UQOP 2019
Uncertainty Quantification & Optimization Conference
18-20 March, Paris, France

An implementation of LASSO using optimization methods for piecewise differentiable functions

Tathagata Basu
Durham University, UK
tathagata.basu@durham.ac.uk

Matthias C. M. Troffaes
Durham University, UK
matthias.troffaes@durham.ac.uk

Jochen Einbeck
Durham University, UK
jochen.einbeck@durham.ac.uk

Keywords: LASSO, Sub-gradient, Proximal Gradient Descent, Co-ordinate Descent

ABSTRACT

We present and compare a novel R implementation of *LASSO* using different optimization methods for piecewise differentiable functions. LASSO (or "Least Absolute Shrinkage and Selection Operator") was introduced by Robert Tibshirani [Tibshirani, 1996]. This is one of the popular regularization methods because of its efficient variable selection property. The variable selection is achieved by adding a penalty term to the least square problem. However, this requires in the minimization of a piecewise differentiable objective function. We review three different optimization techniques to tackle this issue: the sub-gradient method, the proximal gradient descent method, and the co-ordinate descent method. The *sub-gradient* method [Shor, 1985] is a simple variant of the gradient descent method which requires a predetermined sequence of step sizes. If these step sizes are not chosen carefully, convergence can be excessively slow.. However, this is not a descent method in general and we must keep track of the best solution to achieve optimality. The *proximal gradient descent* method [Beck *et al.*, 2009] is applicable for a convex function, which is a sum of two convex functions, where one function is differentiable and the other is non-differentiable but has a simple analytical form. The co-ordinate descent method [Sauer *et al.*, 1993] successively minimizes a multivariable function along each co-ordinate. We compare the speed of convergence of these three methods using LASSO. In addition to this, we extend our implementation to the *adaptive LASSO* [Zou, 2006] (or weighted LASSO), as well as cross-validation or boot-strap method for frequentist model validation tools. We split a dataset into K parts for cross validation and then we fit model for $K-1$ part and test it with the remaining

part. We repeat this method for each part so that each data point is used once for model validation. For boot-strap we randomly sample data points from the dataset with replacement for a given number of times and examine the empirical distribution of the modeling parameters.

We introduce the idea of perturbed weights for assessing the sensitivity of regression parameters using the idea of adaptive LASSO. These weights on the penalty term of LASSO force the predictors to shrink earlier than the other depending on the scale of the weights, i.e. higher weight forces a predictor to shrink early whereas lower weight slows down the shrinkage. For the sensitivity of the predictors we use n number of purturbed weights and obtain the LASSO coefficients. We get a behavior of the predictors w.r.t. these weights. We observe that for these purturbed weights the standard deviations of the values of the predictors are often less for the non-important predictors obtained from the regular LASSO than that of the important ones.

References

[Beck *et al.*, 2009] Beck, A. and Teboulle, M. (2009) A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems *SIAM Journal on Imaging Sciences*, 2(1):183-202.

[Sauer *et al.*, 1993] Sauer K. and Bouman C. (1993) A local update strategy for iterative reconstruction from projections *IEEE Transactions on Signal Processing*, 41(2):534-548.

[Shor, 1985] Shor, N. Z. (1985) Minimization Methods for Non-Differentiable Functions *Springer Series in Computational Mathematics*, 3(1):22-47.

[Tibshirani, 1996] Tibshirani, Robert (1996). Regression shrinkage and selection via the lasso *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58(1):267-288.

[Zou, 2006] Zou, Hui (2006) The Adaptive Lasso and Its Oracle Properties *Journal of the American Statistical Association*, 101(476):1418-1429.