Hunter, David R.; **Li, Runze**

Variable selection using MM algorithms. (English) Zbl 1078.62028

Ann. Stat. 33, No. 4, 1617-1642 (2005).

Summary: Variable selection is fundamental to high-dimensional statistical modeling. Many variable selection techniques may be implemented by maximum penalized likelihood using various penalty functions. Optimizing the penalized likelihood function is often challenging because it may be nondifferentiable and/or nonconcave. This article proposes a new class of algorithms for finding a maximizer of the penalized likelihood for a broad class of penalty functions. These algorithms operate by perturbing the penalty function slightly to render it differentiable, then optimizing this differentiable function using a minorize - maximize (MM) algorithm.

MM algorithms are useful extensions of the well-known class of EM algorithms, a fact that allows us to analyze the local and global convergence of the proposed algorithm using some of the techniques employed for EM algorithms. In particular, we prove that when our MM algorithms converge, they must converge to a desirable point; we also discuss conditions under which this convergence may be guaranteed. We exploit the Newton-Raphson-like aspect of these algorithms to propose a sandwich estimator for the standard errors of the estimators. Our method performs well in numerical tests.

MSC:

62F99	Parametric inference
65C60	Computational problems in statistics (MSC2010)
62J12	Generalized linear models (logistic models)
65C20	Probabilistic models, generic numerical methods in probability and statistics
62F10	Point estimation

Keywords:

AIC; BIC; LASSO; oracle estimator; SCAD; environment data; penalized likelihood; MM algorithms; EM algorithms

Full Text: DOI arXiv

References:

[1] Antoniadis, A. (1997). Wavelets in statistics: A review (with discussion). J. Italian Statistical Society 6 97–144.

[2] Antoniadis, A. and Fan, J. (2001). Regularization of wavelets approximations (with discussion). J. Amer. Statist. Assoc. 96 939–967. · Zbl 1072.62561 · doi:10.1198/016214501753208942

[3] Cai, J., Fan, J., Li, R. and Zhou, H. (2005). Variable selection for multivariate failure time data. Biometrika . · Zbl 1094.62123 · doi:10.1093/biomet/92.2.303

[4] Cox, D. R. (1975). Partial likelihood. Biometrika 62 269–276. · Zbl 0312.62002 · doi:10.1093/biomet/62.2.269

[5] Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. Numer. Math. 31 377–403. · Zbl 0377.65007 · doi:10.1007/BF01404567

[6] Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (with discussion). J. Roy. Statist. Soc. Ser. B 39 1–38. · Zbl 0364.62022

[7] Fan, J. and Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. 96 1348–1360. · Zbl 1073.62547 · doi:10.1198/016214501753382273

[8] Fan, J. and Li, R. (2002). Variable selection for Cox's proportional hazards model and frailty model. Ann. Statist. 30 74–99. · Zbl 1012.62106 · doi:10.1214/aos/1015362185

[9] Fan, J. and Li, R. (2004). New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. J. Amer. Statist. Assoc. 99 710–723. · Zbl 1117.62329 · doi:10.1198/016214504000001060

[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] Heiser, W. J. (1995). Convergent computation by iterative majorization: Theory and applications in multidimensional data

analysis. In Recent Advances in Descriptive Multivariate Analysis (W. J. Krzanowski ed.) 157–189. Clarendon Press, Oxford.

[13] Hestenes, M. R. (1975). Optimization Theory : The Finite Dimensional Case . Wiley, New York. · Zbl 0327.90015

[14] Hunter, D. R. and Lange, K. (2000). Rejoinder to discussion of "Optimization transfer using surrogate objective functions." J. Comput. Graph. Statist. 9 52–59.

[15] Kauermann, G. and Carroll, R. J. (2001). A note on the efficiency of sandwich covariance matrix estimation. J. Amer. Statist. Assoc. 96 1387–1396. · Zbl 1073.62539 · doi:10.1198/016214501753382309

[16] Lange, K. (1995). A gradient algorithm locally equivalent to the EM algorithm. J. Roy. Statist. Soc. Ser. B 57 425–437. · Zbl 0813.62021

[17] Lange, K., Hunter, D. R. and Yang, I. (2000). Optimization transfer using surrogate objective functions (with discussion). J. Comput. Graph. Statist. 9 1–59.

[18] McLachlan, G. and Krishnan, T. (1997). The EM Algorithm and Extensions . Wiley, New York. · Zbl 0882.62012

[19] Meng, X.-L. (1994). On the rate of convergence of the ECM algorithm. Ann. Statist. 22 326–339. JSTOR: · Zbl 0803.65146 · doi:10.1214/aos/1176325371

[20] Meng, X.-L. and Van Dyk, D. A. (1997). The EM algorithm—An old folk song sung to a fast new tune (with discussion). J. Roy. Statist. Soc. Ser. B 59 511–567. · Zbl 1090.62518 · doi:10.1111/1467-9868.00082

[21] Miller, A. J. (2002). Subset Selection in Regression , 2nd ed. Chapman and Hall, London. · Zbl 1051.62060

[22] Ortega, J. M. (1990). Numerical Analysis : A Second Course , 2nd ed. SIAM, Philadelphia. · Zbl 0701.65002

[23] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. J. Roy. Statist. Soc. Ser. B 58 267–288. · Zbl 0850.62538

[24] Wu, C.-F. J. (1983). On the convergence properties of the EM algorithm. Ann. Statist. 11 95–103. JSTOR: · Zbl 0517.62035