Electronic Journal of Statistics > Vol. 1 (2007) open journal systems

Smoothing ℓ_1 -penalized estimators for high-dimensional time-course data

Lukas Meier, *Seminar für Statistik, ETH Zürich*
Peter Bühlmann, *Seminar für Statistik, ETH Zürich*

Abstract

When a series of (related) linear models has to be estimated it is often appropriate to combine the different data-sets to construct more efficient estimators. We use ℓ_1 -penalized estimators like the Lasso or the Adaptive Lasso which can simultaneously do parameter estimation and model selection. We show that for a time-course of high-dimensional linear models the mean squared error rate of the Lasso and of the Adaptive Lasso can be improved by combining the different time-points in a suitable way. Moreover, the Adaptive Lasso still enjoys oracle properties and consistent variable selection. The finite sample properties of the proposed methods are illustrated on simulated data and on a real problem of motif finding in DNA sequences.

AMS 2000 subject classifications: Primary 62J07; secondary 62J99, 62H12.

Keywords: Lasso, Local least squares, Multivariate regression, Variable selection, Weighted likelihood.

Full Text: PDF

Meier, Lukas, Bühlmann, Peter, Smoothing ℓ_1 -penalized estimators for high-dimensional time-course data, Electronic Journal of Statistics, 1, (2007), 597-615 (electronic). DOI: 10.1214/07-EJS103.

References

[1] Beer, M. A. and Tavazoie, S. (2004). Predicting gene expression from sequence. Cell 117, 185–198.

[2] Breiman, L. and Friedman, J. H. (1997). Predicting multivariate responses in multiple linear regression. Journal of the Royal Statistical Society Series B 59, 1, 3–54. MR1436554

[3] Bunea, F., Tsybakov, A., and Wegkamp, M. H. (2007). Sparsity oracle inequalities for the lasso. Electronic Journal of Statistics 1, 169–194. MR2312149

[4] Conlon, E. M., Liu, X. S., Lieb, J. D., and Liu, J. S. (2003). Integrating regulatory motif discovery and genome-wide expression analysis. Proceedings of the National Academy of Science 100, 3339 – 3344.

[5] Donoho, D. L. and Johnstone, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. Biometrika 81, 3, 425–455. MR1311089

[6] Greenshtein, E. and Ritov, Y. (2004). Persistency in high dimensional linear predictor-selection and the virtue of over-parametrization. Bernoulli 10, 971–988. MR2108039

[7] Hu, F. and Zidek, J. V. (2002). The weighted likelihood. The Canadian Journal of

Statistics 30, 3, 347–371. MR1944367

[8] Huang, J., Ma, S., and Zhang, C.-H. (2006). Adaptive lasso for sparse high-dimensional regression models. Tech. rep., The University of Iowa.

[9] Lokhorst, J. (1999). The lasso and generalised linear models. Honors Project, The University of Adelaide, Australia.

[10] Meier, L. and Bühlmann, P. (2007). Discussion of “One-step sparse estimates in nonconcave penalized likelihood models” (auths Zou, H. and Li, R.). Annals of Statistics. To appear.

[11] Meinshausen, N. (2007). Lasso with relaxation. Computational Statistics and Data Analysis 52, 1, 374 – 393.

[12] Meinshausen, N. and Bühlmann, P. (2006). High dimensional graphs and variable selection with the lasso. Annals of Statistics 34, 3, 1436–1462. MR2278363

[13] Meinshausen, N. and Yu, B. (2006). Lasso-type recovery of sparse representations for high-dimensional data. Tech. rep., University of California, Berkeley.

[14] Park, M.-Y. and Hastie, T. (2007). An l1 regularization-path algorithm for generalized linear models. Journal of the Royal Statistical Society Series B 69, 4, 659–677.

[15] Similä, T. and Tikka, J. (2007). Input selection and shrinkage in multiresponse linear regression. Computational Statistics and Data Analysis 52, 1, 406–422.

[16] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B 58, 1, 267–288. MR1379242

[17] Tibshirani, R. (1997). The lasso method for variable selection in the cox model. Statistics in Medicine 16, 385–395.

[18] Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005). Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society Series B 67, 1, 91–108. MR2136641

[19] Turlach, B., Venables, W., and Wright, S. (2005). Simultaneous variable selection. Technometrics 47, 3, 349–363. MR2164706

[20] van de Geer, S. (2007). High-dimensional generalized linear models and the lasso. Annals of Statistics. To appear.

[21] Wasserman, L. and Roeder, K. (2007). High dimensional variable selection. Tech. rep., Carnegie Mellon University.

[22] Zhang, C.-H. and Huang, J. (2007). The sparsity and bias of the lasso selection in high-dimensional linear regression. Annals of Statistics. To appear.

[23] Zidek, J. V. (1997). Discussion of “Predicting multivariate responses in multiple linear regression”(auths. Breiman, L. and Friedman, J.). Journal of the Royal Statistical Society Series B 59, 1, 42–43. MR1436554

[24] Zou, H. (2006). The adaptive lasso and its oracle properties. Journal of the American Statistical Association 101, 476, 1418–1429. MR2279469

[25] Zou, H. and Li, R. (2007). One-step sparse estimates in nonconcave penalized likelihood models. Annals of Statistics. To appear.

Electronic Journal of Statistics. ISSN: 1935-7524