Electronic Journal of Statistics > Vol. 3 (2009) open journal systems

Penalized empirical risk minimization over Besov spaces

Sébastien Loustau, *Université de Provence, LATP*

Abstract

Kernel methods are closely related to the notion of reproducing kernel Hilbert space (RKHS). A kernel machine is based on the minimization of an empirical cost and a stabilizer (usually the norm in the RKHS). In this paper we propose to use Besov spaces as alternative hypothesis spaces. We study statistical performances of a penalized empirical risk minimization for classification where the stabilizer is a Besov norm. More precisely, we state fast rates of convergence to the Bayes rule. These rates are adaptive with respect to the regularity of the Bayes.

Keywords:

Full Text: PDF

Loustau, Sébastien, Penalized empirical risk minimization over Besov spaces, Electronic Journal of Statistics, 3, (2009), 824-850 (electronic). DOI: 10.1214/08-EJS316.

References

[1] J.Y. Audibert and A.B. Tsybakov. Fast learning rates for plug-in classifiers. The Annals of Statistics, 35 (2):608–633, 2007. MR2336861

[2] P.L. Bartlett, S. Boucheron, and G. Lugosi. Model selection and error estimation. Machine Learning, 48:85–113, 2002.

[3] P.L. Bartlett, O. Bousquet, and S. Mendelson. Local rademacher complexities. The Annals of Statistics, 33 (4):1497–1537, 2005. MR2166554

[4] P.L. Bartlett, M.I. Jordan, and J.D. McAuliffe. Convexity, classification, and risk bounds. J. Amer. Statist. Assoc., 101 (473):138–156, 2006. MR2268032

[5] P.L. Bartlett and S. Mendelson. Rademacher and Gaussian complexities: risk bounds and structural results. Journal of Machine Learning Research, 3:463–482, 2002. MR1984026

[6] G. Blanchard, O. Bousquet, and P. Massart. Statistical performance of support vector machines. Annals of Statistics, 36 (2), 2008. MR2396805

[7] 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

[8] B.E. Boser, I. Guyon, and V. Vapnik. A training algorithm for optimal margin classifiers. In Computational Learning Theory, pages 144–152, 1992.

[9] S. Canu, X. Mary, and A. Rakotomamonjy. Functional learning through kernel. Advances in Learning Theory: Methods, Models and Applications, 190:89–110, 2003.

[10] I. Daubechies. Orthonormal bases of compactly supported wavelets. Communications on Pure and Applied Mathematics, 41 (7):909–996, 1988. MR0951745

[11] L. Devroye, L. Györfi, and G. Lugosi. A Probabilistic Theory of Pattern Recognition. Springer-Verlag, 1996. MR1383093

[12] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Density estimation by wavelet thresholding. The Annals of Statistics, 24 (2):508–539, 1996. MR1394974

[13] W. Härdle, G. Kerkyacharian, D. Picard, and A. Tsybakov. Wavelets, Approximation, and Statistical Applications. Lecture Notes in Statistics, 1997. MR1618204

[14] V. Koltchinskii. Rademacher penalties and structural risk minimization. IEEE Transactions on Information Theory, 47 (5):1902–1914, 2001. MR1842526

[15] G. Lecué. Classification with minimax fast rates for classes of Bayes rules with sparse representation. Electronic Journal of Statistics, 2:741–773, 2008. MR2430253

[16] S. Loustau. Aggregation of SVM classifiers using Sobolev spaces. Journal of Machine Learning Research, 9:1559–1582, 2008. MR2426051

[17] S. Mallat. Une exploration des signaux en ondelettes. Ellipses, 2000.

[18] X. Mary, D. De Brucq, and S. Canu. Sou-dualités et noyaux (reproduisants) associés. C. R. Acad. Sc. Paris, 336 (1):949–954, 2003. MR1994600

[19] P. Massart and E. Nédélec. Risk bounds for statistical learning. The Annals of Statistics, 34 (5):2326–2366, 2006. MR2291502

[20] S. Mendelson. On the performance of kernel classes. Journal of Machine Learning Research, 4:759–771, 2003. MR2075996

[21] Y. Meyer. Ondelettes et Opérateurs 1: Ondelettes. Hermann, 1990. MR1085487

[22] J. Peetre. New thoughts on Besov spaces. Mathematics Department, Duke University, Durham, N.C., 1976. MR0461123

[23] H.P. Rosenthal. On the span in l_p of sequences of independent random variables. Israël J. Math., 8:273–303, 1972. MR0271721

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

[25] S. Smale and D.X. Zhou. Estimating the approximation error in learning theory. Analysis and Applications, 1 (1):17–41, 2003.

MR1959283

[26] I. Steinwart and C. Scovel. Fast rates for support vector machines using Gaussian kernels. The Annals of Statistics, 35 (2):575–607, 2007. MR2336860

[27] H. Triebel. Interpolation Theory, Function Spaces, Differential Operators. North-Holland Publishing Company, 1978. MR0503903

[28] H. Triebel. Theory of Functions Spaces II. Birkhauser, 1992. MR1163193

[29] A.B. Tsybakov. Optimal aggregation of classifiers in statistical learning. The Annals of Statistics, 32 (1):135–166, 2004. MR2051002

[30] V.N. Vapnik and A.Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and its Applications, 16 (2):264–280, 1971.

[31] V.N. Vapnik and A.Ya. Chervonenkis. Theory of Pattern Recognition. Nauka, Moscow, 1974. MR0474638

Electronic Journal of Statistics. ISSN: 1935-7524