MAP Estimation via Simulated Annealing for Sparse Bayesian Regression using MCMC sampling

Sudhir Raman
University of Basel, Switzerland
sudhir.raman@unibas.ch

Volker Roth
University of Basel, Switzerland
volker.roth@unibas.ch

Abstract

Problem Context and Motivation. Consider the context of a ℓ_1 penalized regression problem also referred to as the classical Lasso problem. The classical Lasso provides a sparse estimate but is unable to provide more information regarding the posterior of regression coefficients (like variance estimates). In this respect, a Bayesian approach is more advantageous since it is able to quantify the full posterior over the regression coefficients. Now consider the context of a real-world experiment setting (like biological experiments) where there's a constant feedback loop between the application group, which generates experimental data, and the computational group (which provides analysis for the data) and this in turn may trigger further experiments on the application side and so on. In such a scenario, there is a need to provide summarized information of the data analysis in a form that can be effectively used by the application side. It is desirable to have the information in a form that is simple to comprehend and easily interpretable. We can summarize the posterior information in traditional ways like providing an estimate of the first moment, but such estimates are not sparse and sparse point estimates can additionally convey the result of the analysis in a more comprehensible form to the application side. Hence, in addition to the traditional ways of summarizing information of the posterior of regression coefficients, we also need a sparse point estimate.

Proposed model and justification. This requirement to have additional information in the form of a sparse point estimate gives rise to the need for having one framework which can give us both the full posterior information and also additionally a sparse point estimate. We use an existing Bayesian framework for sparse regression defined by introducing auxiliary variables, which is necessary to make posterior inference feasible through MCMC sampling. It is also trivial to extend this framework in order to generate a MAP estimate by using simulated annealing. The annealing is again implemented using MCMC sampling since all the conditional posterior distributions are of standard form. It is however non-trivial to assume that this MAP estimate will also be sparse with respect to the regression coefficients. This is due to the fact that the optimization problem has changed from the original MAP estimation of the regression coefficients to the new MAP estimation of a joint distribution of regression coefficients and auxiliary variables. Apart from the extension of the Bayesian model for generating a MAP estimate, a key contribution of this work involves showing that such an estimate will also be sparse (with respect to the regression coefficients) and hence justifying the extension of this model to generate a MAP estimate.

Non-Convexity. Another problem associated with ℓ_1 penalized regression is that of having too many non-zero coefficients in the solution. This problem can be addressed by solving the regression problem with a flexible constraint of the form ℓ_p-norm with $p < 1$ using p to adjust the level of sparsity. This formulation makes the problem non-convex. We show that the same Bayesian framework can be used for solving this problem by defining a flexible prior which is capable of altering the level of sparsity desired in the solution and applying simulated annealing to generate a sparse point estimate.