Sparse Methods for Automatic Relevance Determination and Applications to Dynamical Systems

Samuel H. Rudy* and Themistoklis P. Sapsis *

*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract. We consider several variations of automatic relevance determination (ARD), a Bayesian method for linear regression often yielding sparse models, and highlight applications in nonlinear system identification. We review the motivation for ARD and show analytically that it often fails to learn accurate sparsity patterns when applied to orthogonal systems of linear equations. We subsequently discuss two classes of methods, regularization based and thresholding based, which build on ARD to learn parsimonious solutions to linear problems. For each method we discuss the motivation and implicit assumptions used. In the case of orthogonal covariates we analytically demonstrate favorable performance with regards to learning the correct set of active terms in a sparse linear system when compared to ARD. We illustrate the relative merits of each of the discussed methods using several example problems including library based approaches to system identification.

Introduction

Motivated by an abundance of data collection across many fields and novel machine learning techniques, nonlinear system identification has been the subject of significant recent interest and research progress [1]. Many recent works have used sparse regression to select a parsimonious set of terms from a large library, yielding interpretable dynamical models. While most sparse regression methods have employed frequentest approaches, the desire for uncertainty quantification has driven interest in Bayesian methods for sparse regression and parameter estimation. Current approaches include Monte Carlo methods for sampling the posterior distribution of the dynamical model's parameters [2] and Bayesian linear regression [7, 5, 6, 3]. The latter approach, while less rigorous, is less computationally expensive and allows for selecting active terms simultaneously with approximating their posterior distribution. In this work we review several variations of Automatic Relevance Determination (ARD) [4], which is empirical Bayes for linear regression with a zero-mean Gaussian prior with diagonal covariance. Analytical comparisons between five variations of ARD are made for orthogonal problems and each variation is tested on system identification using the 40 dimensional Lorenz 96 system [8].

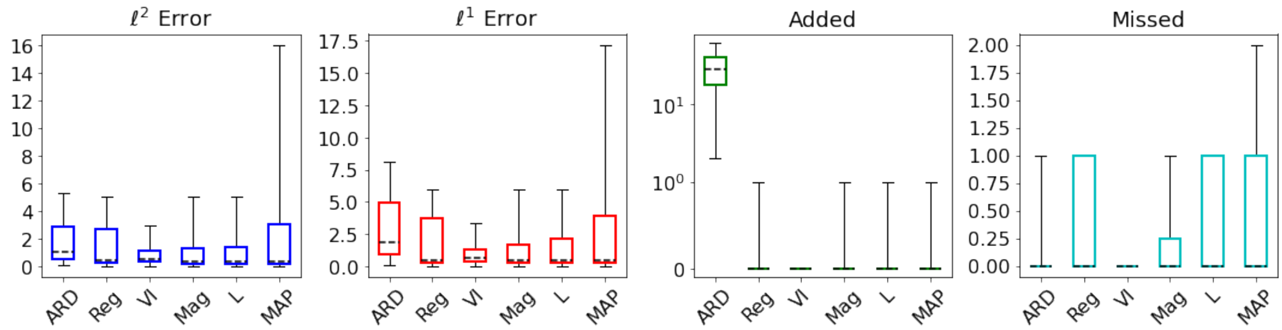

Figure 1: Error statistics for identification of the Lorenz 96 equations from data using the methods discussed in this work across each of 40 dimensions. Included are the ℓ^1 and ℓ^2 errors of the learned coefficients, the number of added terms (out of 857 possible) and the number of missing terms (out of four). Boxes/whiskers indicate inter-quartile range and full range of errors across 40 dimensions.

Results and Discussion

We test ARD and five variations including regularization (Reg), variance inflation (VI), magnitude based thresholding (Mag), likelihood based thresholding (L), and maximum-a-posteriori based thresholding (MAP) on a random sample of linear systems with fixed condition number, and for system identification of the 40 dimensional Lorenz 96 system using 861 library terms. Analytical expressions for the performance of each variant on orthogonal problems are also shown. It is demonstrated that the methods are equivalent with regards to expected sparsity on orthonormal problems. This work serves to compare and better explain five variations of ARD and to serve as a guide for their application to system identification and other sparse regression problems.

References

[1] Brunton, Proctor, and Kutz. *PNAS*, 113(15):3932–3927, 2016.
[2] Yang, Bhouri, and Perdikaris. *Proceedings of the Royal Society A* 476(2243): 20200290, 2020
[3] Rudy and Sapsis, *Physica D: Nonlinear Phenomena*, 132843. 2021.
[4] Wipf and Nagarajan. *Advances in neural information processing systems*, pages 1625–1632, 2008.
[5] Zhang and Lin *Proceedings of the Royal Society A*, 474(2217):20180305, 2018.
[6] Zhang and Lin *Journal of Computational Physics*, page 109962, 2020.
[7] Yuan et al. (2019). arXiv preprint arXiv:1909.06730.
[8] Lorenz. *Proc. Seminar on predictability*, volume 1, 1996.