Linear Algebra and its Applications 473 (2015) 316–338

Contents lists available at ScienceDirect

Linear Algebra and its Applications

www.elsevier.com/locate/laa

Computing Frechet derivatives in partial least squares regression

Lars Eldén [1]

Department of Mathematics, Linköping University, SE-58183 Linköping, Sweden

ARTICLE INFO

Article history:
Received 30 September 2013
Accepted 8 September 2014
Available online 17 September 2014
Submitted by Lek-Heng Lim

MSC:
62J05
65F10

Keywords:
Partial least squares
PLS
Regression
Least squares
Prediction
Golub–Kahan bidiagonalization
Krylov method
Frechet derivative
Recursion
Perturbation theory
Degrees of freedom

ABSTRACT

Partial least squares is a common technique for multivariate regression. The procedure is recursive and in each step basis vectors are computed for the explaining variables and the solution vectors. A linear model is fitted by projection onto the span of the basis vectors. The procedure is mathematically equivalent to Golub–Kahan bidiagonalization, which is a Krylov method, and which is equivalent to a pair of matrix factorizations. The vectors of regression coefficients and prediction are non-linear functions of the right hand side. An algorithm for computing the Frechet derivatives of these functions is derived, based on perturbation theory for the matrix factorizations. From the Frechet derivative of the prediction vector one can compute the number of degrees of freedom, which can be used as a stopping criterion for the recursion. A few numerical examples are given.

E-mail address: lars.elden@liu.se.
[1] Tel.: +46 13 282183.

http://dx.doi.org/10.1016/j.laa.2014.09.017

1. Introduction

Partial least squares regression (PLSR) [1,2] is a frequently applied technique for multivariate regression in the case when the explaining variables (predictor variables) are highly correlated. It iteratively constructs an orthonormal sequence of latent components (basis vectors) from the explaining variables, which have maximal covariance with the response variable. In each step of the procedure, the data and the solution vectors are projected onto subspaces of low dimension, where a linear model is fitted. PLSR can be used as an alternative to principal components regression (PCR), and often a good fit is obtained with a model of considerably smaller dimension than with PCR, see, e.g., [3].

The PLS procedure is mathematically equivalent to a Krylov method, Golub–Kahan bidiagonalization [4,5]. While the so-called NIPALS variant of PLS [5] constructs the basis vectors by successively deflating the data matrix (the predictor variables) and the right hand side (the response variable), the Krylov method generates them by a recursion without modifying the data matrix, see e.g. [3]. The Krylov recursion is equivalent to a pair of matrix factorizations.

A basic problem in PLSR is to determine the "optimal" number of components, i.e. to derive a stopping criterion for the recursion. There are two alternatives, essentially. The standard approach is to use cross validation. Alternatively, in [6] an information criterion is applied and the complexity of the fitted model is defined as the number of degrees of freedom (DOF).

Let $y \in \mathbb{R}^m$ be a vector of observations of the response variable, and $X \in \mathbb{R}^{m\times n}$ be a matrix, whose columns are the observations of the explaining variables. Consider the least squares problem

$$\min_{\beta} \|X\beta - y\|, \tag{1}$$

to which an approximate solution is computed by PLS. Denote the solution after k steps of PLS by β_k, and the prediction by $y_k = X\beta_k$. It turns out that y_k and β_k are non-linear functions of y; we write $y_k = F_k(y)$ and $\beta_k = H_k(y)$. The number of degrees of freedom of the model, D_k, is defined

$$D_k = 1 + \operatorname{tr}\left(\frac{\partial F_k}{\partial y}\right) = 1 + \operatorname{tr}\left(X\frac{\partial \beta_k}{\partial y}\right), \tag{2}$$

where $\partial F_k/\partial y$ is the Frechet derivative of the function. Note that, with $\bar{y} = y + \epsilon\delta y$ a perturbed data vector,

$$\|y_k - \bar{y}_k\| \le \epsilon\left\|\frac{\partial F}{\partial y}\right\|\|\delta y\| + O\left(\epsilon^2\right).$$

Thus the Frechet derivative defines a condition number of the function, which is a measure of the sensitivity to perturbations in the data.