Genuer, Robin
Variance reduction in purely random forests. (English) Zbl 1254.62050
J. Nonparametric Stat. 24, No. 3, 543-562 (2012).

Summary: Random forests (RFs), introduced by *L. Breiman* [Mach. Learn. 45, No. 1, 5–32 (2001; Zbl 1007.68152)], are a very effective statistical method. The complex mechanism of the method makes the theoretical analysis difficult. Therefore, simplified versions of RF, called purely RFs (PRFs), which can be theoretically handled more easily, have been considered. We study the variance of such forests. First, we show a general upper bound which emphasises the fact that a forest reduces the variance. We then introduce a simple variant of PRFs, that we call purely uniformly RFs. For this variant and in the context of regression problems with a one-dimensional predictor space, we show that both random trees and RFs reach the minimax rate of convergence. In addition, we prove that compared with random trees, RFs improve accuracy by reducing the estimator variance by a factor of three-fourths.

MSC:

62G08	Nonparametric regression and quantile regression
68T05	Learning and adaptive systems in artificial intelligence
62G20	Asymptotic properties of nonparametric inference
65C60	Computational problems in statistics (MSC2010)

Cited in **10** Documents

Keywords:

nonparametric regression; rates of convergence; randomisation; ensemble methods

Software:

GeneSrF

Full Text: DOI

References:

[1] Arlot, S. (2008), 'V-Fold Cross-Validation Improved: V-Fold Penalization', preprint, arXiv:0802.0566v2.

[2] Biau, G. (2012), 'Analysis of a random forests model',Journal of Machine Learning Research, in press. · Zbl 1283.62127

[3] Biau G., Journal of Machine Learning Research 9 pp 2039– (2008)

[4] DOI: 10.1023/A:1010933404324 · Zbl 1007.68152 · doi:10.1023/A:1010933404324

[5] Cutler A., Computing Science and Statistics 33 pp 490– (2001)

[6] DOI: 10.1002/0471722162 · Zbl 1053.62060 · doi:10.1002/0471722162

[7] Díaz-Uriarte, R., and Alvarez de Andrés, S. (2006), 'Gene Selection and Classification of Microarray Data using Random Forest',BMC Bioinformatics, 7, 3.

[8] DOI: 10.1016/j.patrec.2010.03.014 · doi:10.1016/j.patrec.2010.03.014

[9] DOI: 10.1186/1471-2156-11-49 · doi:10.1186/1471-2156-11-49

[10] Graham R. L., Concrete Mathematics (1989)

[11] DOI: 10.1007/978-0-387-84858-7 · Zbl 1273.62005 · doi:10.1007/978-0-387-84858-7

[12] Ibragimov I. A., Statistical Estimation: Asymptotic Theory (1981) · doi:10.1007/978-1-4899-0027-2

[13] Liaw A., R News 2 (3) pp 18– (2002)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.