Electronic Journal of Statistics > Vol. 2 (2008) open journal systems

Classification with minimax fast rates for classes of Bayes rules with sparse representation

Guillaume Lecué, *LPMA, Paris 6*

Abstract

We consider the classification problem on the cube $[0,1]^d$ when the Bayes rule is known to belong to some new functions classes. These classes are made of prediction rules satisfying some conditions regarding their coefficients when developed over the (overcomplete) basis of indicator functions of dyadic cubes of $[0,1]^d$. The main concern of the paper is on the thorough analysis of the approximation term, which is in general bypassed in the classification literature. An adaptive classifier is designed to achieve the minimax rate of convergence (up to a logarithmic factor) over these functions classes. Lower bounds on the convergence rate over these classes are established when the underlying marginal of the design is comparable to the Lebesgue measure. Connections with some existing models for classification (RKHS and ``boundary fragements'') are established.

AMS 2000 subject classifications: Primary 62G05; secondary 62C20.

Keywords: Classification, Sparsity, Decision dyadic trees, Minimax rates, Aggregation.

Full Text: PDF

Lecué, Guillaume, Classification with minimax fast rates for classes of Bayes rules with sparse representation, Electronic Journal of Statistics, 2, (2008), 741-773 (electronic). DOI: 10.1214/07-EJS015.

References

A. Antos, L. Devroye, and L.Györfi. Lower bounds for Bayes error estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21:643–645, 1999.

N. Aronszajn. Theory of reproducing kernels. Trans. Am. Math. Soc., 68: 337–404, 1950. MR0051437

P. Assouad. Deux remarques sur l'estimation. C. R. Acad. Sci. Paris Sér. I Math., 296 (23):1021–1024, 1983. French. MR0777600

G. Blanchard, G. Lugosi, and N. Vayatis. On the rate of convergence of regularized boosting classifiers. Journal of Machine Learning Research, 4: 861–894, 2003. MR2076000

G. Blanchard, C. Schäfer, Y. Rozenholc, and K-R. Müller. Optimal dyadic decision trees. Machine Learning, 66(2-3):209–242, 2007.

S. Boucheron, O. Bousquet, and G. Lugosi. Theory of classification: a survey of some recent advances. ESAIM: Probability and Statistics, 9:323–375, 2005. MR2182250

L. Breiman, J. Freidman, J. Olshen, and C. Stone. Classification and regression trees. Wadsworth, 1984.

T. M. Cover and J. A. Thomas. Elements of Information Theory. 1991. Second edition,

2006. MR2239987

L. Devroye, L. Györfi, and G. Lugosi. A Probabilistic Theory of Pattern Recognition. Springer, New York, Berlin, Heidelberg, 1996. MR1383093

A.P. Korostelev and A.B. Tsybakov. Minimax Theory of Image Reconstruction, volume 82 of Lecture Notes in Statistics. NY e.a., 1993. MR1226450

G. Lecué. Optimal oracle inequality for aggregation of classifiers under low noise condition. In Proceeding of the 19th Annual Conference on Learning Theory, COLT 2006, 32(4):364–378, 2006. MR2280618

G. Lecué. Optimal rates of aggregation in classification under low noise assumption. Bernoulli, 13(4):1000–1022, 2007. MR2364224

G. Lugosi and N. Vayatis. On the bayes-risk consistency of regularized boosting methods. Ann. Statist., 32(1):30–55, 2004. MR2051000

E. Mammen and A.B. Tsybakov. Smooth discrimination analysis. Ann. Statist., 27:1808–1829, 1999. MR1765618

P. Massart and E. Nédélec. Risk Bound for Statistical Learning. Ann. Statist., 34(5), 2006.

Y. Meyer. Ondelettes et Opérateurs. Hermann, Paris, 1990. MR1085487

S. Murthy. Automatic construction of decision trees from data: A multi-disciplinary survey. Data Mining and knowledge Discovery, 2(4): 345–389, 1998.

J.R. Quinlan. C4.5: Programs for Machine Learning. Morgan Kaufmann, San Mateo, 1993.

C. Scott and R. Nowak. Minimax-optimal classification with dyadic decision trees. IEEE Transactions on Information Theory, 52(4):1335–1353, April 2006. MR2241192

I. Steinwart, D. Hush, and C. Scovel. Function classes that approximate the Bayes risk. In Proceeding of the 19th Annual Conference on Learning Theory, COLT 2006, 32(4):79–93, 2006. MR2277920

I. Steinwart and C. Scovel. Fast Rates for Support Vector Machines using Gaussian Kernels. Ann. Statist., 35(2), April 2007. MR2336860

R. Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series BB 58, pages 267–288, 1996. MR1379242

A.B. Tsybakov. Optimal aggregation of classifiers in statistical learning. Ann. Statist., 32 (1):135–166, 2004. MR2051002

A.B. Tsybakov and S.A. van de Geer. Square root penalty: adaptation to the margin in classification and in edge estimation. Ann. Statist., 33:1203–1224, 2005. MR2195633

Y. Yang. Minimax nonparametric classification—part I: Rates of convergence. IEEE Transaction on Information Theory, 45:2271–2284, 1999a. MR1725115

Y. Yang. Minimax nonparametric classification—partII: Model selection for adaptation. IEEE Transaction on Information Theory, 45:2285–2292, 1999b. MR1725116