Bibliography

Alon, N., Ben-David, S., Cesa-Bianchi, N. and Haussler, D. (1997). Scale sensitive dimensions, uniform convergence and learnability. *J. ACM* **44** 615–631.

Audibert, J.-Y. (2004a). Aggregated estimators and empirical complexity for least square regression. *Ann. Inst. H. Poincaré Probab. Statist.* **40** 685–736.

Audibert, J.-Y. (2004b). PAC-Bayesian statistical learning theory. Ph.D. thesis, Univ. Paris 6. Available at http://cermics.enpc.fr/ audibert/.

Barron, A. (1987). Are Bayes rules consistent in information? In *Open Problems in Communication and Computation* (T. M. Cover and B. Gopinath, eds.) 85–91. Springer, New York.

Barron, A. and Yang, Y. (1999). Information-theoretic determination of minimax rates of convergence. *Ann. Statist.* **27** 1564–1599.

Barron, A., Birgé, L. and Massart, P. (1999). Risk bounds for model selection by penalization. *Probab. Theory Related Fields* **113** 301–413.

Blanchard, G. (1999). The "progressive mixture" estimator for regression trees. *Ann. Inst. H. Poincaré Probab. Statist.* **35** 793–820.

Blanchard, G. (2001). Mixture and aggregation of estimators for pattern recognition. Application to decision trees. Ph.D. thesis, Univ. Paris 13. Available at http://ida.first.fraunhofer.de/ blanchard/.

Blanchard, G. (2004). Un algorithme accéléré d'échantillonnage Bayésien pour le modèle CART. *Rev. Intell. Artificielle* **18** 383–410.

Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam* (D. Pollard, ed.) 55–87. Springer, New York.

Birgé, L. and Massart, P. (1998). Minimum contrast estimators on sieves. *Bernoulli* **4** 329–375.

Birgé, L. and Massart, P. (2001a). A generalized C_p criterion for Gaussian model selection. Preprint. Available at http://www.math.u-psud.fr/ massart/.

Birgé, L. and Massart, P. (2001b). Gaussian model selection. *J. Eur. Math. Soc.* **3** 203–268.

Blum, A. and Langford, J. (2003). PAC-MDL bounds. *Computational Learning Theory and Kernel Machines. 16th Annual Conference on Computational Learning Theory and 7th Kernel Workshop. COLT/Kernel 2003, Washington, DC, USA, August 24-27, 2003, Proceedings. Lecture Notes in Comput. Sci.* **2777** 344–357. Springer, New York.

Catoni, O. (2002). Data compression and adaptive histograms. In *Foundations of Computational Mathematics. Proceedings of the Smalefest 2000* (F. Cucker and J. M. Rojas eds.) 35–60. World Scientific.

Catoni, O. (2003). Laplace transform estimates and deviation inequalities. *Ann. Inst. H. Poincaré Probab. Statist.* **39** 1–26.

Catoni, O. (2004). Statistical learning theory and stochastic optimization. *Ecole*