Huang, Jian; **Horowitz, Joel L.**; **Wei, Fengrong**

Variable selection in nonparametric additive models. (English) Zbl 1202.62051

Ann. Stat. 38, No. 4, 2282-2313 (2010).

Summary: We consider a nonparametric additive model of a conditional mean function in which the number of variables and additive components may be larger than the sample size but the number of nonzero additive components is "small" relative to the sample size. The statistical problem is to determine which additive components are nonzero. The additive components are approximated by truncated series expansions with B-spline bases. With this approximation, the problem of component selection becomes that of selecting the groups of coefficients in the expansion. We apply the adaptive group Lasso to select nonzero components, using the group Lasso to obtain an initial estimator and reduce the dimension of the problem. We give conditions under which the group Lasso selects a model whose number of components is comparable with the underlying model, and the adaptive group Lasso selects the nonzero components correctly with probability approaching one as the sample size increases and achieves the optimal rate of convergence. The results of Monte Carlo experiments show that the adaptive group Lasso procedure works well with samples of moderate size. A data example is used to illustrate the application of the proposed method.

MSC:

62G08 Nonparametric regression and quantile regression
62G20 Asymptotic properties of nonparametric inference
65D07 Numerical computation using splines
65C05 Monte Carlo methods

Keywords:

adaptive group Lasso; component selection; high-dimensional data; nonparametric regression; selection consistency

Software:

hgam

Full Text: DOI arXiv

References:

[1] Antoniadis, A. and Fan, J. (2001). Regularization of wavelet approximation (with discussion). J. Amer. Statist. Assoc. 96 939-967. JSTOR: · Zbl 1072.62561 · doi:10.1198/016214501753208942 · links.jstor.org

[2] Bach, F. R. (2007). Consistency of the group Lasso and multiple kernel learning. J. Mach. Learn. Res. 9 1179-1225. · Zbl 1225.68147 · www.jmlr.org

[3] Bunea, F., Tsybakov, A. and Wegkamp, M. (2007). Sparsity oracle inequalities for the Lasso. Electron. J. Stat. 169-194. · Zbl 1146.62028 · doi:10.1214/07-EJS008 ·

[4] Chen, J. and Chen, Z. (2008). Extended Bayesian information criteria for model selection with large model space. Biometrika 95 759-771. · Zbl 1437.62415 · doi:10.1093/biomet/asn034

[5] Chen, J. and Chen, Z. (2009). Extended BIC for small- n -large- P sparse GLM. Available at http:// www.stat.nus.edu.sg/~stachenz/ChenChen.pdf. · Zbl 1238.62080

[6] Chiang, A. P., Beck, J. S., Yen, H.-J., Tayeh, M. K., Scheetz, T. E., Swiderski, R., Nishimura, D., Braun, T. A., Kim, K.-Y., Huang, J., Elbedour, K., Carmi, R., Slusarski, D. C., Casavant, T. L., Stone, E. M. and Sheffield, V. C. (2006). Homozygosity mapping with SNP arrays identifies a novel gene for Bardet-Biedl syndrome (BBS10). Proc. Natl. Acad. Sci. USA 103 6287-6292.

[7] de Boor, C. (2001). A Practical Guide to Splines , revised ed. Springer, New York. · Zbl 0987.65015

[8] Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression (with discussion). Ann. Statist. 32 407-499. · Zbl 1091.62054 · doi:10.1214/009053604000000067 ·

[9] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist.

Assoc. 96 1348-1360. JSTOR: · Zbl 1073.62547 · doi:10.1198/016214501753382273 · links.jstor.org

[10] Fan, J. and Peng, H. (2004). Nonconcave penalized likelihood with a diverging number of parameters. Ann. Statist. 32 928-961. · Zbl 1092.62031 · doi:10.1214/009053604000000256 ·

[11] Frank, I. E. and Friedman, J. H. (1993). A statistical view of some chemometrics regression tools (with discussion). Technometrics 35 109-148. · Zbl 0775.62288 · doi:10.2307/1269656

[12] Horowitz, J. L., Klemelä, J. and Mammen, E. (2006). Optimal estimation in additive regression models. Bernoulli 12 271-298. · Zbl 1098.62043 · doi:10.3150/bj/1145993975 · euclid:bj/1145993975

[13] Horowitz, J. L. and Mammen, E. (2004). Nonparametric estimation of an additive model with a link function. Ann. Statist. 32 2412-2443. · Zbl 1069.62035 · doi:10.1214/009053604000000814 ·

[14] Huang, J., Horowitz, J. L. and Ma, S. G. (2008). Asymptotic properties of bridge estimators in sparse high-dimensional regression models. Ann. Statist. 36 587-613. · Zbl 1133.62048 · doi:10.1214/009053607000000875 ·

[15] Huang, J., Ma, S. and Zhang, C.-H. (2008). Adaptive Lasso for high-dimensional regression models. Statist. Sinica 18 1603-1618. · Zbl 1255.62198

[16] Irizarry, R. A., Hobbs, B., Collin, F., Beazer-Barclay, Y. D., Antonellis, K. J., Scherf, U. and Speed, T. P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics 4 249-264. · Zbl 1141.62348 · doi:10.1093/biostatistics/4.2.249

[17] Lin, Y. and Zhang, H. (2006). Component selection and smoothing in multivariate nonparametric regression. Ann. Statist. 34 2272-2297. · Zbl 1106.62041 · doi:10.1214/009053606000000722 ·

