Lasso type classifiers with a reject option

Marten H. Wegkamp, *FSU*

Abstract
We consider the problem of binary classification where one can, for a particular cost, choose not to classify an observation. We present a simple proof for the oracle inequality for the excess risk of structural risk minimizers using a lasso type penalty.

AMS 2000 subject classifications: Primary 62C05; secondary 62G05, 62G08.

Keywords: Bayes classifiers, classification, convex surrogate loss, empirical risk minimization, hinge loss, large margin classifiers, ℓ_1 penalties, local mutual coherence, margin condition, reject option, support vector machines.

Full Text: PDF

References

[1]   P.L. Bartlett and M.H. Wegkamp (2007). Classification with a reject option using a hinge loss. Technical report M980, Department of Statistics, Florida State University.

[2]   S. Boucheron, O. Bousquet and G. Lugosi (2004). Introduction to statistical learning theory. In Advanced Lectures in Machine Learning (O. Bousquet, U. von Luxburg, and G. Rätsch, Editors), 169–207. Springer, New-York.

[3]   S. Boucheron, O. Bousquet and G. Lugosi (2005). Theory of Classification: a Survey of Recent Advances. ESAIM: Probability and Statistics, 9:323-375. MR2182250

[4]   F. Bunea, A.B. Tsybakov and M.H. Wegkamp (2004). Aggregation for Gaussian regression. Technical Report M972, Department of Statistics, Florida State University. To appear in the Annals of Statistics.

[5]   F. Bunea, A.B. Tsybakov and M.H. Wegkamp (2006). Aggregation and sparsity via ℓ_1-penalized least squares. Proceedings of 19th Annual Conference on Learning Theory, COLT 2006. Lecture Notes in Artificial Intelligence, 4005, 379–391. Springer-Verlag, Heidelberg. MR2280619

[6]   F. Bunea, A.B. Tsybakov and M.H. Wegkamp (2006). Sparsity oracle inequalities for the Lasso. Technical Report M979, Department of Statistics, Florida State University.

[7]   F. Bunea, A.B. Tsybakov and M.H. Wegkamp (2007). Sparse density estimation with ℓ_1 penalties. Proceedings of 20th Annual Conference on Learning Theory, COLT 2007. Lecture Notes in Artificial Intelligence, 4539, 530–543. Springer-Verlag, Heidelberg.

[8]   F. Bunea and M.H. Wegkamp (2004). Two-stage model selection procedures in partially linear regression. Canadian Journal of Statistics, 32(2), 105–118. MR2064395

[9]   C.K. Chow (1970). On optimum error and reject trade-off. IEEE Transactions on

Information Theory, 16, 41–46.

[10]   L. Devroye and G. Lugosi (2000). Combinatorial Methods in density estimation. Springer, New York. MR1843146

[11]   R. Herbei and M. H. Wegkamp (2006). Classification with reject option. Canadian Journal of Statistics, 34(4), 709–721.

[12]   M. Ledoux and M. Talagrand (1991). Probability in Banach Spaces. Springer, New York. MR1102015

[13]   B. D. Ripley (1996). Pattern recognition and neural networks. Cambridge University Press, Cambridge. MR1438788

[14]   B. Tarigan and S. A. van de Geer (2006). Classifiers of support vector machine type with ℓ_1 complexity regularization. Bernoulli, 12(6), 1045–1076 MR2274857

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

[16]   S.A. van de Geer (2000). Empirical processes in M-estimation. Cambridge Series in statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

[17]   S.A. van de Geer (2006). High dimensional generalized linear models and the Lasso. Research report No.133. Seminar für Statistik, ETH.