Least angle and L_1 penalized regression: A review

Tim C. Hesterberg, *Insightful Corp.*
Nam Hee Choi, *University of Michigan*
Lukas Meier, *ETH Zuerich*
Chris Fraley, *Insightful Corp.*

Abstract

Least Angle Regression is a promising technique for variable selection applications, offering a nice alternative to stepwise regression. It provides an explanation for the similar behavior of LASSO (L_1-penalized regression) and forward stagewise regression, and provides a fast implementation of both. The idea has caught on rapidly, and sparked a great deal of research interest. In this paper, we give an overview of Least Angle Regression and the current state of related research.

Full Text: PDF

Hesterberg, Tim C., Choi, Nam Hee, Meier, Lukas, Fraley, Chris, Least angle and L_1 penalized regression: A review, Statistics Surveys, , (), 61-93 (electronic). DOI: 10.1214/08-SS035.

References

Adams, J. L. (1990) A computer experiment to evaluate regression strategies. In Proceedings of the Statistical Computing Section, 55–62. American Statistical Association.

Avalos, M., Grandvalet, Y. and Ambroise, C. (2007) Parsimonious additive models. Computational Statistics and Data Analysis, 51, 2851–2870. MR2345610

Bakin, S. (1999) Adaptive regression and model selection in data mining problems. Ph.D. thesis, The Australian National University.

Balakrishnan, S. and Madigan, D. (2007) Finding predictive runs with laps. International Conference on Machine Learning (ICML), 415–420.

Banerjee, O., El Ghaoui, L. and d'Aspremont, A. (2008) Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. Journal of Machine Learning Research. (to appear).

Breiman, L. (1995) Better subset regression using the nonnegative garrote. Technometrics, 37, 373–384. MR1365720

Bühlmann, P. and Meier, L. (2008) Discussion of "One-step sparse estimates in nonconcave penalized likelihood models" by H. Zou and R. Li. Annals of Statistics. (to appear).

Bühlmann, P. and Yu, B. (2006) Sparse boosting. Journal of Machine Learning Research, 7, 1001–1024. MR2274395

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

Candes, E. and Tao, T. (2007) The Dantzig selector: statistical estimation when p is much larer than n. Annals of Statistics, 35, 2313–2351. MR2382644

Candes, E. J., Wakin, M. and Boyd, S. (2007) Enhancing sparsity by reweighted L1 minimization. Tech. rep., California Institute of Technology.

Chen, S., Donoho, D. and Saunders, M. (1998) Atomic decomposition by basis pursuit. SIAM Journal on Scientific Computing, 20, 33–61. MR1639094

Choi, N. H. and Zhu, J. (2006) Variable selection with strong heredity/marginality constraints. Tech. rep., Department of Statistics, University of Michigan.

Dahl, J., Vandenberghe, L. and Roychowdhury, V. (2008) Covariance selection for non-chordal graphs via chordal embedding. Optimization Methods and Software. (to appear).

Draper, N. R. and Smith, H. (1998) Applied regression analysis. Wiley, 3rd edn. MR1614335

Efron, B. and Hastie, T. (2003) LARS software for R and Splus. http://www-stat.stanford.edu/~hastie/Papers/LARS.

Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004) Least angle regression. Annals of Statistics, 32, 407–451. MR2060166

Efron, B., Hastie, T. and Tibshirani, R. (2007) Discussion of “the Dantzig selector” by E. Candes and T. Tan. Annals of Statistics, 35, 2358–2364. MR2382646

Efroymson, M. A. (1960) Multiple regression analysis. In Mathematical Methods for Digital Computers (eds. A. Ralston and H. S. Wilf), vol. 1, 191–203. Wiley. MR0117923

Fan, J. and Li, R. (2001) Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association, 96, 1348–1360. MR1946581

Fraley, C. and Hesterberg, T. (2007) Least-angle regression and Lasso for large datasets. Tech. rep., Insightful Corporation.

Frank, I. E. and Friedman, J. H. (1993) A statistical view of some chemometrics regression tools, with discussion. Technometrics, 35, 109–148.

Freund, Y. and Schapire, R. E. (1997) A decision-theoretic generalization of online learning and an application to boosting. Journal of Computer and System Sciences, 55, 119–139. MR1473055

Friedman, J. (2006) Herding lambdas: fast algorithms for penalized regression and classification. Manuscript.

Friedman, J. H. (1991) Mulivariate adaptive regression splines. Annals of Statistics, 19, 1–67. MR1091842

Friedman, J. H., Hastie, T., Höfling, H. and Tibshirani, R. (2007a) Pathwise coordinate optimization. Annals of Applied Statistics, 1, 302–332.