[18] Meier, L., van de Geer, S. and Bühlmann, P. (2009). High-dimensional additive modeling. Ann. Statist. 37 3779-3821. · Zbl 1360.62186

[19] Meinshausen, N. and Bühlmann, P. (2006). High dimensional graphs and variable selection with the Lasso. Ann. Statist. 34 1436-1462. · Zbl 1113.62082 · doi:10.1214/009053606000000281

[20] Meinshausen, N. and Yu, B. (2009). Lasso-type recovery of sparse representations for high-dimensional data. Ann. Statist. 37 246-270. · Zbl 1155.62050 · doi:10.1214/07-AOS582 · www.projecteuclid.org

[21] Ravikumar, P., Liu, H., Lafferty, J. and Wasserman, L. (2009). Sparse additive models. J. Roy. Statist. Soc. Ser. B 71 1009-1030.

[22] Scheetz, T. E., Kim, K.-Y. A., Swiderski, R. E., Philp, A. R., Braun, T. A., Knudtson, K. L., Dorrance, A. M., DiBona, G. F., Huang, J., Casavant, T. L., Sheffield, V. C. and Stone, E. M. (2006). Regulation of gene expression in the mammalian eye and its relevance to eye disease. Proc. Natl. Acad. Sci. USA 103 14429-14434.

[23] Schwarz, G. (1978). Estimating the dimension of a model. Ann. Statist. 6 461-464. · Zbl 0379.62005 · doi:10.1214/aos/1176344136 ·

[24] Schumaker, L. (1981). Spline Functions: Basic Theory . Wiley, New York. · Zbl 0449.41004

[25] Shen, X. and Wong, W. H. (1994). Convergence rate of sieve estimates. Ann. Statist. 22 580-615. · Zbl 0805.62008 · doi:10.1214/aos/1176325486 ·

[26] Stone, C. J. (1985). Additive regression and other nonparametric models. Ann. Statist. 13 689-705. · Zbl 0605.62065 · doi:10.1214/aos/1176349548 ·

[27] Stone, C. J. (1986). The dimensionality reduction principle for generalized additive models. Ann. Statist. 14 590-606. · Zbl 0603.62050 · doi:10.1214/aos/1176349940 ·

[28] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. J. Roy. Statist. Soc. Ser. B 58 267-288. JSTOR: · Zbl 0850.62538 · links.jstor.org

[29] van de Geer, S. (2008). High-dimensional generalized linear models and the Lasso. Ann. Statist. 36 614-645. · Zbl 1138.62323 · doi:10.1214/009053607000000929 ·

[30] Van der Vaart, A. W. (1998). Asymptotic Statistics . Cambridge Univ. Press, Cambridge. · Zbl 0910.62001 · doi:10.1017/CBO9780511802256

[31] van der Vaart, A. W. and Wellner, J. A. (1996). Weak Convergence and Empirical Processes: With Applications to Statistics . Springer, New York. · Zbl 0862.60002

[32] Wang, L., Chen, G. and Li, H. (2007). Group SCAD regression analysis for microarray time course gene expression data. Bioinformatics 23 1486-1494.

[33] Wang, H. and Xia, Y. (2008). Shrinkage estimation of the varying coefficient model. J. Amer. Statist. Assoc. 104 747-757. · Zbl 1388.62213 · doi:10.1198/jasa.2009.0138

[34] Wei, F. and Huang, J. (2008). Consistent group selection in high-dimensional linear regression. Technical Report #387, Dept. Statistics and Actuarial Science, Univ. Iowa. Available at http:// www.stat.uiowa.edu/techrep/tr387.pdf.

[35] Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. J. R. Stat. Soc. Ser. B Stat. Methodol. 68 49-67. · Zbl 1141.62030 · doi:10.1111/j.1467-9868.2005.00532.x

[36] Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. Ann. Statist. 38 894-942. · Zbl 1183.62120 · doi:10.1214/09-AOS729 ·

[37] Zhang, H., Wahba, G., Lin, Y., Voelker, M., Ferris, M., Klein, R. and Klein, B. (2004). Variable selection and model building via likelihood basis pursuit. J. Amer. Statist. Assoc. 99 659-672. · Zbl 1117.62459 · doi:10.1198/016214504000000593 · masetto.asa.catchword.org

[38] Zhang, C.-H. and Huang, J. (2008). The sparsity and bias of the Lasso selection in high-dimensional linear regression. Ann. Statist. 36 1567-1594. · Zbl 1142.62044 · doi:10.1214/07-AOS520 ·

[39] Zhang, H. H. and Lin, Y. (2006). Component selection and smoothing for nonparametric regression in exponential families. Statist. Sinica 16 1021-1041. · Zbl 1107.62036

[40] Zhao, P. and Yu, B. (2006). On model selection consistency of LASSO. J. Mach. Learn. Res. 7 2541-2563. · Zbl 1222.62008 · www.jmlr.org

[41] Zhou, S., Shen, X. and Wolf, D. A. (1998). Local asymptotics for regression splines and confidence regions Ann. Statist. 26 1760-1782. · Zbl 0929.62052 · doi:10.1214/aos/1024691356 ·

[42] Zou, H. (2006). The adaptive Lasso and its oracle properties. J. Amer. Statist. Assoc. 101 1418-1429. · Zbl 1171.62326 · doi:10.1198/016214506000000735