

Feng, Zhenghui; Lin, Lu; Zhu, Ruoqing; Zhu, Lixing

Nonparametric variable selection and its application to additive models. (English)

Zbl 07205440

Ann. Inst. Stat. Math. 72, No. 3, 827-854 (2020).

Summary: Variable selection for multivariate nonparametric regression models usually involves parameterized approximation for nonparametric functions in the objective function. However, this parameterized approximation often increases the number of parameters significantly, leading to the “curse of dimensionality” and inaccurate estimation. In this paper, we propose a novel and easily implemented approach to do variable selection in nonparametric models without parameterized approximation, enabling selection consistency to be achieved. The proposed method is applied to do variable selection for additive models. A two-stage procedure with selection and adaptive estimation is proposed, and the properties of this method are investigated. This two-stage algorithm is adaptive to the smoothness of the underlying components, and the estimation consistency can reach a parametric rate if the underlying model is really parametric. Simulation studies are conducted to examine the performance of the proposed method. Furthermore, a real data example is analyzed for illustration.

MSC:

- 62G08 Nonparametric regression and quantile regression
- 62P20 Applications of statistics to economics
- 62P25 Applications of statistics to social sciences

Keywords:

nonparametric regression; variable selection; nonparametric additive model; adaptive estimation

Full Text: [DOI](#)

References:

- [1] Candès, E.; Tao, T., The Dantzig selector: Statistical estimation when $\|(p)\|$ is much larger than $\|(n)\|$ (with discussion), *The Annals of Statistics*, 35, 2313-2404 (2007) · [Zbl 1139.62019](#)
- [2] Chaudhuri, P.; Huang, M-C; Loh, W-Y; Yao, R., Piecewise-polynomial regression trees, *Statistica Sinica*, 4, 143-167 (1994) · [Zbl 0824.62032](#)
- [3] Cook, R. D. (1998). *Regression graphics: Ideas for studying regressions through graphics*. New York: Wiley. · [Zbl 0903.62001](#)
- [4] Cook, RD; Weisberg, S., Discussion of ‘Sliced inverse regression for dimension reduction’, *Journal of the American Statistical Association*, 86, 28-33 (1991)
- [5] Cui, X.; Peng, H.; Wen, SQ; Zhu, LX, Component selection in the additive regression model, *Scandinavian Journal of Statistics*, 40, 3, 491-510 (2013) · [Zbl 1364.62091](#)
- [6] Fan, J.; Li, R., Variable selection via nonconcave penalized likelihood and its oracle properties, *Journal of the American Statistical Association*, 96, 1348-1360 (2001) · [Zbl 1073.62547](#)
- [7] Fan, J.; Lv, J., Sure independence screening for ultrahigh dimensional feature space (with discussion), *Journal of the Royal Statistical Society Series B*, 70, 849-911 (2008) · [Zbl 1411.62187](#)
- [8] Fan, J.; Feng, Y.; Song, R., Nonparametric independence screening in sparse ultra-high-dimensional additive models, *Journal of the American Statistical Association*, 106, 544-557 (2011) · [Zbl 1232.62064](#)
- [9] Guan, Y.; Xie, C.; Zhu, L., Sufficient dimension reduction with mixture multivariate skew elliptical distributions, *Statistica Sinica*, 27, 1, 335-355 (2017) · [Zbl 06686362](#)
- [10] Hall, P.; Li, K-C, On almost linearity of low dimensional projections from high dimensional data, *Annals of Statistics*, 21, 867-889 (1993) · [Zbl 0782.62065](#)
- [11] Härdle, W., *Applied nonparametric regression*, econometric society monograph series, 19 (1990), Cambridge: Cambridge University Press, Cambridge
- [12] Härdle, W.; Marron, JS, Optimal bandwidth selection in nonparametric regression function estimation, *Annals of Statistics*, 13, 1465-1481 (1985) · [Zbl 0594.62043](#)
- [13] Hastie, T.; Tibshirani, R., Generalized additive models, *Statistical Science*, 1, 297-318 (1986) · [Zbl 0645.62068](#)

