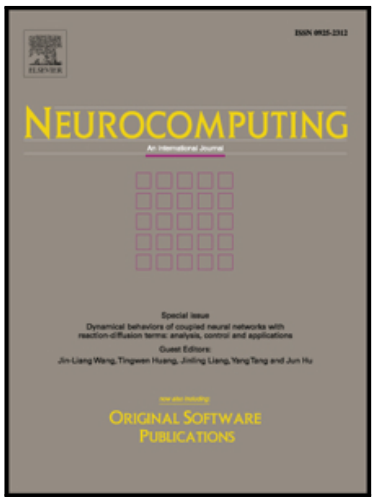

Sparse Bayesian linear regression with latent masking variables

Yohei Kondo, Kohei Hayashi, Shin-ichi Maeda

PII: S0925-2312(17)30422-8
DOI: 10.1016/j.neucom.2016.12.080
Reference: NEUCOM 18172

To appear in: *Neurocomputing*

Received date: 29 April 2016
Revised date: 1 December 2016
Accepted date: 12 December 2016

Please cite this article as: Yohei Kondo, Kohei Hayashi, Shin-ichi Maeda, Sparse Bayesian linear regression with latent masking variables, *Neurocomputing* (2017), doi: 10.1016/j.neucom.2016.12.080

Sparse Bayesian linear regression with latent masking variables

Yohei Kondo[a], Kohei Hayashi[b,c], Shin-ichi Maeda[a]

[a] *Graduate School of Informatics, Kyoto University, Kyoto, Japan*
[b] *National Institute of Advanced Industrial Science and Technology, Tokyo, Japan*
[c] *Kawarabayashi Large Graph Project, ERATO, JST*

Abstract

Extracting a small number of relevant features for the task, i.e., feature selection, is often a crucial step in supervised learning problems. Sparse linear regression provides a fast and convenient option for feature selection, where regularization facilitates reducing the weight parameters of irrelevant features. However, the regularization also induces undesirable shrinkage in the weights of relevant features.

Here, we propose Bayesian masking (BM) in order to resolve the trade-off problem between sparsity and shrinkage. Our strategy is not to directly impose any regularization on the weights; instead, BM introduces binary latent variables, called masking variables, into a regression model to keep the sparsity; each feature and sample has a binary variable whose value determines if the feature is masked or not at the sample. We derive a variational Bayesian inference algorithm for the augmented model based on the factorized information criterion (FIC), a recently-proposed asymptotic approximation of the marginal log-likelihood. We analyze the one-dimensional estimators of Lasso, automatic relevance determination (ARD), and BM, and thus show the superiority of BM in terms of the sparsity-shrinkage trade-off. Finally, we confirm our theoretical analyses through experiments and, demonstrate that BM achieves higher feature selection accuracy compared with Lasso and ARD.

Keywords: Sparse estimation, Factorized information criterion, Lasso, Automatic relevance determination