

11.2 Random Forests

Machine Learning 1: Foundations

Marius Kloft (TUK)

Kaiserslautern, 30 June – 7 July 2020

1 Decision Trees

2 Random Forests

Bagging

One major problem with trees is their high variance:

- ▶ often a small change in the data can result in a very different series of splits

Bagging averages many trees to reduce this variance

Bagging [1]

Definition

- 1: randomly draw data sets **with replacement** from the training data, each having the same size as the original training set
- 2: grow a tree on each data set
- 3: average the resulting prediction functions (regression) or perform majority vote (classification)

Note:

- ▶ in principle, this trick can be applied to any arbitrary learning machine algorithm (e.g., SVM, KRR, neural network)
- ▶ but in practice it works best with (small) decision trees
- ▶ the resulting ensemble of trees is called a **forest**

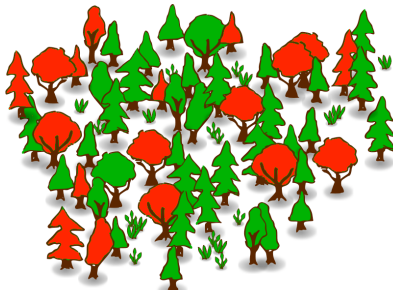
Random Forests [2]

Definition

Pretty much the same as bagging (of decision trees) except for one difference:

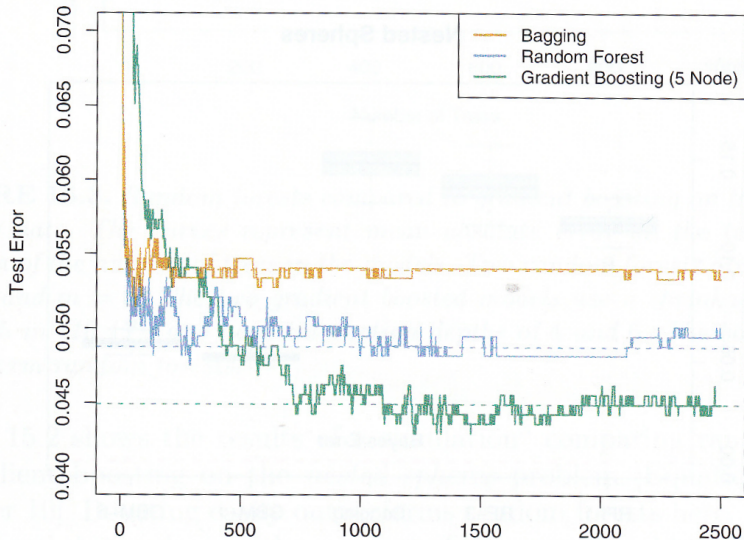
- ▶ when growing a tree of the forest, for each split of an internal node in the tree, we randomly select a subset of $m < d$ many features.
- ▶ The optimal split feature j is determined only among those m features.

Note: each node uses a new random draw of m features



Example (SPAM data)

Spam Data



Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?

Manuel Fernández-Delgado

Eva Cernadas

Senén Barro

CITIUS: Centro de Investigación en Tecnologías da Información da USC

University of Santiago de Compostela

Campus Vida, 15872, Santiago de Compostela, Spain

MANUEL.FERNANDEZ.DELGADO@USC.ES

EVA.CERNADAS@USC.ES

SENEN.BARRO@USC.ES

Dinani Amorim

Departamento de Tecnologia e Ciências Sociais- DTCS

Universidade do Estado da Bahia

Av. Edgard Chastinet S/N - São Geraldo - Juazeiro-BA, CEP: 48.305-680, Brasil

DINANIAMORIM@GMAIL.COM

Editor: Russ Greiner

Abstract

We evaluate 179 classifiers arising from 17 families (discriminant analysis, Bayesian, neural networks, support vector machines, decision trees, rule-based classifiers, boosting, bagging, stacking, random forests and other ensembles, generalized linear models, nearest-neighbors, partial least squares and principal component regression, logistic and multinomial regression, multiple adaptive regression splines and other methods), implemented in Weka, R (with and without the caret package), C and Matlab, including all the relevant classifiers available today. We use 121 data sets, which represent the whole UCI data base (excluding the large-scale problems) and other own real problems, in order to achieve significant conclusions about the classifier behavior, not dependent on the data set collection. The classifiers most likely to be the bests are the random forest (RF) versions, the best of which (implemented in R and accessed via caret) achieves 94.1% of the maximum accuracy overcoming 90% in the 84.3% of the data sets. However, the difference is not statistically significant with the second best, the SVM with Gaussian kernel implemented in C using LibSVM, which achieves 92.3% of the maximum accuracy. A few models are clearly better than the remaining ones: random forest, SVM with Gaussian and polynomial kernels, extreme learning machine with Gaussian kernel, C5.0 and avNNet (generalized linear model).



Acknowledgment

This lecture is based on [3], Chapter 9 and 15. The figure on Slide 5 is taken from [3].

Refs I



Bagging predictors, *Machine Learning*, vol. 24, no. 2, pp. 123–140, 1996.



L. Breiman, Random forests, *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.



T. Hastie, R. Tibshirani, and J. Friedman, The elements of statistical learning, 2nd edition. Springer series, 2009.