Discussion on the Paper "A Bayesian Information Criterion for Singular Models"

Pierre-Alexandre Mattei *pierre-alexandre.mattei@parisdescartes.fr*
Laboratoire MAP5, UMR CNRS 8145
Université Paris Descartes & Sorbonne Paris Cité
45 rue des Saints-Pères, 75006 Paris, France

Abstract

This is a contribution to the discussion on the paper *A Bayesian information criterion for singular models*, by M. Drton and M. Plummer [J. R. Stat. Soc. B **79** (2017), pp. 323–380, arXiv:1309.0911].

Drton and Plummer's paper has many impressive features: it clearly introduces Watanabe's theory and revisits it cleverly, it is faithful to the spirit of Schwarz's original article, and it provides an in-depth visit of possible applications of their new criterion. In short, this is a formidable piece of work.

I believe that the main attract of sBIC is that it can be considered a valuable marginal likelihood proxy. An important practical application of this uncertainty knowledge would be Bayesian model averaging (BMA). I present here a proof of concept that sBIC may be used for efficient model averaging.

Consider the reduced-rank regression framework (Example 2.2). Given a new covariate value, the BMA estimate of the response is a weighted average of the posterior means obtained for each model. The weights are posterior model probabilities, and may be replaced by BIC-based approximations (Hoeting et al., 1999). Here, I propose to use posterior model probabilities approximated by sBIC. The posterior means are approximated by maximum-likelihood predictions. The predictive performance is evaluated on three real data sets, "eyedata" (Scheetz et al., 2006), "feedstock" (Liebmann et al., 2009) and "vélibs" (Bouveyron et al., 2015). To obtain multivariate regression problems, the following preprocessing step was used. The variables were ranked according to the unsupervised feature selection technique of Bouveyron et al. (2016). The first 20 variables were considered as response and the 30 last were considered as covariates. The data are then split equally between training and test set and the performance is assessed (Table 1) using the mean-squared error (MSE). The code for this experiment is available from *http://pamattei.github.io*. Five estimators are considered: the ordinary least-squares estimator (OLSE) obtained with the full model, OLSEs obtained with models selected by BIC and sBIC, and two BMA estimators.

Table 1: MSE over 1000 replications

	OLSE	BIC	sBIC	BMA-BIC	BMA-sBIC
eyedata	10.8 (1.07)	8.67 (0.536)	8.67 (0.541)	8.67 (0.536)	**8.60 (0.584)**
feedstock	10.5 (1.72)	10.5 (1.53)	**9.79 (1.42)**	10.4 (1.52)	**9.79 (1.42)**
vélibs	14.9 (0.980)	16.5 (0.672)	14.7 (0.624)	16.4 (0.694)	**14.5 (0.612)**

The sBIC-based BMA estimator outperforms all other competitors, illustrating that sBIC provides a more reliable proxy for posterior probabilities than does BIC.

The main drawback of sBIC is that it can only be computed for models where learning coefficients (or some bounds) are available. Can the authors provide some insight on what models others than the ones mentioned in the article might be compatible with sBIC? In particular, would it be possible to exploit the work of Aoyagi (2009) and Aoyagi and Nagata (2012), to calibrate deep neural networks?

Acknowledgement

This discussion was written during a visit to University College Dublin funded by the Fondation Sciences Mathématiques de Paris (FSMP).

Additional references

M. Aoyagi and K. Nagata. Learning coefficient of generalization error in Bayesian estimation and Vandermonde matrix-type singularity. *Neural computation*, 24(6):1569–1610, 2012.

C. Bouveyron, E. Côme, and J. Jacques. The discriminative functional mixture model for a comparative analysis of bike sharing systems. *The Annals of Applied Statistics*, 9(4): 1726–1760, 2015.

C. Bouveyron, P. Latouche, and P.-A. Mattei. Bayesian variable selection for globally sparse probabilistic PCA. Technical report, HAL-01310409, 2016.

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: a tutorial. *Statistical science*, pages 382–401, 1999.

B. Liebmann, A. Friedl, and K. Varmuza. Determination of glucose and ethanol in bioethanol production by near infrared spectroscopy and chemometrics. *Analytica Chimica Acta*, 642 (1):171–178, 2009.

T. E. Scheetz, K.-Y. A. Kim, R. E. Swiderski, A. R. Philp, T. A. Braun, K. L. Knudtson, A. M. Dorrance, G. F. DiBona, J. Huang, and T. L. Casavant. Regulation of gene expression in the mammalian eye and its relevance to eye disease. *Proceedings of the National Academy of Sciences*, 103(39):14429–14434, 2006.