Shrinkage priors for Bayesian penalized regression: An overview and tutorial using Stan.

<mark>Van Erp, S.</mark>[1*], Oberski, D.L.[2], Mulder, J.[3]

[1] Tilburg University, The Netherlands; PhD-student under supervision of Dr. Ir. J. Mulder, Dr. D.L. Oberski, and Prof. Dr. J.K. Vermunt

[2] Utrecht University, The Netherlands

[3] Tilburg University, The Netherlands

Suggested talk duration (15-60 minutes)

25 minutes

Summary (max. 500 words)

This presentation focuses on the problem of a small sample size, relative to the number of predictors in a regression model. If the sample size is smaller than the number of predictors, the model is not identified and cannot be estimated using traditional regression approaches, such as ordinary least squares. Penalized regression methods such as the lasso (least absolute shrinkage and selection operator; Tibshirani, 1996) and the ridge (Hoerl & Kennard, 1970) are well-known solutions to this identification problem. The central idea of penalized regression approaches is to add an ad hoc penalty term to the minimization problem that will shrink small coefficients towards zero. However, it can be difficult to obtain valid standard errors and to determine the penalty parameter, which is a central parameter in these methods. These problems can be solved using a Bayesian approach in a straightforward manner. In a Bayesian analysis, a prior distribution is specified for all parameters, which, combined with the likelihood of the data, results in a posterior distribution. It is well known that specific shrinkage priors imply penalties equivalent to classical penalization methods, such as the ridge and lasso. Moreover, certain Bayesian penalization methods have been shown to perform similarly to or better than classical penalization methods (Hans, 2009; Kyung et al., 2010; Li & Lin, 2010), and additionally

Bayesian methods straightforwardly result in credibility intervals with a clear Bayesian interpretation.

Due to these advantages, Bayesian penalization is becoming increasingly popular and many different prior distributions have been proposed that have desirable properties in terms of prediction and variable selection. However, the extensive technical Bayesian literature and subtle differences between the priors can make it difficult for researchers to navigate the options and make sensible choices for the problem at hand. The goal of this talk is to aid researchers in this endeavor by presenting an overview and comparison of the different prior options and, importantly, providing insight in the characteristics and behaviors of the priors. Two methods will be presented to determine the penalty parameter: a full Bayesian method and an empirical Bayes method. All priors have been implemented in the freely available software package Stan (Stan development team, 2017), and all code will be made available online so that researchers can easily use the different priors.

Relevance to conference theme

This presentation focuses on Bayesian penalization methods as a solution to samples which are small relative to the number of predictors. Given the extensive literature on Bayesian penalization, this presentation is especially relevant for applied researchers who wish to use Bayesian penalization since it will provide insight in the characteristics and behaviors of the many priors that are available. Although the focus lies mainly on the simple linear regression model, the methods presented can be used in more complex models as well, e.g., factor analysis, which is currently an active area of research.

Keywords (max. 3)

Penalized regression, shirinkage priors, Bayesian credibility intervals