

## LARS / LASSO / Forward Stagewise Regression

Least-angle regression (LARS) is a regression technique for high-dimensional data. It is related to both the Least Absolute Shrinkage and Selection Operator (LASSO) [3] and forward stagewise regression [2]. LARS was introduced by Efron *et al.* in [1] along with an algorithm that can efficiently fit all three type of models. Of these three regression techniques LASSO can be described the most compactly as follows:

Given a vector of  $n$  observed values,  $y = \{y_i : i = 1, 2, \dots, n\}$  and an  $n \times p$  design matrix  $X$ , where the  $j$ th column of  $X$ , denoted  $x_j$ , is a vector of length  $n$  representing the  $j$ th independent variable  $x_j$ , standardized such that  $\sum_{i=1}^n x_{ij} = 0$ , and  $\sum_{i=1}^n x_{ij}^2 = 1$  and a set of model parameters  $\beta$  to be estimated from the observed values, the LASSO model of [3] is given by

$$\underset{\alpha, \beta \in \mathbb{R}^p}{\text{minimize}} \|y - \alpha - X^T \beta\|^2 \quad \text{subject to} \quad \|\beta\|_1 \leq t \quad (1)$$

for a given value of  $t$ , where  $\alpha = \bar{y} = n^{-1} \sum_{i=1}^n y_i$ .

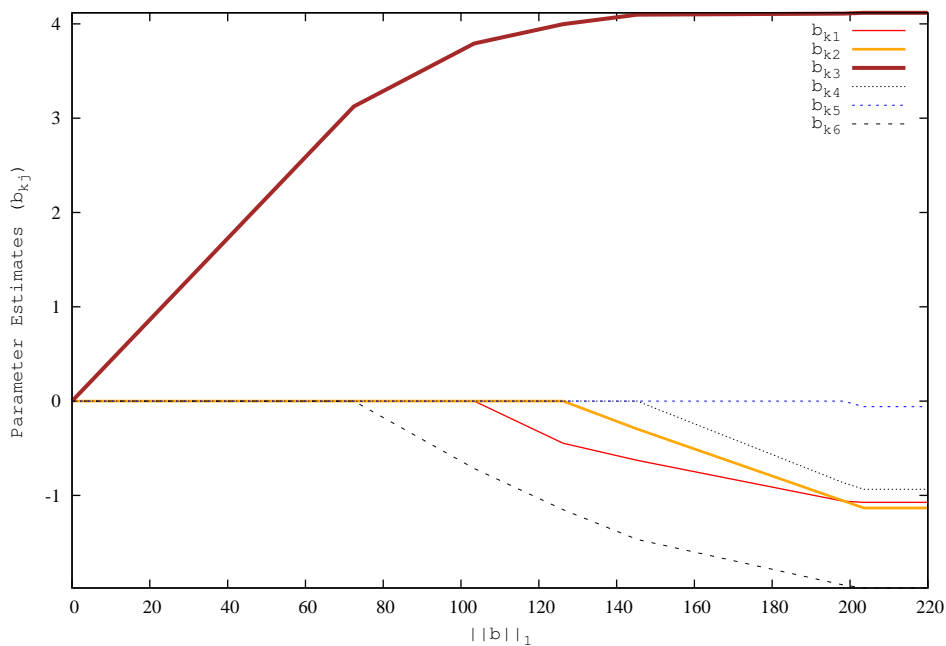


Figure 1: Plot of the parameter estimates,  $\beta_{kj}$ , against their  $L_1$ -norm,  $\|\beta\|_1$  for a LASSO model fit to simulated data.

Rather than solve equation (1) for a given value of  $t$ , the algorithm presented in [1] returns a full solution path for all possible values of  $t$ . It turns out that this path is piecewise linear with a finite number of pieces, denoted  $K$ , corresponding to  $K$  sets of parameter estimates. An example solution path, for a set of simulated data with 20 observations and 6 variables can be seen in Figure 1

---

At Mark 25, the algorithm (and adjustments) described in [1] has been introduced into Chapter G02 of the NAG Library allowing a LARS, LASSO or forward stagewise model to be fit to data. In addition a utility routine has been included that returns the parameter estimates at any arbitrary point along the solution path a requirement for, amongst other things, the use of cross-validation techniques.

## References

- [1] B Efron, T Hastie, I Johnstone, and R Tibshirani. Least Angle Regression. *The Annals of Statistics (Volume 32)*, 2:407–499, 2004.
- [2] T Hastie, R Tibshirani, and J Friedman. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer (New York), 2001.
- [3] R Tibshirani. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B (Methodological) (Volume 58)*, 1:267–288, 1996.