relaxnet and widenet: Extending the glmnet R package with Relaxation, Basis Expansions and Aggressive Cross-Validation

Stephan J. Ritter* Alan E. Hubbard†

Abstract
Two new R packages for prediction will be described. **relaxnet** applies the idea of the relaxed lasso (Meinshausen, 2007) to glmnet models (as provided by the **glmnet** R package, Friedman et al., 2010), leading to increased prediction accuracy in certain cases, and greater sparsity of the selected model. **widenet** adds the capability of applying polynomial basis expansion to the input data and then selecting a subset of the basis functions using **relaxnet**. The intent with both of these packages is for the user to make aggressive use of cross-validation to select tuning parameters, and this is encouraged by providing options to easily parallelize the execution over different parameter values and over cross-validation folds.

Key Words: R packages, penalized regression, cross-validation, basis expansion

1. Introduction

Meinshausen (2007) proposed the relaxed lasso, a modification of the lasso procedure (Tibshirani, 1996), which overcomes the slow convergence rates for sparse, high-dimensional problems. The **glmnet** R package (Friedman et al., 2010) provides efficient lasso and elastic net (Zou and Hastie, 2005) penalized generalized linear models. Here we introduce the **relaxnet** and **widenet** R packages, available from the Comprehensive **R** Archive Network (CRAN, `http://cran.r-project.org`). **relaxnet** provides an implementation of the relaxed lasso which is based on the **glmnet** package. This makes it more efficient than the existing **relaxo** package (Meinshausen, 2012), especially for large problems. **widenet** expands the size of the statistical model by adding polynomial basis expansions. A subset of the basis functions is selected using **relaxnet**, and cross-validation may be used to select the degree of basis expansion as well as the **relaxnet** tuning parameters.

2. The Relaxed Lasso

The relaxed lasso (Meinshausen, 2007) improves on the convergence rate of the lasso for sparse, high-dimensional data. It also leads to sparser models than standard lasso. An alternative method is the adaptive lasso of Zou (2006), which has similar performance to the relaxed lasso (Bühlmann et al., 2011). The set of relaxed lasso solutions, indexed by the parameters λ and ϕ, is a super set of the set of lasso solutions:

$$\hat{\boldsymbol{\beta}}^{relaxed} = \text{argmin}_{\boldsymbol{\beta}} \frac{1}{n} \sum_{i=1}^{n} (y_i - x_i^T \{\boldsymbol{\beta} \cdot \mathbf{1}_{\mathscr{M}_\lambda}\})^2 + \phi\lambda \sum_{j=1}^{p} |\beta_j|. \tag{1}$$

Here, y_i is the ith element of y and x_i is the ith row of X, for $i = 1, \ldots, n$; $\mathscr{M}_\lambda$ is the set of indices corresponding to predictors selected by the lasso at a certain λ value and $\mathbf{1}_{\mathscr{M}_\lambda}$ is the indicator function on that set, so that

*Division of Biostatistics, School of Public Health, University of California, Berkeley, 101 Haviland Hall, MC 7358, Berkeley, CA, 94720
†Division of Biostatistics, School of Public Health, University of California, Berkeley, 101 Haviland Hall, MC 7358, Berkeley, CA, 94720

