Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computational Statistics & Data Analysis 50 (2006) 446–462
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Local minima in categorical multiple regression

Anita J. van der Kooij[a,*], Jacqueline J. Meulman[a], Willem J. Heiser[b]

[a] *Department of Educational Sciences, Data Theory Group, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands*
[b] *Department of Psychology, Psychometrics Group, Leiden University, The Netherlands*

Received 6 November 2003; received in revised form 27 August 2004; accepted 30 August 2004
Available online 18 September 2004

Abstract

CATREG is a program for categorical multiple regression, applying optimal scaling methodology to quantify categorical variables, including the response variable, simultaneously optimizing the multiple regression coefficient. The scaling levels that can be applied are nominal, nonmonotonic spline, ordinal, monotonic spline or numerical. When ordinal or monotonic spline scaling levels are applied, local minima can occur. With ordinal or monotonic spline scaling levels, the transformations are required to be monotonically increasing, but this can also be achieved by reflecting a monotonic decreasing transformation. A monotonic transformation is obtained by restricting a nonmonotonic transformation, but the direction of the monotonic restriction (increasing or decreasing) is undefined, and it will be shown that this is the cause of local minima. Several strategies to obtain the global minimum for the ordinal scaling level will be presented. Also, results of a simulation study to assess the performance of these strategies are given. The simulation study is also used to identify data conditions under which local minima are more likely to occur and are more likely to be severe. It was found that local minima more often occur with low to moderately low R^2 values, with higher number of categories and with higher multicollinearity.

Keywords: Alternating least squares; Backfitting; Categorical data; CATREG; Local minima; Linearization; Monotonic; Multiple regression; Nonlinear regression; Nonmonotonic; Optimal scaling; Quantification; Multiple systematic starts; Transformation

* Corresponding author. Tel.: +31-71-527-3827; fax: +31-71-527-3865.
E-mail address: kooij@fsw.leidenuniv.nl (A.J. van der Kooij).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2004.08.009

1. Introduction

With numerical data multiple regression is the most often used method to predict a dependent or response variable from a set of predictor variables. CATREG is a nonparametric method to perform multiple regression when data are categorical or a mix of numerical and categorical variables. CATREG allows for nonlinear transformations of the variables, including the response variable. CATREG can also be used with numerical data to explore the existence of nonlinear relationships. The program is available in SPSS (SPSS, 1998; Van der Kooij and Meulman, 1999).

Transformation of variables has become an important tool in data analysis over the last decades. Various models and programs have been developed, for example the Box–Cox model (Box and Cox, 1964), using parametric families of transformations. More general methods include monotone transformations (Kruskal, 1965), MORALS (Young et al., 1976) (implemented in TRANSREG (SAS/STAT, 1989)), spline transformations (Winsberg and Ramsay, 1980; Ramsay, 1988), Projection Pursuit Regression (Friedman and Stuetzle, 1981), ACE (Breiman and Friedman, 1985), and GAIM (Hastie and Tibshirani, 1990) for fitting generalized additive models using smoothers; see also Gaudart et al. (2004) for a comparison of Neural Networks and linear regression.

CATREG applies the optimal scaling methodology as developed in the Gifi system (Gifi, 1990) to quantify categorical variables, simultaneously optimizing the multiple regression coefficient. In the quantification process, information in the observed variable is retained in the quantified variable. The kind of information that is retained, and thereby the form of the transformation, depends upon the scaling level. The numerical scaling level results in a linear transformation, that is, the data are treated as numerical, and are only transformed into standardized variables. The non-numerical scaling levels allow for nonlinear transformations: the nominal and nonmonotonic spline scaling levels allow for nonmonotonic transformations; the ordinal and monotonic spline scaling levels allow for monotonic transformations. The scaling level can be chosen for each variable separately.

When all or some of the scaling levels are of the ordinal or monotonic spline type, local, nonglobal minima (suboptimal solutions) can occur. Ordinal or monotonic spline quantifications are obtained by applying a restriction to unrestricted (nominal, nonmonotonic) quantifications. In this paper, we argue that the occurrence of local minima in CATREG is due to the fact that the direction of the restriction is undefined. CATREG is closely related to MORALS (Young et al., 1976), TRANSREG (SAS/STAT, 1989) and ACE (Breiman and Friedman, 1985). With both these methods we have observed local minima as well. MORALS, TRANSREG, and ACE also use monotonic restriction of nonmonotonic transformations to obtain monotonic transformations, so it is very likely that the cause of local minima in MORALS and ACE is the same as in CATREG.

In this paper we present a strategy to obtain the global minimum, using systematic multiple starts. That is, each start uses the same initial values for the quantifications, but with different signs. Applying a negative sign results in a monotonic increasing transformation that is equal to a reflected decreasing transformation. Using all possible sign patterns will yield the global minimum. However, the number of possible sign patterns is a power of two, where the power is the number of predictor variables with ordinal or monotonic spline scaling level. So, for each additional predictor the number of systematic multiple starts is doubled. To reduce the