Statistical Science
2007, Vol. 22, No. 1, 36–39
DOI: 10.1214/088342307000000050
Main article DOI: 10.1214/088342306000000682

Comment: Fisher Lecture: Dimension Reduction in Regression

Lexin Li and Christopher J. Nachtsheim

1. INTRODUCTION

Professor Cook is to be congratulated for his groundbreaking work in dimension reduction in regression. The paper develops a general theoretical foundation for studying principal components and other dimension reduction methods in a regression context. This framework yields a basis for elucidating the strengths, weaknesses and relationships among the various dimension reduction methods, including ordinary least squares (OLS), principal components regression (PCR), sliced inverse regression (SIR), parametric inverse regression and partial least squares. The promising new method, principal fitted components (PFC), appears to outperform some long-standing approaches such as PCR, OLS and SIR. Finally, as a result of this contribution, the standard approach to regression, with its emphasis on fixed predictors and the need to assume away the randomness of $\mathbf{X}$, and the standard approach to principal components, with its focus on the correlation matrix rather than the covariance matrix, both seem to be under question.

Specific contributions of Professor Cook's paper include the following: (1) It provides a theoretical foundation for the widely used principal components regression. (2) It resorts to a model and thus a likelihood function, through the inverse regression of predictors given response, to study sufficient reduction in a forward regression problem. Consequently, likelihood-based inferences can be developed, and the inferential capabilities of dimension reduction are moved closer to mainstream regression methodology. (3) It permits extension to categorical or mixtures of continuous and categorical predictors, an area that most existing model-free dimension reduction approaches do not handle effectively. (4) It seems to be applicable to problems where the number of predictors exceeds the number of observational units, which is perhaps one of the most challenging current frontiers in statistical methodology. In general, we believe that this paper has paved the way for substantive research in dimension reduction, and that it will surely be the subject of much future application and elaboration.

In what follows, we focus on two issues raised but not thoroughly addressed in Professor Cook's paper: (1) the role of predictor screening and a connection with the recently proposed supervised principal components method and (2) dimension reduction in the presence of binary predictors.

2. SUPERVISED PRINCIPAL COMPONENTS

Professor Cook has suggested a combination of predictor screening and principal fitted components analysis when the number of predictors p is large—in particular when $n < p$. As Professor Cook noted, traditional studies have based predictor screening on the univariate forward regressions of Y on individual X_j, $j = 1, \dots, p$. In view of the results presented in this paper, it appears that the screening at the outset should probably be based on the univariate inverse regressions of X_j on $\mathbf{f}_y$. This in turn suggests the following algorithm for PFC in conjunction with predictor screening:

1. Compute the inverse univariate regressions X_j on $\mathbf{f}_y$, for $j = 1, \dots, p$.
2. Form a reduced matrix $\mathbf{X}_\theta$ composed of only those predictors whose regression coefficient on $\mathbf{f}_y$ is determined to surpass a level-of-significance threshold θ.
3. Obtain $\widehat{\boldsymbol{\Gamma}}$ using PFC based on the reduced $\mathbf{X}_\theta$.
4. Pass $\widehat{\boldsymbol{\Gamma}}$ to the forward regression.

Recently, Bair et al. (2006) proposed an approach called supervised principal components (SPC), which can be described as follows:

1. Compute the forward univariate regressions Y on individual X_j, $j = 1, \dots, p$.

Lexin Li is Assistant Professor, Department of Statistics, North Carolina State University, Box 8203, Raleigh, North Carolina 27695, USA (e-mail: li@stat.ncsu.edu). Christopher J. Nachtsheim is Professor, School of Statistics, University of Minnesota, 224 Church Street S. E., Minneapolis, Minnesota 55455, USA (e-mail: cnachtsheim@csom.umn.edu).