cases in branching node n_i

- How to measure impurity? Choices of splitting criteria:

$$\text{Entropy:} I(n) = -\sum_{j}^{J} p(c_j|n) * \log_2(p(c_j|n)) \tag{2}$$

$$\text{Gini impurity:} I(n) = 1 - \sum_{j}^{J} p(c_j|n)^2 \tag{3}$$

$$\text{Misclassification impurity:} I(n) = 1 - \max_j p(c_j), \tag{4}$$

for classes $c_j, j \in J = \{1, 2, ...\}$

Decision Tree for classification

- Choice of stopping rule: A fully grown tree has pure leaf nodes and may overfit the data
  - Pre-pruning
  - Post-pruning

Bagging

blalala

Random Forest

blalala

Boosting

blalala

Gradient Boosting

blalala

Bayes??

blalala

Stacked Generalization

blalala

Potentials and Problems of Ensemble Learning

blalala

Sources

Kuncheva, L. I. and Whitaker, C. J. (2003).
Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy.
Machine learning, 51(2):181–207.

Surowiecki, J. (2005).
The Wisdom of Crowds.
Anchor.