Neuronized Priors - Initial Report

Chistopher Crabtree, Lanxin Yang, Yunshan Duan

November 11, 2021

Abstract

In linear regression settings it is often desirable to examine a reduced model by removing several of the predictor variables. Doing so can simplify subsequent analysis and/or increase efficiency of inference and prediction. In the Bayesian framework, selecting the best subset of features can be accomplished by using *shrinkage priors* on the regression coefficients that encourage them to shrink to zero. Several shrinkage priors have been previously proposed in the literature. These include *continuous shrinkage priors* (such as the horseshoe prior [1] and Bayesian Lasso [4]) as well as another class of priors known as *spike-and-slab* priors [2, 3].

Our assigned paper aims to unify these shrinkage priors under a simple framework called *neuronized priors* inspired by activation functions used in neural networks. The paper shows that many previous shrinkage priors have simple reductions to to neuronized priors. The unification of shrinkage priors has the benefit of reducing practical hurdles inherent in comparing performance among several classes of priors. Additionally, the paper provides efficient algorithms to sample from the posterior when using neuronized priors as well as theoretical results showing these algorithms attain an optimal posterior contraction rate and exponentially fast convergence for neuronized priors.

1 Progress

This paper compares several broad classes of shrinkage priors and for each of those, experiments are conducted using the *original* implementation of each prior as well as the *neuronized* version. We have therefore adopted a strategy of dividing work among our group members based on shrinkage prior classes. Each group member is taking the lead in replicating results of the *original* version of a single class of shrinkage priors. These results will then be compared to the *neuronized* versions of each sparse prior, which will be produced by a single group member. We chose this approach under the assumption that re-running our paper's code to produce the neuronized results would be a simple task.

Additionally, the paper contains two distinct groups of experiments: one conducted on *synthetic* data, and one on *real* data.

1.1 Original Version Progress

As our paper lies tangential to many previous works, much of our progress in replicating the original versions up to this point has been in background research and literature review. The ideas behind each prior are quite dense and the paper did not spend much time giving a thorough description of them. This made it difficult initially to grasp some sections of the paper and prevented us from starting our replication efforts as quickly as we would have liked.

Currently, we plan to start utilizing the referenced implementations of the original priors (in R packages) starting next week.

1.2 Neuronized Version Progress

We have recently made our first efforts to duplicate the results of the neuronized priors using a synthetic data generation method (briefly) discussed in the paper. We found that the code for the data generation was not included in the paper's code repository. Nevertheless, we have tried to duplicate the paper's description of the data generation, but currently our results using the neuronized priors based on this data differ markedly from what was reported in the paper.

At this point it is unclear if the error lies in: our data generation process, our sampling of the posteriors, or in the calculation of the reporting metrics (MSE and cosine similarity). We therefore plan on running our inference code on the real dataset, which will hopefully help us to at least assess whether the error lies in the synthetic data generation or the rest of our implementation.

References

[1] Carlos M Carvalho, Nicholas G Polson, and James G Scott. The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480, 2010.

[2] Edward I George and Robert E McCulloch. Variable selection via gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889, 1993.

[3] Toby J Mitchell and John J Beauchamp. Bayesian variable selection in linear regression. *Journal of the american statistical association*, 83(404):1023–1032, 1988.

[4] Trevor Park and George Casella. The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008.