Friedman, J. H., Hastie, T. and Tibshirani, R. (2007b) Sparse inverse covariance estimation with the graphical lasso. Biostatistics. (published online December 12, 2007).

Fu, W. (1998) Penalized regressions: the Bridge versus the Lasso. Journal of Computational and Graphical Statistics, 7, 397–416. MR1646710

Fu, W. (2000) S-PLUS package brdgrun for shrinkage estimators with bridge penalty. http://lib.stat.cmu.edu/S/brdgrun.shar.

Furnival, G. M. and Wilson, Jr., R. W. (1974) Regression by leaps and bounds. Technometrics, 16, 499–511.

Gao, H.-Y. (1998) Wavelet shrinkage denoising using the non-negative garrote. Journal of Computational and Graphical Statistics, 7, 469–488. MR1665666

Genkin, A., Lewis, D. D. and Madigan, D. (2007) Large-scale Bayesian logistic regression for text categorization. Technometrics, 49, 291–304.

Ghosh, S. (2007) Adaptive elastic net: An improvement of elastic net to achieve oracle properties. Tech. rep., Department of Mathematical Sciences, Indiana University-Purdue University, Indianapolis.

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

Gui, J. and Li, H. (2005) Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data. Bioinformatics, 21, 3001–3008.

Guigue, V., Rakotomamonjy, A. and Canu, S. (2006) Kernel basis pursuit. Revue d'Intelligence Artificielle, 20, 757–774.

Gunn, S. R. and Kandola, J. S. (2002) Structural modeling with sparse kernels. Machine Learning, 10, 581–591.

Hamada, M. and Wu, C. (1992) Analysis of designed experiments with complex aliasing. Journal of Quality Technology, 24, 130–137.

Hastie, T., Rosset, S., Tibshirani, R. and Zhu, J. (2004) The entire regularization path for the support vector machine. Journal of Machine Learning Research, 5, 1391–1415. 3/5/04. MR2248021

Hastie, T., Taylor, J., Tibshirani, R. and Walther, G. (2007) Forward stagewise regression and the monotone lasso. Electronic Journal of Statistics, 1, 1–29. MR2312144

Hastie, T., Tibshirani, R. and Friedman, J. (2001) The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer Verlag. MR1851606

Hesterberg, T. and Fraley, C. (2006a) Least angle regression. Proposal to NIH, http://www.insightful.com/lars.

Hesterberg, T. and Fraley, C. (2006b) S-PLUS and R package for least angle regression. In Proceedings of the American Statistical Association, Statistical Computing Section [CD-ROM], 2054–2061. Alexandria, VA: American Statistical Association.

Huang, J., Ma, S. and Zhang, C.-H. (2008) Adaptive Lasso for sparse high-dimensional regression models. Statisica Sinica. (to appear).

Hurvich, C. M. and Tsai, C.-L. (1990) The impact of model selection on inference in linear regression. The American Statistician, 44, 214–217.

Insightful Corportation (2006) GLARS: Generalized Least Angle Regression software for R and S-PLUS. http://www.insightful.com/lars.

Ishwaran, H. (2004) Discussion of "Least Angle Regression" by Efron et al. Annals of Statistics, 32, 452–458. MR2060166

Jolliffe, I., Trendafilov, N. and Uddin, M. (2003) A modified principal component technique based on the LASSO. Journal of Computational and Graphical Statistics, 12, 531–547. MR2002634

Keerthi, S. and Shevade, S. (2007) A fast tracking algorithm for generalized lars/lasso. IEEE Transactions on Neural Networks, 18, 1826–1830.

Khan, J. A., Van Aelst, S. and Zamar, R. H. (2007) Robust linear model selection based on least angle regression. Journal of the American Statistical Association, 102, 1289–1299.

Kim, J., Kim, Y. and Kim, Y. (2005a) glasso: R-package for Gradient LASSO algorithm. http://idea.snu.ac.kr/Research/glassojskim/glasso.htm.
R package version 0.9, http://idea.snu.ac.kr/Research/glassojskim/ glasso.htm.

Kim, J., Kim, Y. and Kim, Y. (2005b) Gradient LASSO algorithm. Technical report, Seoul National University.

