Forward stagewise regression and the monotone lasso

Trevor J. Hastie, *Stanford Univ.*
Jonathan Taylor, *Stanford Univ.*
Rob Tibshirani, *Stanford*
Guenther Walther, *Stanford Univ.*

Abstract
We consider the least angle regression and forward stagewise algorithms for solving penalized least squares regression problems. In Efron, Hastie, Johnstone & Tibshirani (2004) it is proved that the least angle regression algorithm, with a small modification, solves the lasso regression problem. Here we give an analogous result for incremental forward stage-wise regression, showing that it solves a version of the lasso problem that enforces monotonicity. One consequence of this is as follows: while lasso makes optimal progress in terms of reducing the residual sum-of-squares per unit increase in L_1-norm of the coefficient $beta$, forward stage-wise is optimal per unit L_1 arc-length traveled along the coefficient path. We also study a condition under which the coefficient paths of the lasso are monotone, and hence the different algorithms coincide. Finally, we compare the lasso and forward stagewise procedures in a simulation study involving a large number of correlated predictors.

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

Keywords: regression, lasso, stagewise.

Full Text: PDF

References

Buhlmann, P. (2006), 'Boosting for high-dimensional linear models', Annals of Statistics 34(2), 559–583. MR2281878

Chen, S. S., Donoho, D. L. and Saunders, M. A. (1998), 'Atomic decomposition by basis pursuit', SIAM Journal on Scientific Computing pp. 33–61.

Donoho, D. and Tsaig, Y. (2006), Fast solution of ℓ_1-norm minimization problems when the solution may be sparse, Technical report, Stanford University.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004), 'Least angle regression', Annals of Statistics 32(2), 407–499. (with discussion).

Friedman, J. (2001), 'Greedy function approximation: a gradient boosting machine', Annals of Statistics 29, 1180. MR1873328

Friedman, J., Hastie, T. and Tibshirani, R. (2000), 'Additive logistic regression: a statistical view of boosting (with discussion)', Annals of Statistics 28, 337–307.

Friedman, J. and Popescu, B. (2004), Gradient directed regularization, Technical report, Stanford University.

Hastie, T., Tibshirani, R. and Friedman, J. (2001), The Elements of Statistical Learning; Data mining, Inference and Prediction, Springer Verlag, New York.

Osborne, M., Presnell, B. and Turlach, B. (2000), 'A new approach to variable selection in least squares problems', IMA Journal of Numerical Analysis 20, 389–404.

Park, M.-Y. and Hastie, T. (2006), l_1 regularization path algorithms for generalized linear models, Technical report, Stanford University.

Rosset, S. (2005), Tracking curved regularized optimization solution paths, in ' Advances in Neural Information Processing Systems (NIPS*2004)', Vol. 17, MIT Press, Cambridge, MA. to appear.

Rosset, S. and Zhu, J. (2004), 'Piecewise linear regularized solution paths', Annals of Statistics. (to appear 2007).

Schapire, R. E. and Freund, Y. (1997), 'Boosting the margin: A new explanation for the effectiveness of voting methods', Proceedings of the Fourteenth International Conference on Machine Learning .

Schapire, R., Freund, Y., Bartlett, P. and Lee, W. (1998), 'Boosting the margin: a new explanation for the effectiveness of voting methods', Annals of Statistics 26(5), 1651–1686.

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', J. Royal. Statist. Soc. B. 58, 267–288. MR1379242

Tropp, J. (2006), 'Just relax: convex programming methods for identifying sparse signals in noise', IEEE Transactions on Information Theory 52, 1030–1051. MR2238069

Tropp, J. A. (2004), 'Greed is good: algorithmic results for sparse approximation', IEEE Transactions on Information Theory 50, 2231–2242. MR2097044

Zhao, P. and Yu, B. (2004), Boosted lasso, Technical Report 678, Statistics Department, University of California, Berkeley.