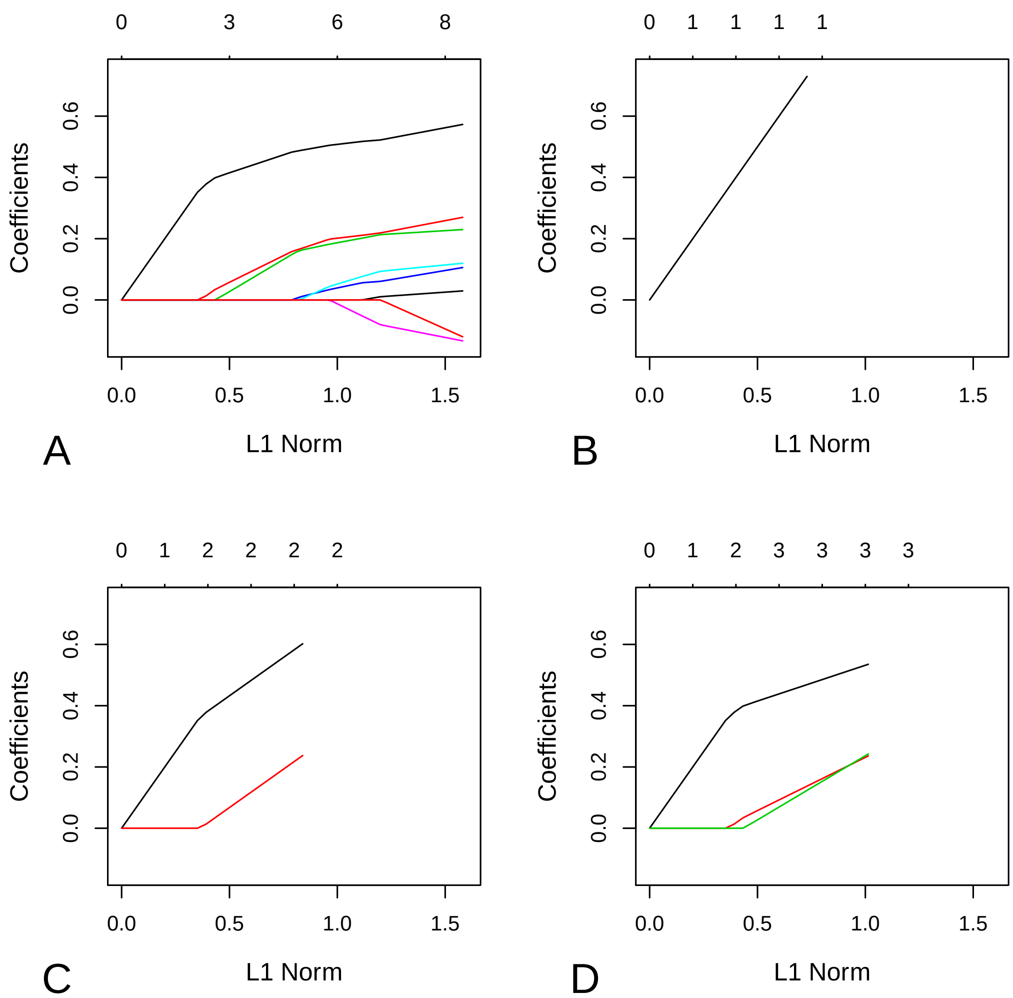

Figure 1: Relaxed lasso example. The data are from Stamey et al. (1989), made available through the **ElemStatLearn** R package (Hastie et al., 2009; Halvorsen, 2012). A: Full lasso coefficient profile. B-D: Coefficient profiles for the first three "relaxed models." Notice that the coefficients for the variables present in the relaxed models reach higher values than they do for the full lasso model.

$$\{\boldsymbol{\beta} \cdot \mathbf{1}_{\mathscr{M}_\lambda}\}_k = \begin{cases} 0, & k \notin \mathscr{M}_\lambda \\ \beta_k, & k \in \mathscr{M}_\lambda \end{cases}$$

(Meinshausen, 2007). One way to find the relaxed lasso solutions is to first find the lasso solutions for all λ, then run lasso again on each distinct subset of predictors from along the solution path. This allows the values of the coefficients for these subsets to "relax" back up from their lasso values, to the values of the OLS solution for just that subset. An example is given in figure 1. An R package implementing the relaxed lasso, **relaxo** (Meinshausen, 2012), is available on CRAN. This package depends on the **lars** package.

