*The Annals of Statistics*
2004, Vol. 32, No. 1, 12


# THREE PAPERS ON BOOSTING: AN INTRODUCTION

BY VLADIMIR KOLTCHINSKII AND BIN YU

*University of New Mexico and University of California, Berkeley*

The notion of boosting originated in the Machine Learning literature in the 1980's [4]. The goal of boosting is to improve the generalization performance of weak (or base) learning algorithms by combining them in a certain way. The first algorithm of this type was discovered by Schapire [3] and then the second one by Freund [1]. Schapire and Freund [2] came up with the idea of a more practical version of boosting and invented the algorithm called AdaBoost that combines simple classification rules into much more powerful and precise classification algorithms. For a fixed number of iterations, AdaBoost runs the weak (or base) learning algorithm on resampled original data sets in a sequential manner and then combines the resulting learning algorithms through a weighted summation at the end of the iteration. Gradually, it became clear that AdaBoost is a special case of a more general statistical methodology of combining simple estimates in classification or regression into more complex and more precise ones. The study of statistical properties of these methods has been conducted in several directions since then in both the machine learning and statistics communities.

The problem of consistency of AdaBoost is posed by Leo Breiman in the first paper in this issue of *The Annals of Statistics.* Breiman studies one ingredient needed to prove the consistency, the convergence properties of AdaBoost as a numerical method in the population case. This paper has been circulated for a couple of years as a preprint and its results were also covered in the Wald Lectures delivered by Breiman at the IMS Annual Meeting in 2002 in Banff, Canada. The papers by Jiang, Lugosi and Vayatis, and Zhang, published below with discussions, consider various versions of boosting and give answers to the consistency question posed by Breiman.

## REFERENCES

[1] FREUND, Y. (1995). Boosting a weak learning algorithm by majority. *Inform. and Comput.* **121** 256–285.

[2] FREUND, Y. and SCHAPIRE, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. System Sci.* **55** 119–139.

[3] SCHAPIRE, R. E. (1990). The strength of weak learnability. *Machine Learning* **5** 197–227.

[4] VALIANT, L. G. (1984). A theory of the learnable. In *Proc. 16th Annual ACM Symposium on Theory of Computing* 436–445. ACM Press, New York.

DEPARTMENT OF MATHEMATICS AND STATISTICS
UNIVERSITY OF NEW MEXICO
ALBUQUERQUE, NEW MEXICO 87131-1141
USA
E-MAIL: vlad@math.unm.edu

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720-3860
USA
E-MAIL: binyu@stat.berkeley.edu

Received May 2003; revised May 2003.