Kim, Y., Kim, J. and Kim, Y. (2006) Blockwise sparse regression. Statistica Sinica, 16, 375–390. MR2267240

Knight, K. (2004) Discussion of “Least Angle Regression” by Efron et al. Annals of Statistics, 32, 458–460. MR2060166

Leng, C., Lin, Y. and Wahba, G. (2006) A note on the LASSO and related procedures in model selection. Statistica Sinica, 16, 1273–1284. MR2327490

Lin, Y. and Zhang, H. (2006) Component selection and smoothing in multivariate nonparametric regression. Annals of Statistics, 34, 2272–2297. MR2291500

Lokhorst, J. (1999) The LASSO and Generalised Linear Models. Honors Project, The University of Adelaide, Australia.

Lokhorst, J., Venables, B. and Turlach, B. (1999) Lasso2: L1 Constrained Estimation Routines. http://www.maths.uwa.edu.au/ berwin/software/ lasso.html.

Loubes, J. and Massart, P. (2004) Discussion of “least angle regression” by Efron et al. Annals of Statistics, 32, 460–465. MR2060166

Lu, W. and Zhang, H. (2007) Variable selection for proportional odds model. Statistics in Medicine, 26, 3771–3781.

Madigan, D. and Ridgeway, G. (2004) Discussion of “least angle regression” by Efron et al. Annals of Statistics, 32, 465–469. MR2060166

McCullagh, P. and Nelder, J. A. (1989) Generalised Linear Models. London: Chapman and Hall.

Meier, L. and Bühlmann, P. (2007) Smoothing ℓ_1-penalized estimators for high-dimensional time-course data. Electronic Journal of Statistics, 1, 597–615. MR2369027

Meier, L., van de Geer, S. and Bühlmann, P. (2008) The group lasso for logistic regression. Journal of the Royal Statistical Society, Series B, 70, 53–71.

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

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

Meinshausen, N., Rocha, G. and Yu, B. (2007) A tale of three cousins: Lasso, L2Boosting, and Dantzig. Annals of Statistics, 35, 2373–2384. MR2382649

Meinshausen, N. and Yu, B. (2008) Lasso-type recovery of sparse representations for high-dimensional data. Annals of Statistics. (to appear).

Miller, A. (2002) Subset Selection in Regression. Chapman & Hall, second edn. MR2001193

Monahan, J. F. (2001) Numerical Methods of Statistics. Cambridge University Press. MR1813549

Osborne, M. R., Presnell, B. and Turlach, B. A. (2000a) A new approach to variable selection in least squares problems. IMA Journal of Numerical Analysis, 20, 389–403. MR1773265

Osborne, M. R., Presnell, B. and Turlach, B. A. (2000b) On the LASSO and its dual. Journal of Computational and Graphical Statistics, 9, 319–337. MR1822089

Owen, A. (2006) A robust hybrid of lasso and ridge regression. From the web.

Park, M. Y. and Hastie, T. (2006a) glmpath: L1 Regularization Path for Generalized Linear Models and Proportional Hazards Model. URL http://cran.r-project.org/src/contrib/Descriptions/glmpath.html. R package version 0.91.

Park, M. Y. and Hastie, T. (2006b) Regularization path algorithms for detecting gene interactions. Tech. rep., Department of Statistics, Stanford University.

Park, M.-Y. and Hastie, T. (2007) An L_1 regularization-path algorithm for generalized linear models. Journal of the Royal Statistical Society Series B, 69, 659–677. MR2370074

Roecker, E. B. (1991) Prediction error and its estimation for subset-selected models. Technometrics, 33, 459–468.

Rosset, S. (2005) Following curved regularized optimization solution paths. In Advances in Neural Information Processing Systems 17 (eds. L. K. Saul, Y. Weiss and L. Bottou), 1153–1160. Cambridge, MA: MIT Press.

Rosset, S. and Zhu, J. (2004) Discussion of “Least Angle Regression” by Efron et al. Annals of Statistics, 32, 469–475. MR2060166

Rosset, S. and Zhu, J. (2007) Piecewise linear regularized solution paths. Annals of Statistics, 35, 1012–1030. MR2341696

Roth, V. (2004) The generalized LASSO. IEEE Transactions on Neural Networks, 15, 16–28.

Segal, M. R., Dahlquist, K. D. and Conklin, B. R. (2003) Regression approaches for microarray data analysis. Journal of Computational Biology, 10, 961–980.

Shi, W., Wahba, G., Wright, S., Lee, K., Klein, R. and Klein, B. (2008) Lasso-patternsearch algorithm with application to ophthalmology and genomic data. Statistics and Its Interface. (to appear).

