L_1-Statistical Procedures and Related Topics
IMS Lecture Notes – Monograph Series (1997) Volume 31

Dimension reduction via parametric inverse regression

Efstathia Bura

The George Washington University, Washington, USA

Abstract: In this paper, a linear subspace containing part or all of the information for the regression of a m-vector Y on a p-vector X and its dimension are estimated via the means of inverse regression. Smooth parametric curves are fitted to the p inverse regressions through a multivariate linear model, without imposing any strict assumptions on the error distribution. This method is expected to be more powerful in reducing the dimension of a regression problem when compared to SIR, the estimation procedure proposed by Li (1991), that is based on fitting piecewise constant functions to the inverse regression curves.

Key words: Dimension reduction, regression, linear subspace estimation.

AMS subject classification: 62A99, 62H05.

1 Introduction

Let $Y \in \mathrm{R}^m$ and $X \in \mathrm{R}^p$ with joint cumulative distribution function (c.d.f.) $F(Y, X)$. In a regression setting the behavior of the conditional cumulative distribution function of Y given X, $F(Y|X)$, as the value of X varies in its marginal sample space is under study. As a means of characterizing the regression structure, consider replacing X by $k \leq p$ linear combinations of its components, $\eta_1^T X, \ldots, \eta_k^T X$, without losing information on $F(Y|X)$ so that, for all values of X,

$$F(Y|X) = F(Y|\eta_1^T X, \ldots, \eta_k^T X) = F(Y|\eta^T X) \tag{1}$$

where η is the $p \times k$ matrix with columns η_j, and $F(\cdot|\cdot)$ denotes the conditional c.d.f. of the first argument given the second. Equation (1) holds trivially when $\eta = I_p$, where I_p denotes the identity matrix of dimension