Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 52 (2008) 2129–2157
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Semi-parametric nonlinear regression and transformation using functional networks

Enrique Castillo[a], Ali S. Hadi[b,*], Beatriz Lacruz[c], Rosa E. Pruneda[d]

[a]Department of Applied Mathematics and Computational Sciences, University of Cantabria and University of Castilla La Mancha, Spain
[b]Department of Mathematics, The American University in Cairo, 113 Kasr El-Aini Street, P.O. Box 2511, Cairo 11511, Egypt
[c]Department of Statistical Methods, University of Zaragoza, Spain
[d]Department of Mathematics, University of Castilla-La Mancha, Spain

Received 14 August 2006; received in revised form 6 July 2007; accepted 10 July 2007
Available online 19 July 2007

Abstract

Functional networks are used to solve some nonlinear regression problems. One particular problem is how to find the optimal transformations of the response and/or the explanatory variables and obtain the best possible functional relation between the response and predictor variables. After a brief introduction to functional networks, two specific transformation models based on functional networks are proposed. Unlike in neural networks, where the selection of the network topology is arbitrary, the selection of the initial topology of a functional network is problem driven. This important feature of functional networks is illustrated for each of the two proposed models. An equivalent, but simpler network may be obtained from the initial topology using functional equations. The resultant model is then checked for uniqueness of representation. When the functions specified by the transformations are unknown in form, families of linear independent functions are used as approximations. Two different parametric criteria are used for learning these functions: the constrained least squares and the maximum canonical correlation. Model selection criteria are used to avoid the problem of overfitting. Finally, performance of the proposed method are assessed and compared to other methods using a simulation study as well as several real-life data.

Keywords: Adjusted correlation coefficient; Akaike information criterion; Alternating conditional expectation; Canonical correlation; Constrained least squares; Minimum description length measure; Transforming both sides

1. Introduction

The problem of investigating the functional relationship between a *response* variable Y and one or more *predictor* variables $X_1, X_2, \ldots, X_k$ is of interest. In many practical situations, the form of such a functional relationship is unknown. Additionally, discovering such a relationship may also require transformation of the response and/or the predictor variables. The problem of transformation in regression modeling has attracted considerable attention in the literature. See, for example, Atkinson (1985), Carroll and Ruppert (1988), and the references therein.

In this paper, we are mainly interested in solving two problems:

* Corresponding author. Tel.: +202 2797 5185; fax: +202 2797 7565.
E-mail addresses: castie@unican.es (E. Castillo), ahadi@aucegypt.edu (A.S. Hadi), lacruz@unizar.es (B. Lacruz), rosa.pruneda@uclm.es (R.E. Pruneda).

0167-9473/$ - see front matter
doi:10.1016/j.csda.2007.07.006

Problem 1: Transforming prediction variables (modeling). When the goal of the analysis is to model Y as a function of $X_1, X_2, \ldots, X_k$, the transformation involves only the predictor variables. Here we assume that the relationship between the response and predictor variables can be expressed as

$$Y = h(X_1, X_2, \ldots, X_k) + \varepsilon, \quad (1)$$

where ε is a random *error* whose expected value is assumed to be 0. Then, the problem is to discover the structure of the function h in (1).

Problem 2: Transforming response variable. In some applications transformation may involve both the response and predictor variables. In this case, we assume that the relationship between the response and predictor variables can be written as

$$f(Y) = h(X_1, X_2, \ldots, X_k) + \varepsilon, \quad (2)$$

where ε is a random *error* whose expected value is assumed to be 0. Our task here is to discover the structure of transformations f and h in (2).

There are several ways for the formulation and estimation of the models in (1) and (2). On the one hand, one may first consider the classical *multiple linear regression*

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon, \quad (3)$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are the *parameters* to be estimated from the data. Then, if needed, a nonlinear transformations of the response and/or the explanatory variables are commonly used in regression problems to achieve certain desirable conditions such as linearity, normality, and homoscedasticity. In this case, (3) can be written as

$$f(Y) = \beta_0 + \beta_1 h_1(X_1) + \beta_2 h_2(X_2) + \cdots + \beta_k h_k(X_k) + \varepsilon, \quad (4)$$

and now the problem consists of estimating the parameters $\beta_0, \beta_1, \ldots, \beta_k$ and to discover the transformations $h_1, h_2, \ldots, h_k$, and f (some of them can be the identity function).

The models in (3) and (4) are parametric models and they have considerable success in a variety of applications, despite their rigid linear form for h in (1) or (2). These forms, however, may not be appropriate for many more complex situations.

On the other hand, nonparametric methods, which make minimal assumptions about the form of h in (1) or (2), can be used. For example, a nonparametric *additive model* is obtained by replacing the linear function $\beta_j X_j$ in (3), $j = 1, \ldots, k$, by a nonlinear function $h_j(X_j)$ to get

$$Y = h_1(X_1) + h_2(X_2) + \cdots + h_k(X_k) + \varepsilon, \quad (5)$$

or, when nonlinear transformation of the response is necessary,

$$f(Y) = h_1(X_1) + h_2(X_2) + \cdots + h_k(X_k) + \varepsilon. \quad (6)$$

This generalization retains some important features of the linear approximation: It is additive in the explanatory variables effect and it allows us to separately examine the roles of the explanatory variables in modeling the response. The nonlinear functions h_j, $j = 1, \ldots, k$, in (5) are *smooth* functions and an iterative procedure is needed to fit the model. Also, more complex than simple component-wise additive functions can be used in the additive model (see, e.g., Hastie and Tibshirani, 1990; Friedman and Stuetzle, 1981).

When the transformation of the response variable is also needed, the function f in (6) can also assumed to be an arbitrary smooth function. In this case, techniques for fitting the model in (6) include *alternating conditional expectation* (ACE) algorithm (Breiman and Friedman, 1985) and *additivity and variance stability* (AVAS) transformation procedure (Tibshirani, 1988).

Other nonparametric techniques such as *piecewise* and *local* multiple regression, *regression trees* and *multivariate adaptive regression splines* (MARS) tackle the approximation of h in (1) by fitting several simple parametric functions in different pieces of the observed domain (see, for example, Friedman, 1991; Fox, 2000; and the references therein).