Generalization error for multi-class margin classification

Xiaotong Shen, *University of Minnesota*
Lifeng Wang, *University of Pennsylvania*

Abstract

In this article, we study rates of convergence of the generalization error of multi-class margin classifiers. In particular, we develop an upper bound theory quantifying the generalization error of various large margin classifiers. The theory permits a treatment of general margin losses, convex or nonconvex, in presence or absence of a dominating class. Three main results are established. First, for any fixed margin loss, there may be a trade-off between the ideal and actual generalization performances with respect to the choice of the class of candidate decision functions, which is governed by the trade-off between the approximation and estimation errors. In fact, different margin losses lead to different ideal or actual performances in specific cases. Second, we demonstrate, in a problem of linear learning, that the convergence rate can be arbitrarily fast in the sample size n depending on the joint distribution of the input/output pair. This goes beyond the anticipated rate $O(n^{-1})$. Third, we establish rates of convergence of several margin classifiers in feature selection with the number of candidate variables p allowed to greatly exceed the sample size n but no faster than $exp(n)$.

AMS 2000 subject classifications: Primary 68T10, 62H30.

Keywords: Convex and nonconvex losses, import vector machines, psi-learning, small n and large p, sparse learning, support vector machines.

Full Text: PDF

Shen, Xiaotong, Wang, Lifeng, Generalization error for multi-class margin classification, Electronic Journal of Statistics, 1, (2007), 307-330 (electronic). DOI: 10.1214/07-EJS069.

References

[1] Alexander, K. (1984) Probability inequalities for empirical processes and a law of the iterated logarithm The Annals of Probability, 12, 1041-1067. MR0757769

[2] Bartlett, P., and Shawe-Taylor, J. (1999). Generalization performance of support vector machines and other pattern classifiers. In B. Scholkopf, et. al (ed.) Advances in Kernel Methods - Support Vector Learning, 43-54, MIT Press.

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

[4] Bradley, P. S. and Mangasarian, O. L. (1998) Feature selection via concave minimization and support vector machines. in J. Shaclik (ed.), Machine Learning Proc the Fifteenth International Conf., Morgan Kaufmann, San Francisco, Ca, 82-90.

[5] Bredensteiner, E. J. and Bennett, K. P. (1999) Multicategory classification by support vector machines. Comput. Optim. and Appli., 12, 35-46. MR1704101

[6] Corduneanu, A. and Jaakkola, T. (2003) On information regularization. Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence.

[7] Cortes, C., and Vapnik, V. (1995). Support-vector networks. Machine Learning. 20, 73-297.

[8] Crammer, K., and Singer, Y. (2001). On the algorithmic implementation of multi-class kernel-based vector machines. J. Machine Learning Res., 2, 265-292.

[9] Cucker, F. and Smale, S. (2002) On the Mathematical foundations of learning. Bulletin Amer. Math. Soc., 39, 1-49. MR1864085

[10] Gu, C. (2000). Multidimension smoothing with splines. Smoothing and Regression: Approaches, Computation and Application, edited by M.G. Schimek. MR1795148

[11] Guermeur, Y. (2002). Combining discriminant models with new multi-class SVMs. Pattern Analy. Appli., 5, 168-179. MR1922450

[12] Lafferty, J. and Wasserman, L. (2006) Challenges in Statistical Machine Learning. Statistica Sinica, 16, 307-323. MR2267237

[13] Lee, Y., Lin, Y. and Wahba, G. (2004) Multicategory support vector machines, theory, and application to the classification of microarray data and satellite radiance data. J. Amer. Statist. Assoc., 99, 67-81. MR2054287

[14] Lin, Y. (2000). Some asymptotic properties of the support vector machine. Technical report 1029, Department of Statistics, University of Wisconsin-Madison.

[15] Lin, Y. (2002) A Note on margin-based loss functions in classification. Statistics and Probability Letters, 68, 73-82. MR2064687

[16] Liu, Y. and Shen, X. (2006). On multicategory ψ-learning and support vector machine. J. Amer. Statist. Assoc., 101, 500-509. MR2256170

[17] Mercer, J. (1909). Functions of positive and negative type and their connection with the theory of integral equations. Philosophical Transactions of the Royal Society, London A, 209 415-446.

[18] Pollard, D. (1984). Convergence of Stochastic Processes. Springer-Verlag, New York. MR0762984

[19] Shen, X., Tseng, G., Zhang, X. and Wong, W. (2003) On ψ-Learning. J. Amer. Statist. Assoc., 98, 724-734. MR2011686

[20] Shen, X. and Wong, W. (1994). Convergence rate of sieve estimates. Ann. Statist., 22, 580-615. MR1292531

[21] Tewari, A., and Bartlett, P. (2005). On the consistency of multiclass classification methods. In Proceedings of the 18th Annual Conference on Learning Theory, 3559, 143-157. Springer. MR2203259

[22] Tarigan, B., and van de Geer, SA (2004). Adaptivity of support vector machines with l_1 penalty. Technical Report MI 2004-14, University of Leiden.

[23] Vapnik, V. (1998). Statistical Learning Theory, Wiley, New York. MR1641250

[24] Van De Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. Ann. of Statist., 21,14-44. MR1212164

[25] Van der Vaart, A. W. and Wellner, J. A. (2000) Weak Convergence and Empirical Processes with Application to Statistics, Springer, New York.

[26] Weston, J., and Watkins, C. (1999). Support vector machines for multi-class pattern recognition. Proceedings of the Seventh European Symposium On Artificial Neural Networks.

[27] Wang, L., Zhu, J. and Zou, H. (2006) The doubly regularized support vector

machine. Statistica Sinica, 16, 617-634. MR2267251

[28] Wang, LF., and Shen, X. (2007). On L_1-norm multi-class support vector machines: methodology and theory. J. Amer. Statist. Asso, 102, 595-602.

[29] Zhang, T. (2004a). Statistical behavior and consistency of classification methods based on convex risk minimization. Ann. Statist., 32, 56-85. MR2051001

[30] Zhang, T. (2004b). Statistical analysis of some multi-category large Margin classification Methods. J. Machine Learning Res., 5, 1225-1251.

[31] Zou, H. and Hastie, T. (2005). Regularization and Variable Selection via the Elastic Net. J. R. Statist. Soc. B, 67, 301-320. MR2137327

[32] Zou, H., Zhu, J. and Hastie, T. (2005) The margin vector, admissible losses and multi-class margin-based classifiers. Technical Report, University of Minnesota.

[33] Zhu, J. and Hastie, T. (2005) Kernel logistic regression and the import vector machine. J. Comput. and Graph. Statist., 14, 185-205. MR2137897

Electronic Journal of Statistics. ISSN: 1935-7524