Penalized Regression: LASSO & Cousins

Emmanuel Ogundimu

General description: Penalized regression methods, particularly the LASSO (Least Absolute Shrinkage and Selection Operator), have gained significant attention in data science and statistical modelling. The method introduces a penalty term based on the absolute values of regression coefficients, encouraging sparsity in the model. By shrinking some coefficients to zero, LASSO effectively performs feature selection, identifying the most relevant predictors while discarding irrelevant ones. It has been extended to practically all statistical models and has been used for regularization in many well-known machine learning algorithms such as support vector machines. Consider a least squares (LS) regression problem. Suppose we are given a set of responses $y_i, \ldots, y_n$ and associated vectors of p predictors $\mathbf{x_i} \in \mathbf{R^p}$, where $i = 1, \ldots, n$. Then the LASSO estimator is given by

$$\arg\min_\beta \left\{ \sum_{i=1}^{n} \left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\},$$

where λ is the regularization parameter that handles a trade-off between goodness-of-fit and sparsity.

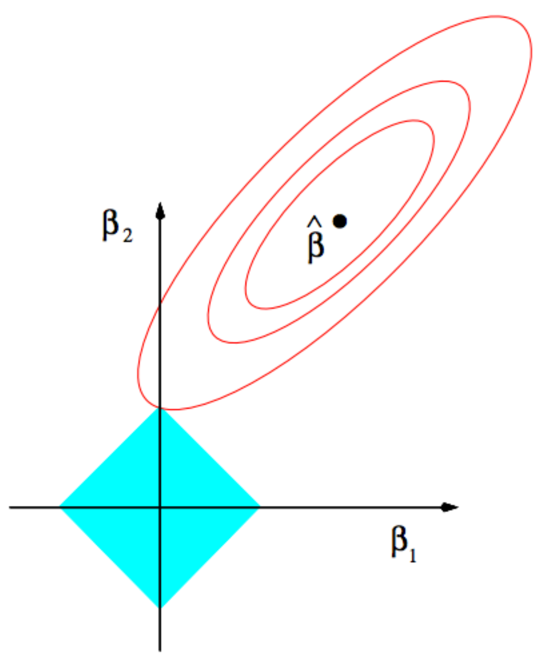

Figure 1: Geometric interpretation of Lasso via elliptical contours (Tibshirani, 1996)

Figure 1 shows the geometrical interpretation of the LASSO problem for LS regression. The contour lines depict the LS problem, and the diamond corresponds to the restriction enforced by the penalty $\sum_{j=1}^{p} |\beta_j|$. For a fixed value of λ, we aim to find the point on the diamond closest to the LS solution. This point is likely to lie on the axes, thus setting unnecessary parameters equal to zero.

LASSO has close cousins (methods with similar appearances to LASSO but with improved properties) such as Adaptive LASSO, Elastic net (a combination of LASSO and Ridge regression), Fused LASSO, Smooth LASSO, Group LASSO, Relaxed LASSO, Square-root LASSO, etc. It also has distant cousins such as BAR (broken adaptive ridge), MCP (minimax concave penalty), SCAD (smoothly clipped absolute deviation), SELO (seamless-L0), SICA (Smooth integration of counting and absolute deviation), etc.

Objectives: This project aims to introduce students to the inner workings of the algorithms developed for penalized regression for variable selection. Most importantly, it will explore the use of these methods for various data types such as continuous, binary, survival, and zero-inflated count data, with a focus on predictive accuracy in the context of predictive modelling and drawing causal inference in policy evaluation. Some potential directions for this project include:

- Analysis of various regularization parameter tuning methods for LASSO and its extensions via cross-validation, BIC (Bayesian information criterion), AIC (Akaike information criterion), GCV (generalized cross-validation), bootstrap, etc.
- Performance of LASSO-type methods in the presence of rare events in binary and survival data.
- Regularized missing data imputation.
- Variable selection in observational studies to address confounding and improve causal estimates.
- Statistical inference and post-selection inference.
- Ensemble and stacked LASSO for improved predictions.

Essential modules/Skills: Familiarity with the R statistical software and interest in data science.

References

1. Breheny, P. and J. Huang (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. Ann Appl Stat 5(1), 232–253
2. Freijeiro-González, L., M. Febrero-Bande, and W. González-Manteiga (2022). A critical review of LASSO and its derivatives for variable selection under dependence among covariates. International Statistical Review 90(1), 118–145.
3. Friedman J,Hastie T, Tibshirani RJ (2010) Regularization paths for generalized linear models via coordinate descent. J Stat Softw 33(1):1–22.
4. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Series B. 58(1), 267–288.
5. Vidaurre, D., Bielza, C. & Larranaga, P. (2013). A survey of L1 regression. Int. Stat. Rev., 81 361 387.