Newer Approaches to Regression and Tree-Based Modeling

These approaches may not require data reduction before modeling. But recent research has shown them to be "data hungry"[5]

- lasso (shrinkage using L1 norm favoring zero regression coefficients) [9, 8]
- elastic net (combination of L1 and L2 norms that handles the $p > n$ case better than the lasso) [16]
- adaptive lasso [14, 10]
- more flexible lasso to differentially penalize for variable selection and for regression coefficient estimation [6]
- group lasso to force selection of all or none of a group of related variables (e.g., dummy variables representing a polytomous predictor)
- group lasso-like procedures that also allow for variables within a group to be removed [11]
- sparse-group lasso using L1 and L2 norms to achieve spareness on groups and within groups of variables [7]
- adaptive group lasso (Wang & Leng)
- Breiman's nonnegative garrote [13]
- "preconditioning", i.e., model simplification after developing a "black box" predictive model [4]
- sparse principal component analysis to achieve parsimony in data reduction [12, 15, 3, 2]
- bagging, boosting, and random forests. [1]

One problem prevents most of these methods from being ready for everyday use: they require scaling predictors before fitting the model. When a predictor is represented by nonlinear basis functions, the scaling recommendations in the literature are not sensible. There are also computational issues and difficulties obtaining hypothesis tests and confidence intervals.

0.1 Some Useful Links

- `http://freakonometrics.hypotheses.org/19424` has beautiful demonstrations of several methods using R to approximate a smooth 3-dimensional surface
- `http://freakonometrics.hypotheses.org/19874` has beautiful demonstrations of *boosting* when there is one continuous predictor

Annotated Bibliography

[1] Trevor Hastie, Robert Tibshirani, and Jerome H. Friedman. *The Elements of Statistical Learning*. second. ISBN-10: 0387848576; ISBN-13: 978-0387848570. New York: Springer, 2008.

[2] Seokho Lee, Jianhua Z. Huang, and Jianhua Hu. "Sparse logistic principal components analysis for binary data". In: *Ann Appl Stat* 4.3 (2010), pp. 1579–1601.

[3] Chenlei Leng and Hansheng Wang. "On general adaptive sparse principal component analysis". In: *J Comp Graph Stat* 18.1 (2009), pp. 201–215.

[4] Debashis Paul et al. ""Preconditioning" for feature selection and regression in high-dimensional problems". In: *Ann Stat* 36.4 (2008), pp. 1595–1619. DOI: `10.1214/009053607000000578`. URL: `http://dx.doi.org/10.1214/009053607000000578`.

develop consistent $\hat{Y}$ using a latent variable structure, using for example supervised principal components. Then run stepwise regression or lasso predicting $\hat{Y}$ (lasso worked better). Can run into problems when a predictor has importance in an adjusted sense but has no marginal correlation with Y;model approximation;model simplification
.

[5] Tjeerd van der Ploeg, Peter C. Austin, and Ewout W. Steyerberg. "Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints". In: *BMC Medical Research Methodology* 14.1 (Dec. 22, 2014), pp. 137+. ISSN: 1471-2288. DOI: `10.1186/1471-2288-14-137`. URL: `http://dx.doi.org/10.1186/1471-2288-14-137`.

Would be better to use proper accuracy scores in the assessment. Too much emphasis on optimism as opposed to final discrimination measure. But much good practical information. Recursive partitioning fared poorly.
.

[6] Peter Radchenko and Gareth M. James. "Variable inclusion and shrinkage algorithms". In: *J Am Stat Assoc* 103.483 (2008), pp. 1304–1315.

solves problem caused by lasso using the same penalty parameter for variable selection and shrinkage which causes lasso to have to keep too many variables in the model to avoid overshrinking the remaining predictors;does not handle scaling issue well
.

[7] Noah Simon et al. "A sparse-group lasso". In: *J Comp Graph Stat* 22.2 (2013), pp. 231–245. DOI: 10.1080/10618600.2012.681250. eprint: http://www.tandfonline.com/doi/pdf/10.1080/10618600.2012.681250. URL: http://www.tandfonline.com/doi/abs/10.1080/10618600.2012.681250.

sparse effects both on a group and within group levels;can also be considered special case of group lasso allowing overlap between groups
.

[8] Ewout W. Steyerberg et al. "Prognostic modelling with logistic regression analysis: A comparison of selection and estimation methods in small data sets". In: *Stat Med* 19 (2000), pp. 1059–1079.

[9] Robert Tibshirani. "Regression shrinkage and selection via the lasso". In: *J Roy Stat Soc B* 58 (1996), pp. 267–288.

[10] Hansheng Wang and Chenlei Leng. "Unified LASSO estimation by least squares approximation". In: *J Am Stat Assoc* 102 (2007), pp. 1039–1048. DOI: 10.1198/016214507000000509. URL: http://dx.doi.org/10.1198/016214507000000509.

[11] S. Wang et al. "Hierarchically penalized Cox regression with grouped variables". In: *Biometrika* 96.2 (2009), pp. 307–322.

[12] Daniela M. Witten and Robert Tibshirani. "Testing significance of features by lassoed principal components". In: *Ann Appl Stat* 2.3 (2008), pp. 986–1012.

reduction in false discovery rates over using a vector of t-statistics;borrowing strength across genes;"one would not expect a single gene to be associated with the outcome, since, in practice, many genes work together to effect a particular phenotype. LPC effectively down-weights individual genes that are associated with the outcome but that do not share an expression pattern with a larger group of genes, and instead favors large groups of genes that appear to be differentially-expressed.";regress principal components on outcome;sparse principal components
.

[13] Shifeng Xiong. "Some notes on the nonnegative garrote". In: *Technometrics* 52.3 (2010), pp. 349–361.

"... to select tuning parameters, it may be unnecessary to optimize a model selectin criterion repeatedly";natural selection of penalty function
.

[14] Hao H. Zhang and Wenbin Lu. "Adaptive lasso for Cox's proportional hazards model". In: *Biometrika* 94 (2007), pp. 691–703.

penalty function has ratios against original MLE;scale-free lasso
.

[15] Hui Zhou, Trevor Hastie, and Robert Tibshirani. "Sparse principal component analysis". In: *J Comp Graph Stat* 15 (2006), pp. 265–286.

principal components analysis that shrinks some loadings to zero
.

[16] Hui Zou and Trevor Hastie. "Regularization and variable selection via the elastic net". In: *J Roy Stat Soc B* 67.2 (2005), pp. 301–320.