Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Computational Statistics & Data Analysis 49 (2005) 757 – 770
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Inverse boosting for monotone regression functions☆

Yuwon Kim[a,*], Ja-Yong Koo[b]

[a] *Statistical Research Center for Complex Systems, Seoul National University, Kwan Ak Gu, Sillim dong, Seoul 151-742, Republic of Korea*
[b] *Department of Statistics, Inha University, Incheon 402-751,Republic of Korea*

Received 25 June 2003; received in revised form 24 May 2004; accepted 31 May 2004
Available online 2 July 2004

Abstract

A new method for the estimation of smooth monotone regression functions is proposed. It is assumed that the monotonicity may come from some physical or economic reason. A monotone estimator of an integral (of a positive function) using a gradient boosting method is derived. The proposed method generates a sequence of fits without monotone constraints and combines them to form a monotone estimate. An advantage of the proposed algorithm is that one can use a popular smoothing technique without the monotone constraint as the base learner. The performance of the proposed procedure is demonstrated on both simulated and real data sets.

Keywords: Base learner; Gradient boosting; Inverse problems; Monotonicity; Nonparametric regression; Regression spline

1. Introduction

This paper proposes a new method for the estimation of a smooth monotone regression function, where the monotone assumption may be due to some physical or economic reason. For example, Fig. 1 shows the height of a single child over a 312-day period. A purpose of

☆ This research was supported by KOSEF through The Statistical Research Center for Complex Systems at Seoul National University.

* Corresponding author. Tel.: 82168442812; fax: 8228885834.
E-mail address: gary@stats.snu.ac.kr (Y. Kim)

0167-9473/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2004.05.038

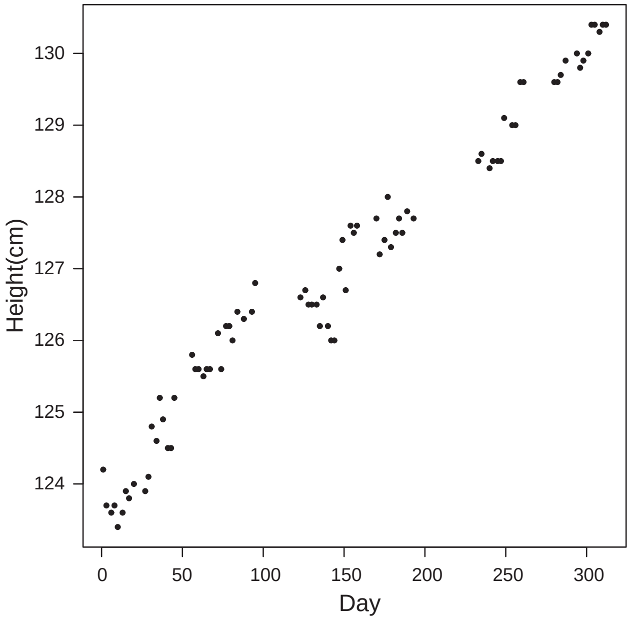

Fig. 1. Height of a child recorded over a 312-day period.

such a study maybe to estimate the dependence of height as a function of day, where it is reasonable to assume monotonicity.

A summary of existing methods is as follows. Based on isotonic regression, Barlow et al. (1972) proposed a technique which provides a monotone pointwise estimate at the observed points. Friedman and Tibshirani (1984) introduced a hybrid method which combines the local average smoothing and isotonic regression; see also Delecroix et al. (1993) and Mukerjee (1988). Villalobos and Wahba (1987) and Kelly and Rice (1990) used constrained optimization for smoothing splines and regression splines, respectively. Mammen (1991b) proposed a least-squares procedure in a class of functions satisfying some moderate smoothness conditions and a monotone restriction. The solution turned out to be a least-squares spline with knot points that depend on the observations. Ramsay (1988a) considered a regression spline estimator, where the coefficients of the B-splines used in the estimator are constrained to be non-negative; recently, Ramsay (1998b) studied a different method based on a characterization of monotone functions through a differential operator. Delecroix et al. (1996) introduced a two-step procedure consisting of a smoothing step followed by an isotonizing step; in the first step, a smoothing spline estimate or a convolution-type kernel estimate can be used and the second step is a projection in a Hilbert space onto the cone of restrictions. He and Shi (1998) adopted a monotone B-spline smoothing method based on L_1 optimization. For more details of recent works, see Delecroix and Thomas-Agnan (2000).

We formulate the problem of estimating a monotone regression function as an inverse problem. To make a regression function monotone, one can define the regression function