

Wang, Tao; Zhu, Lixing

Consistent tuning parameter selection in high dimensional sparse linear regression. (English)

Zbl 1216.62103

J. Multivariate Anal. 102, No. 7, 1141-1151 (2011).

Summary: An exhaustive search as required for traditional variable selection methods is impractical in high dimensional statistical modeling. Thus, to conduct variable selection, various forms of penalized estimators with good statistical and computational properties, have been proposed during the past two decades. The attractive properties of these shrinkage and selection estimators, however, depend critically on the size of regularization which controls model complexity. We consider the problem of consistent tuning parameter selection in high dimensional sparse linear regression where the dimension of the predictor vector is larger than the size of the sample. First, we propose a family of high dimensional Bayesian Information Criteria (HBIC), and then investigate the selection consistency, extending the results of the extended Bayesian Information Criterion (EBIC), of *J. Chen* and *Z. Chen* [Biometrika 95, No. 3, 795–771 (2008; Zbl 1437.62415)] to ultra-high dimensional situations. Second, we develop a two-step procedure, the SIS + AENET, to conduct variable selection in $p > n$ situations. The consistency of tuning parameter selection is established under fairly mild technical conditions. Simulation studies are presented to confirm theoretical findings, and an empirical example is given to illustrate the use in the internet advertising data.

MSC:

- 62J05 Linear regression; mixed models
- 62F15 Bayesian inference
- 62H12 Estimation in multivariate analysis
- 65C60 Computational problems in statistics (MSC2010)
- 62G20 Asymptotic properties of nonparametric inference

Cited in 8 Documents

Keywords:

adaptive elastic net; Bayesian information criterion; high dimensionality; sure independence screening; variable selection

Full Text: DOI

References:

- [1] Broman, K.W.; Speed, T.P., A model selection approach for the identification of quantitative trait loci in experimental crosses, J. roy. statist. soc. ser. B, 64, 641-656, (2002) · Zbl 1067.62108
- [2] Chen, J.; Chen, Z., Extended Bayesian information criteria for model selection with large model spaces, Biometrika, 95, 759-771, (2008) · Zbl 1437.62415
- [3] Donoho, D., For most large underdetermined systems of linear equations, the minimal ℓ^1 -norm near-solution approximates the sparsest near-solution, Comm. pure appl. math., 59, 797-829, (2006) · Zbl 1113.15004
- [4] Donoho, D.; Johnstone, I., Ideal spatial adaptation by wavelet shrinkage, Biometrika, 81, 425-455, (1994) · Zbl 0815.62019
- [5] Efron, B.; Hastie, T.; Johnstone, I.; Tibshirani, R., Least angle regression, Ann. statist., 32, 407-499, (2004) · Zbl 1091.62054
- [6] Fan, J.; Li, R., Variable selection via nonconcave penalized likelihood and its oracle properties, J. amer. statist. assoc., 96, 1348-1360, (2001) · Zbl 1073.62547
- [7] Fan, J.; Li, R., Statistical challenges with high-dimensionality: feature selection in knowledge discovery, (), 595-622 · Zbl 1117.62137
- [8] Fan, J.; Lv, J., Sure independence screening for ultra-high-dimensional feature space, J. roy. statist. soc. ser. B, 70, 849-911, (2008)
- [9] Fan, J.; Lv, J., A selective overview of variable selection in high dimensional feature space, Statist. sinica, 20, 101-148, (2010) · Zbl 1180.62080
- [10] Fan, J.; Peng, H., Non-concave penalized likelihood with a diverging number of parameters, Ann. statist., 32, 928-961, (2004) · Zbl 1092.62031

- [11] Foster, D.; George, E., The risk inflation criterion for multiple regression, *Ann. statist.*, 22, 1947-1975, (1994) · [Zbl 0829.62066](#)
- [12] Meinshausen, N.; Yu, B., LASSO-type recovery of sparse representations for high dimensional data, *Ann. statist.*, 37, 246-270, (2009) · [Zbl 1155.62050](#)
- [13] Schwarz, G., Estimating the dimension of a model, *Ann. statist.*, 6, 461-464, (1978) · [Zbl 0379.62005](#)
- [14] Siegmund, D., Model selection in irregular problems: application to mapping quantitative loci, *Biometrika*, 91, 785-800, (2004) · [Zbl 1064.62114](#)
- [15] Tibshirani, R., Regression shrinkage and selection via the lasso, *J. roy. statist. soc. ser. B*, 58, 267-288, (1996) · [Zbl 0850.62538](#)
- [16] Wang, H., Forward regression for ultra-high dimensional variable screening, *J. amer. statist. assoc.*, 104, 1512-1524, (2009) · [Zbl 1205.62103](#)
- [17] Wang, H.; Li, B.; Leng, C., Shrinkage tuning parameter selection with a diverging number of parameters, *J. roy. statist. soc. ser. B*, 71, 671-683, (2009) · [Zbl 1250.62036](#)
- [18] Wang, H.; Li, R.; Tsai, C.L., Tuning parameter selectors for the smoothly clipped absolute deviation method, *Biometrika*, 94, 553-568, (2007) · [Zbl 1135.62058](#)
- [19] Zhang, C.H.; Huang, J., The sparsity and bias of the LASSO selection in high-dimensional linear regression, *Ann. statist.*, 36, 1567-1594, (2008) · [Zbl 1142.62044](#)
- [20] Zou, H., The adaptive LASSO and its oracle properties, *J. amer. statist. assoc.*, 101, 1418-1429, (2006) · [Zbl 1171.62326](#)
- [21] Zou, H.; Hastie, T., Regularization and variable selection via the elastic net, *J. roy. statist. soc. ser. B*, 67, 301-320, (2005) · [Zbl 1069.62054](#)
- [22] Zou, H.; Zhang, H.H., On the adaptive elastic-net with a diverging number of parameters, *Ann. statist.*, 37, 1733-1751, (2009) · [Zbl 1168.62064](#)

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.