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

Variable selection for multicategory SVM via adaptive sup-norm regularization

Hao Helen Zhang, *North Carolina State University*
Yufeng Liu, *University of North Carolina*
Yichao Wu, *Princeton University*
Ji Zhu, *University of Michigan*

Abstract

Support Vector Machine (SVM) is a popular classification paradigm in machine learning and has achieved great success in real applications. However, the standard SVM can not select variables automatically and therefore its solution typically utilizes all the input variables without discrimination. This makes it difficult to identify important predictor variables, which is often one of the primary goals in data analysis. In this paper, we propose two novel types of regularization in the context of the multicategory SVM (MSVM) for simultaneous classification and variable selection. The MSVM generally requires estimation of multiple discriminating functions and applies the argmax rule for prediction. For each individual variable, we propose to characterize its importance by the supnorm of its coefficient vector associated with different functions, and then minimize the MSVM hinge loss function subject to a penalty on the sum of supnorms. To further improve the supnorm penalty, we propose the adaptive regularization, which allows different weights imposed on different variables according to their relative importance. Both types of regularization automate variable selection in the process of building classifiers, and lead to sparse multiclassifiers with enhanced interpretability and improved accuracy, especially for high dimensional low sample size data. One big advantage of the supnorm penalty is its easy implementation via standard linear programming. Numerious examples and one real gene data analysis demonstrate the outstanding performance of the adaptive supnorm penalty in various data settings.

AMS 2000 subject classifications: Primary 62H30.

Keywords: Classification, L_1-norm penalty, multicategory, sup-norm, SVM.

Full Text: PDF

Zhang, Hao Helen, Liu, Yufeng, Wu, Yichao, Zhu, Ji, Variable selection for multicategory SVM via adaptive sup-norm regularization, Electronic Journal of Statistics, 2, (2008), 149-167 (electronic). DOI: 10.1214/08-EJS122.

References

Argyriou, A., Evgeniou, T. and M., P. (2006). Multi-task feature learning. Neural Information Processing Systems, 19.

Argyriou, A., Evgeniou, T. and M., P. (2007). Convex multi-task feature learning. Machine Learning. To appear.

Boser, B. E., Guyon, I. M. and Vapnik, V. (1992). A training algorithm for optimal margin classifiers. In Fifth Annual ACM Workshop on Computational Learning Theory. ACM Press, Pittsburgh, PA, 144–152.

Bradley, P. S. and Mangasarian, O. L. (1998). Feature selection via concave minimization and support vector machines. In Proc. 15th International Conf. on Machine Learning. Morgan Kaufmann, San Francisco, CA, 82–90.

Christianini, N. and Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press.

Crammer, K. and Singer, Y. (2001). On the algorithmic implementation of multiclass kernel-based vector machines. Journal of Machine Learning Research, 2 265–292.

Dudoit, S., Fridlyand, J. and Speed, T. (2002). Comparison of discrimination methods for the classification of tumors using gene expression data. Journal of American Statistical Association, 97 77–87. MR1963389

Fourer, R., Gay, D. and Kernighan, B. (2003). AMPL: A Modeling Language for Mathematical Programming. Duxbury Press.

Khan, J., Wei, J. S., Ringnér, M., , Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C. and Meltzer, P. S. (2001). Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nature Medicine, 7 673–679.

Lafferty, J. and Wasserman, L. (2006). Challenges in statistical machine learning. Statistica Sinica, 16 307–323. MR2267237

Lee, Y., Kim, Y., Lee, S. and Koo, J.-Y. (2006). Structured multicategory support vector machine with anova decomposition. Biometrika, 93 555–571. MR2261442

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. Journal of American Statistical Association, 99 67–81. MR2054287

Liu, Y. and Shen, X. (2006). Multicategory ψ-learning. Journal of the American Statistical Association, 101 500–509. MR2256170

Liu, Y. and Wu, Y. (2007). Variable selection via a combination of the l_0 and l_1 penalties. Journal of Computation and Graphical Statistics, 16 782–798.

Liu, Y., Zhang, H. H., Park, C. and Ahn, J. (2007). Support vector machines with adaptive l_q penalties. Computational Statistics and Data Analysis, 51 6380–6394.

Micchelli, C. and Pontil, M. (2007). Feature space perspectives for learning the kernel. Machine Learning, 66 297–319.

Schölkopf, B. and Smola, A. J. (2002). Learning with Kernels. MIT Press. MR1949972

Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer-Verlag, New York. MR1367965

Vapnik, V. (1998). Statistical learning theory. Wiley. MR1641250

Wang, H., Li, G. and Jiang, G. (2007). Robust regression shrinkage and consistent variable selection via the lad-lasso. Journal of Business and Economics Statistics, 25 347–355.

Wang, J. and Shen, X. (2007a). Large margin semi-supervised learning. Journal of Machine Learning Research, 8 1867–1891.

Wang, L. and Shen, X. (2007b). On l_1-norm multi-class support vector machines: methodology and theory. Journal of the American Statistical Association, 102 583–594.

Wang, L., Zhu, J. and Zou, H. (2006). The doubly regularized support vector machine. Statistica Sinica, 16 589–615. MR2267251

Weston, J., Elisseeff, A., Schölkopf, B. and Tipping, M. (2003). Use of the zero-norm with linear models and kernel methods. Journal of Machine Learning Research, 3 1439–1461. MR2020766

Weston, J. and Watkins, C. (1999). Multiclass support vector machines. In Proceedings of ESANN99 (M. Verleysen, ed.). D. Facto Press.

Wu, Y. and Liu, Y. (2007a). Robust truncated-hinge-loss support vector machines. Journal of the American Statistical Association, 102 974–983.

Wu, Y. and Liu, Y. (2007b). Variable selection in quantile regression. Statistica Sinica. To appear.

Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society, Series B, 68 49–67. MR2212574

Zhang, H. H., Ahn, J., Lin, X. and Park, C. (2006). Gene selection using support vector machines with nonconvex penalty. Bioinformatics, 22 88–95.

Zhang, H. H. and Lu, W. (2007). Adaptive-lasso for cox's proportional hazard model. Biometrika, 94 691–703.

Zhao, P., Rocha, G. and Yu, B. (2006). Grouped and hierarchical model selection through composite absolute penalties. Technical Report 703, Department of Statistics University of California at Berkeley.

Zhu, J., Hastie, T., Rosset, S. and Tibshirani, R. (2003). 1-norm support vector machines. Neural Information Processing Systems, 16.

Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association, 101 1418–1429. MR2279469

Zou, H. and Yuan, M. (2006). The f_∞-norm support vector machine. Statistica Sinica. To appear.