Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Computational and Applied Mathematics 176 (2005) 215–222
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

Ridge estimation of a semiparametric regression model ☆

Hongchang Hu[a,b]

[a] Department of Mathematics, Hubei Normal University, Huangshi 435002, China
[b] College of Mathematics and Statistics, Wuhan University, Wuhan 430072, China

Received 30 July 2003; received in revised form 4 June 2004

Abstract

Considered the semiparametric regression model

$$l_i = A_i^{\mathrm{T}} X + s(t_i) + \Delta_i \quad (i = 1, 2, \ldots, n).$$

Firstly, ridge estimators of both parameters and nonparameters are attained without a restrained design matrix. Secondly, the ridge estimator will be compared with two steps estimation under a mean square error and some conditions in which the former excels the latter are given. Finally, the validity and feasibility of the method are illustrated by a simulating example.

Keywords: Semiparametric regression model; Ridge estimation; Two steps estimation; Mean square error

1. Introduction

Considered the semiparametric regression model

$$l_i = A_i^{\mathrm{T}} X + s(t_i) + \Delta_i \quad (i = 1, 2, \ldots, n), \tag{1a}$$

where $s_i = s(t_i)$ denotes the nonparametric signal of the observation and l_i denotes a number related to the observation at t_i, $A_i \in R^p (n > p)$, $X = (x_1, \ldots, x_p)^{\mathrm{T}}$ is a parameter vector with p denoting

☆ Supported by National Natural Science Foundation of China (No. 40274005), and Big Item of Education Office, Hubei Province (No. 2001Z06003).

E-mail address: retutome@163.com (H. Hu).

0377-0427/$ - see front matter © 2004 Published by Elsevier B.V.
doi:10.1016/j.cam.2004.07.032

the number of parameters or unknowns, and Δ_i denotes the noise and is assumed to be independently $N(0, \sigma^2)$-distributed.

In vector notation, the data model is given by

$$L = AX + S + \Delta, \tag{1b}$$

where $L = (l_1, \ldots, l_n)^{\mathrm{T}}$ and $S = (s_1, s_2, \ldots, s_n)^{\mathrm{T}}$correspond to $t = (t_1, \ldots, t_n)^{\mathrm{T}}$ ($t_i \neq t_j$ for $i \neq j$), design matrix $A = (A_1, \ldots, A_n)^{\mathrm{T}}$ without any restrained conditions, namely, *rank* $(A) < p$ or *rank* $(A) = p$ (ill-conditioned or not).

The model (1) has been used in the discussion of many methods, e.g., penalized least-squares (see [1]), smoothing splines (see [2]), piecewise polynomial (see [6]) and two steps estimation methods (see [5,7,8]). The essential thought of two steps estimation is the following: the first step, $S(t, X)$ is defined with supposition where X is supposed to be known; the second step, the estimator of parametric X is attained by a least-squares method; accordingly, $\hat{S}(t) = S(t, \hat{X})$ is gained. However, they all assume *rank* $(A) = p$. In fact, if full rank A is an ill-conditioned matrix, then the results may not fulfil our wishes, or can even be false in some situations, especially for small samples. Many papers do not consider the case *rank* $(A) < p$, and few people investigate the situation that the design matrix A is rank-deficient.

Although there are many results about ridge estimation of linear models (see [3,4,9]), to the best of my knowledge, nothing is known about a semiparametric regression model. It is noticeable that textual ridge estimation not only solves rank-deficient and ill-conditioned problems, but also offers a new method which can deal with (non)linear and semiparametric regression models for $rank(A) = p$ without ill-conditioning.

2. Ridge estimation method

In the following, one introduces ridge estimation method based on a two steps estimation process.

In the first step, we assume that X is known, and the nonparametric estimator of S is defined by

$$S(t, X) = W(t, \lambda)(L - AX), \tag{2}$$

based on $\{l_i - A_i^{\mathrm{T}} X, t_i\}(i = 1, \ldots, n)$, where λ is an arbitrary parameter and $W(t, \lambda)$ is an $(n \times n)$ matrix. Depending on the particular choice of $W(t, \lambda)$, the two steps estimation process leads to different methods, such as wavelet estimate (see [8]), near neighbour estimation (see [5]), or kernel estimation (see [7]).

Substituting (2) into (1), we have

$$\tilde{L} = \tilde{A}X + \tilde{\Delta}, \tag{3}$$

where

$$\tilde{A} = (I - W)A, \quad \tilde{L} = (I - W)L, \quad \tilde{\Delta} = \tilde{S} + (I - W)\Delta, \quad \tilde{S} = (I - W)S. \tag{4}$$

Though (3) is a linear model, it is different from the generic one because the error $\tilde{\Delta}$ is related to S, t, X and W.

In the second step, with minimal condition

$$V^{\mathrm{T}}V + \beta \hat{X}^{\mathrm{T}} \hat{X} = \min \quad (V = \tilde{A}\hat{X} - \tilde{L}), \tag{5}$$