- [14] Li, B.; Wang, S., On directional regression for dimension reduction, *Journal of the American Statistical Association*, 102, 997-1008 (2007) · [Zbl 0556.4427](#)
- [15] Li, K-C, Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, 86, 316-327 (1991) · [Zbl 0742.62044](#)
- [16] Li, K-C; Duan, N., Regression analysis under link violation, *The Annals of Statistics*, 17, 3, 1009-1052 (1989) · [Zbl 0753.62041](#)
- [17] Li, K-C; Lue, HH; Chen, CH, Interactive tree-truncated regression via principal Hessian directions, *Journal of the American Statistical Association*, 95, 547-560 (2000) · [Zbl 1013.62074](#)
- [18] Li, R.; Zhong, W.; Zhu, LP, Feature screening via distance correlation learning, *Journal of the American Statistical Association*, 107, 1129-1139 (2012) · [Zbl 1443.62184](#)
- [19] Lin, L.; Cui, X.; Zhu, LX, An adaptive two-stage estimation method for additive models, *Scandinavian Journal of Statistics*, 36, 248-269 (2009) · [Zbl 1194.62048](#)
- [20] Lin, L.; Sun, J.; Zhu, LX, Nonparametric feature screening, *Computational Statistics and Data Analysis*, 36, 162-174 (2013) · [Zbl 06970880](#)
- [21] Lin, Y.; Zhang, H., Component selection and smoothing in multivariate nonparametric regression, *The Annals of Statistics*, 34, 2272-2297 (2006) · [Zbl 1106.62041](#)
- [22] Meier, L.; Van der Geer, S.; Bühlmann, P., High-dimensional additive modeling, *The Annals of Statistics*, 37, 3779-3821 (2009) · [Zbl 1360.62186](#)
- [23] Storlie, CB; Bonedll, HD; Reich, BJ; Zhang, HH, Surface estimation, variance selection, and the nonparametric oracle property, *Statistica Sinica*, 21, 679-705 (2011) · [Zbl 1214.62044](#)
- [24] Tibshirani, R., Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society, Series B*, 58, 267-288 (1996) · [Zbl 0850.62538](#)
- [25] Wahba, G. (1990). *Spline models for observational data*, vol. 59. SIAM. CBMSNSF Regional Conference Series in Applied Mathematics. · [Zbl 0813.62001](#)
- [26] Yuan, M.; Lin, Y., Model selection and estimation in regression with grouped variable, *Journal of the Royal Statistical Society, Series B*, 68, 49-67 (2006) · [Zbl 1141.62030](#)
- [27] Zhao, P.; Yu, B., On model selection consistency of Lasso, *Journal of Machine Learning Research*, 7, 2541-2563 (2006) · [Zbl 1222.62008](#)
- [28] Zhu, LP; Wang, T.; Zhu, LX; Ferré, L., Sufficient dimension reduction through discretization-expectation estimation, *Biometrika*, 97, 295-304 (2010) · [Zbl 1205.62048](#)
- [29] Zhu, LP; Li, LX; Li, R.; Zhu, LX, Model-free feature screening for ultrahigh-dimensional data, *Journal of the American Statistical Association*, 106, 1464-1474 (2011) · [Zbl 1233.62195](#)
- [30] Zhu, LX; Miao, BQ; Peng, H., On sliced inverse regression with high dimensional covariates, *Journal of the American Statistical Association*, 101, 630-643 (2006) · [Zbl 1119.62331](#)
- [31] Zou, H., The adaptive lasso and its oracle properties, *Journal of the American Statistical Association*, 101, 1418-1429 (2006) · [Zbl 1171.62326](#)

This reference list is based on information provided by the publisher or from digital mathematics libraries. Its items are heuristically matched to zbMATH identifiers and may contain data conversion errors. It attempts to reflect the references listed in the original paper as accurately as possible without claiming the completeness or perfect precision of the matching.