Silva, J., Marques, J. and Lemos, J. (2005) Selecting landmark points for sparse manifold learning. In Advances in Neural Information Processing Systems 18 (eds. Y. Weiss, B. Schölkopf and J. Platt), 1241–1248. Cambridge, MA: MIT Press.

Similä, T. and Tikka, J. (2006) Common subset selection of inputs in multiresponse regression. In IEEE International Joint Conference on Neural Networks, 1908–1915. Vancouver, Canada.

Stine, R. A. (2004) Discussion of “Least Angle Regression” by Efron et al. Annals of Statistics, 32, 475–481. MR2060166

Thisted, R. A. (1988) Elements of Statistical Computing. Chapman and Hall. MR0940474

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

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

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, 91–108. MR2136641

Trendafilov, N. and Joilliffe, I. (2007) Dlass: Variable selection in discriminant analysis via the lasso. Computational Statistics and Data Analysis, 51, 3718–3736. MR2364486

Turlach, B. A. (2004) Discussion of “Least Angle Regression” by Efron et al. Annals of Statistics, 32, 481–490. MR2060166

Turlach, B. A., Venables, W. N. and Wright, S. J. (2005) Simultaneous variable selection. Technometrics, 47, 349–363. MR2164706

van de Geer, S. (2008) High-dimensional generalized linear models and the lasso. Annals of Statistics, 36, 614–645.

Wang, G., Yeung, D.-Y. and Lochovsky, F. (2007a) The kernel path in kernelized LASSO. In International Conference on Artificial Intelligence and Statistics. San Juan, Puerto Rico.

Wang, H. and Leng, C. (2006) Improving grouped variable selection via aglasso. Tech. rep., Peking University & National University of Singapore.

Wang, H. and Leng, C. (2007) Unified LASSO estimation via least squares approximation. Journal of the American Statistical Association, 102, 1039–1048.

Wang, H., Li, G. and Tsai, C. (2007b) Regression coefficient and autoregressive order shrinkage and selection via the lasso. Journal of the Royal Statistical Society, Series B, 69, 63–78. MR2301500

Yuan, M. (2008) Efficient computation of the ℓ_1 regularized solution path in Gaussian graphical models. Journal of Computational and Graphical Statistics. (to appear).

Yuan, M., Joseph, R. and Lin, Y. (2007) An efficient variable selection approach for analyzing designed experiments. Technometrics, 49, 430–439.

Yuan, M. and Lin, Y. (2006) Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society, Series B, 68, 49–68. MR2212574

Yuan, M. and Lin, Y. (2007a) Model selection and estimation in the Gaussian graphical model. Biometrika, 94, 19–35. MR2367824

Yuan, M. and Lin, Y. (2007b) On the non-negative garrote estimator. Journal of the Royal Statistical Society, Series B, 69, 143–161. MR2325269

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.

Zhang, H. and Lu, W. (2007) Adaptive Lasso for Cox’s proportional hazards model. Biometrika, 94, 691–703.

Zhang, H., Wahba, G., Lin, Y., Voelker, M., Ferris, M., Klein, R. and Klein, B. (2004) Variable selection and model building via likelihood basis pursuit. Journal of the American Statistical Association, 99, 659–672. MR2090901

Zhao, P., Rocha, G. and Yu, B. (2008) Grouped and hierarchical model selection through composite absolute penalties. Annals of Statistics. (to appear).

Zhao, P. and Yu, B. (2006) On model selection consistency of Lasso. Journal of Machine Learning Research, 7, 2541–2567. MR2274449

Zhao, P. and Yu, B. (2007) Stagewise Lasso. Journal of Machine Learning Research, 8, 2701–2726.

Zhu, J., Rosset, S., Hastie, T. and Tibshirani, R. (2003) 1-norm support vector machines. In Advances in Neural Information Processing Systems 16, 49–56. MIT Press. NIPS 2003 Proceedings.

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

Zou, H. and Hastie, T. (2005a) elasticnet: Elastic Net Regularization and Variable Selection. R package version 1.0-3.

Zou, H. and Hastie, T. (2005b) Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society, Series B, 67, 301–320. MR2137327

Zou, H., Hastie, T. and Tibshirani, R. (2006) Sparse principal component analysis. Journal of Computational and Graphical Statistics, 15, 265–286. MR2252527

Zou, H., Hastie, T. and Tibshirani, R. (2007) On the “Degrees of Freedom” of the Lasso. Annals of Statistics, 35, 2173–2192. MR2363967

Zou, H. and Li, R. (2008) One-step sparse estimates in nonconcave penalized likelihood models. Annals of Statistics. (to appear).

Statistics Surveys. ISSN: 1935-7516