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


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.


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.

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


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. regressing principal components on outcome; sparse principal components.


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


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


principal components analysis that shrinks some loadings to zero.