

plasso

Introduction

Michael Knaus
Stefan Glaisner

03/12/2023

This document is meant as a short overview over the *plasso* package by Knaus and Glaisner (2023).

1 General idea

Built on top of the *glmnet* library by Friedman et al. (2010), the *plasso* package follows Knaus (2022) and comes up with two main functions that both estimate least squares Lasso and Post-Lasso models. The *plasso()* function adds coefficient paths for a Post-Lasso model to the *glmnet()* output of a standard Lasso model using the same values for the penalty term λ .

On top of that, *cv.plasso()* cross-validates all Lasso and Post-Lasso models. The function output includes mean squared errors (MSEs) for all models, and, the most optimal λ hyperparameter value (according to minimum MSE or alternative definitions such as the one-standard-error-rule).

Both functions come with several S3 methods attached. *cv.plasso()* allows for a cross-validation focused *summary* output including the λ value and the names of the active set at the minimum. Both functions come with *predict* methods that can be used for prediction of both fitted values and coefficients - either across the whole λ sequence or for a specified λ value. The coefficients can also be obtained by using the *coef* method associated to both functions.

Interesting insights into the models can be inferred from the *plot* methods. For both functions, the *plot* method is quite similar to its respective *glmnet* counterpart. *plasso()* visualises the coefficient paths - either for the Lasso or Post-Lasso model depending on the specification of option *lasso* (default is *FALSE*). *cv.plasso()*, on the other hand, unifies both models by representing the MSEs and number of active coefficients along the whole sequence of λ values.

For an even more comprehensive overview including exemplary code, have a look at the package's [vignette](#).

2 Technical implementation

Heretofore, the code is entirely written in R. The first step in both `cv.plasso()` and `plasso()` is a `glmnet()` call for the estimation of the Lasso model and the extraction of the chosen sequence of lambda values. The code is written such as that the Post-Lasso model is newly estimated every time the active set changes. By doing so the computational burden is reduced as we do not have to estimate the Post-Lasso model for the whole lambda sequence.

3 Contact

For reporting a bug, simply [open an issue](#) on GitHub. For personal contact, you can write an email to michael.knaus@uni-tuebingen.de.

References

- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22.
- Knaus, M. and Glaisner, S. (2023). *plasso: Cross-validated Post-Lasso*. R package version 0.1.1.
- Knaus, M. C. (2022). Double machine learning-based programme evaluation under unfoundedness. *The Econometrics Journal*, 25(3):602–627.