

Comment: Fisher Lecture: Dimension Reduction in Regression

Ronald Christensen

I am pleased to participate in this well-deserved recognition of Dennis Cook’s remarkable career.

Cook points out Fisher’s insistence that predictor variables in regression be chosen without reference to the dependent variable. Reduction by principal components clearly satisfies that dictum. One of my primary objections to partial least squares regression when I first encountered it as an alternative to principal components was that the predictor variables were being chosen with reference to the dependent variable. (I now have other objections to partial least squares.) Yet on the other hand, variable selection in regression is well accepted and it clearly chooses variables based on their relationship to the dependent variable. Perhaps variable selection is better thought of as a form of shrinkage estimation rather than as a process for choosing predictor variables.

Cook also reiterates something that I think is difficult to overemphasize: Fisher’s point that “More or less elaborate forms [for models] will be suitable according to the volume of the data.” We see this now on a regular basis as modern technology provides larger data sets to which elaborate models are regularly fitted.

With regard to Cook’s work, it seems to me that the key issue in the development of Cook’s models (2), (5), (10) and (13) is whether they are broadly reasonable. The question did not seem to be extensively addressed but Cook shows that much can be gained if we can reasonably use them. When they are appropriate, the results in the corresponding propositions are rather stunning. It has long been known that the best regression model available—technically the best predictor of a random variable y based on a p -dimensional random vector x —is the conditional mean $E(y|x)$. The problem with this result is that it requires us to know the joint distribution of (x', y) . Most of what we commonly recognize as regression analysis is an attempt to

model the relationship $E(y|x)$. This includes linear regression, nonlinear regression, generalized linear models and the various approaches to “nonparametric” (actually, highly parametric) regression. Under the models being considered, there exists a $p \times d$ matrix Γ such that

$$y|x \sim y|\Gamma'x.$$

This means that $E(y|x) = E(y|\Gamma'x)$ regardless of what modeling strategy we choose to use. If anything, this dimensionality reduction from p to d is of more importance to nonparametric regression than other forms because, as the number of predictor variables increases, nonparametric regression gets hit harder by the curse of dimensionality than less highly parametric forms. As a result, nonparametric regression should benefit most from the existence of a generally valid reduction in dimensionality.

The issue with these four models is to estimate the column space of Γ , say, $C(\Gamma)$. In the first six sections, the results are all closely tied to the eigenvectors (principal component vectors) of some estimated covariance matrix for the predictor variables x , say $\hat{\Sigma}$. For model (2), the space is spanned by the first d principal component vectors of the usual $\hat{\Sigma}$. For model (5), the space is spanned by the first d principal component vectors of a restricted version of $\hat{\Sigma}$. For models (10) and (13), the estimation procedure is a bit more complicated. The key is that for both models (10) and (13) the population covariance matrix of x can be written as

$$\Sigma = \Gamma V D V' \Gamma + \Gamma_0 V_0 D_0 V_0' \Gamma_0,$$

with D and D_0 diagonal matrices, in such a way that

$$\Sigma(\Gamma V) = (\Gamma V) D, \quad \Sigma(\Gamma_0 V_0) = (\Gamma_0 V_0) D_0.$$

This implies that the eigenvectors of Σ are either in $C(\Gamma)$ or in $C(\Gamma_0) \equiv C(\Gamma)^\perp$, the orthogonal complement of $C(\Gamma)$. The problem is to establish which d out of the p orthogonal eigenvectors belong in $C(\Gamma)$. To estimate $C(\Gamma)$, find the orthogonal eigenvectors of $\hat{\Sigma}$, say, v_1, \dots, v_p , and check the likelihood of every one of the p choose d combinations that has d of the v_i ’s

Ronald Christensen is Professor, Department of Mathematics and Statistics, University of New Mexico, Albuquerque, New Mexico 87131-0001, USA (e-mail: fletcher@stat.unm.edu).