Available online at www.sciencedirect.com
ScienceDirect
Digital Signal Processing 18 (2008) 236–248
Digital Signal Processing
www.elsevier.com/locate/dsp

Empirical Bayes linear regression with unknown model order ☆

Yngve Selén [a,*], Erik G. Larsson [b,1]

[a] *Department of Information Technology, Uppsala University, P.O. Box 337, SE-751 05 Uppsala, Sweden*
[b] *School of EE, Communication Theory, Royal Institute of Technology, Osquldas väg 10, SE-100 44 Stockholm, Sweden*

Available online 30 March 2007

Abstract

We study maximum a posteriori probability model order selection for linear regression models, assuming Gaussian distributed noise and coefficient vectors. For the same data model, we also derive the minimum mean-square error coefficient vector estimate. The approaches are denoted BOSS (Bayesian order selection strategy) and BPM (Bayesian parameter estimation method), respectively. In their simplest form, both BOSS and BPM require a priori knowledge of the distribution of the coefficients. However, under the assumption that the coefficient variance profile is smooth, we derive "empirical Bayesian" versions of our algorithms which estimate the coefficient variance profile from the observations and thus require little or no information from the user. We show in numerical examples that the estimators can outperform several classical methods, including the well-known AICc and BIC for model order selection.

Keywords: Linear regression; Empirical Bayes; MMSE estimation; Parameter estimation; Order selection

1. Introduction

1.1. Problem formulation

Consider the linear regression model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{\epsilon}, \tag{1}$$

where $\boldsymbol{y} \in \mathbb{R}^N$ is the vector of observed data, $\boldsymbol{X} = [\boldsymbol{x}_1 \dots \boldsymbol{x}_n] \in \mathbb{R}^{N\times n}$ is a known matrix of n regressors $\{\boldsymbol{x}_j\}_{j=1}^n$, $\boldsymbol{h} = [h_1 \dots h_n]^T \in \mathbb{R}^n$ is the unknown vector of linear regression coefficients[2] (we will also sometimes call $\boldsymbol{h}$ the *parameter vector*) and $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2\boldsymbol{I})$ is a length N vector of zero-mean Gaussian white noise with variance σ^2. We call (1) the *full model* and assume that the data are generated by a model of the form

$$\mathcal{M}_k\colon\ \boldsymbol{y} = \boldsymbol{X}_k\boldsymbol{h}_k + \boldsymbol{\epsilon}, \tag{2}$$

☆ This work was partly supported by the Swedish Research Council (VR). A conference version of this article has been accepted to ICASSP 2007. Parts of this paper were also presented at EUSIPCO 2006.

* Corresponding author. Fax: +46 18 511925.
E-mail addresses: yngve.selen@it.uu.se (Y. Selén), erik.larsson@ee.kth.se (E.G. Larsson).

[1] E.G. Larsson is a Royal Swedish Academy of Sciences Research Fellow supported by a grant from the Knut and Alice Wallenberg Foundation.

[2] The empirical Bayesian estimators to be presented in Section 3 require $n \leqslant N$, but the theory presented in Sections 1 and 2 does not need this assumption.

1051-2004/$ – see front matter
doi:10.1016/j.dsp.2007.03.005

where $n_{\min} \leqslant k \leqslant n$, $\boldsymbol{X}_k = [\boldsymbol{x}_1 \dots \boldsymbol{x}_k]$ (i.e., $\boldsymbol{X}_k$ consists of the first k columns of $\boldsymbol{X}$), and $\boldsymbol{h}_k = [h_1 \dots h_k]^T$. Furthermore, we make the assumption that the coefficients h_j are zero-mean independent Gaussian random variables, $h_j \sim \mathcal{N}(0, \gamma_j^2)$. In other words, $\boldsymbol{h}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Gamma}_k)$, where $\boldsymbol{\Gamma}_k = \operatorname{diag}[\gamma_1^2 \dots \gamma_k^2]$. We assume that the model order k and the variances $\{\gamma_j^2\}_{j=1}^k$, σ^2 are *unknown*.

We consider the following two classical interrelated problems:

(1) The model order selection problem: to find the correct order k, given $\boldsymbol{X}$ and $\boldsymbol{y}$.
(2) The parameter estimation problem: to estimate $\boldsymbol{h}$ as accurately as possible when the order k is unknown.

1.2. Related work

Bayesian solutions to the above two problems, under the Gaussianity assumption on the coefficients and the noise, are available in the literature. In, e.g., [1], the maximum likelihood (ML) and maximum a posteriori (MAP) model order selection algorithms for the current model were derived, although not numerically evaluated. In [2], the minimum mean square error (MMSE) estimate of a frequency function was derived. Within the same framework, it is easy to derive the MMSE estimate of $\boldsymbol{h}$. In [3,4] derivations of the maximum a posteriori (MAP) model order selection algorithm and the MMSE estimate of $\boldsymbol{h}$ were presented. These derivations will be the basis of the empirical Bayes method that we propose, and they will be summarized in Sections 2.1 and 2.2.

Classical (i.e., non-Bayesian, or "orthodox") approaches to model selection include the *information criteria*, which will be used for comparative purposes in our numerical examples in Section 4. However, they are not based on the Bayesian paradigm and are thus less related to the current approach than the methods discussed above. A short review of the information criterion approach is given in Appendix A.

The problem of determining the number of signal components in a mixture is fundamentally important also for many other types of data models. For example in array processing for determining the number of sources impinging on an array of sensors [5], or in line spectrum analysis for determining the number of sinusoids in noise [6]. In the present article, however, we are only concerned with the linear regression data model (1).

1.3. Contributions of this work

For the Bayesian approaches in the above subsection it is generally assumed that the noise variance σ^2 and the coefficient variances $\{\gamma_j^2\}_{j=1}^n$ are *known*. This assumption is hardly realistic in applications. The goal of the present article is to present methods which do not require knowledge of σ^2 and $\{\gamma_j^2\}_{j=1}^n$. To this end we take an empirical Bayes approach: we estimate σ^2, $\{\gamma_j^2\}_{j=1}^n$ from the data and then use the resulting estimates as if they were the true values; see Section 3.

Note that the models in (2) that we consider are *nested*, in the sense that a lower order model can be obtained as a special case of a higher order model (by setting certain coefficients to zero). This type of models is common in signal processing applications, such as finite-impulse-response (FIR) filter identification or the estimation of polynomial coefficients. By way of contrast, it is common in statistical data analysis to consider *sparse* models for which the true model can consist of *any* subset of the regressors $\{\boldsymbol{x}_k\}_{k=1}^n$. We have treated linear regression for sparse models in [3] where we, among other things, proposed an empirical Bayesian technique tailored to sparse models. However, for the nonsparse unknown model order problem, the technique in [3] is less sophisticated and flexible than the one we provide in this paper.

2. Bayesian model order selection and parameter estimation

2.1. Optimal model order selection

Here we review the MAP probability model order selection algorithm for the problem posed in Section 1, assuming known σ^2, $\{\gamma_j^2\}_{j=1}^n$. This model selection rule has been derived previously in [1]. In the remainder of the paper this specific model selection algorithm will be denoted by BOSS (Bayesian order selection strategy).