3. Advantages of the relaxnet Package

The **relaxnet** package takes advantage of the existing **glmnet** package (Friedman et al., 2010), in order to provide relaxed lasso functionality for generalized linear models. Applying relaxation to glmnet models offers several advantages over glmnet alone:

- Prediction error may be reduced when the data generating distribution is truly sparse, while the error will stay similar to that for the non-relaxed model otherwise
- The resulting model will usually be sparser, with a smaller false positive rate. This is true for a wide range of underlying true levels of sparsity of the data generating model (Ritter, 2013).
- Applying relaxation to elastic net penalized models can be especially effective, since the prediction accuracy improvements of elastic net over pure lasso may be retained without sacrificing the sparsity of the final model (i.e. one keeps the false positive rate low). As shown in Ritter (2013, Section 1.4.3), for the case of a sparse data-generating model with small correlated groups of truly contributing predictors, combining relaxation with the elastic net penalty can result in significant additional improvements in prediction error on top of the improvement due to using just relaxed lasso. Note that the adaptive lasso has also been applied to elastic net penalized models (Zou and Zhang, 2009)

The fact that **relaxnet** is based on **glmnet** offers several advantages over the **relaxo** package, which is based on the **lars** package:

- **glmnet** is based on a very efficient algorithm coded in Fortran, while **lars** is based on an earlier algorithm and is coded entirely in R. **relaxnet** therefore has a considerable speed advantage over **relaxo**, especially for when the number of predictors is large (Ritter, 2013, Section 1.4.4)
- **relaxnet** works for generalized linear models as well as linear models. The idea of applying relaxation to L_1 penalized generalized linear models was discussed in Meinshausen (2007, section 2.2). So far only linear (`family = "gaussian"`) and and binary outcome logistic (`family = "binomial"`) models have been implemented for **relaxnet**
- **relaxnet** allows application of elastic net penalties in addition to the lasso penalty

4. The widenet Package

The **widenet** package seeks to occupy a middle ground between linear methods and non-parametric methods such as random forests (Breiman, 2001). This is accomplished by using polynomial basis expansions. Sinisi and van der Laan (2004) introduced the Deletion/Substitution/Addition (DSA) algorithm, which uses a sequential search through subsets of basis functions combined with cross-validation to select a good subset from among those singled out during the sequential search. **widenet** provides a similar result, in that the final model may contain arbitrary basis functions which need not be nested (i.e. the term $x_1 : x_2$ may be in the model although x_1 and x_2 may not be). However, **widenet** expands the basis first and then runs **relaxnet** in order to select a subset. This avoids the sequential search through basis functions, and leads to a more easily parallelizable algorithm. Simulations and data analyses in Ritter (2013, Chapter 2) show that **widenet** provides improvement in prediction accuracy over the DSA.

References

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Bühlmann, P. L., van de Geer, S. A., and Van de Geer, S. (2011). *Statistics for high-dimensional data*. Springer-Verlag.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1.

Halvorsen, K. (2012). ***ElemStatLearn**: Data sets, functions and examples from the book: "The Elements of Statistical Learning, Data Mining, Inference, and Prediction" by Trevor Hastie, Robert Tibshirani and Jerome Friedman.* R package version 2012.04-0.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer-Verlag, New York, 2nd edition.

Meinshausen, N. (2007). Relaxed lasso. *Computational Statistics & Data Analysis*, 52(1):374–393.

Meinshausen, N. (2012). ***relaxo**: Relaxed Lasso*. R package version 0.1-2.

Ritter, S. J. (2013). *Software for Prediction and Estimation with Applications to High-Dimensional Genomic and Epidemiologic Data*. PhD thesis, University of California, Berkeley.

Sinisi, S. E. and van der Laan, M. J. (2004). Deletion/substitution/addition algorithm in learning with applications in genomics. *Statistical Applications in Genetics and Molecular Biology*, 3(1):Article 18.

Stamey, T., Kabalin, J., McNeal, J., Johnstone, I., Freiha, F., Redwine, E., and Yang, N. (1989). Prostate specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate. ii. radical prostatectomy treated patients. *The Journal of urology*, 141(5):1076.

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

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, 67(2):301–320.

Zou, H. and Zhang, H. H. (2009). On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics*, 37(4):1733.