Computational Statistics and Data Analysis 81 (2015) 89–106

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

A convex version of multivariate adaptive regression splines

Diana L. Martinez [a], Dachuan T. Shih [b], Victoria C.P. Chen [c,*], Seoung Bum Kim [d]

[a] TMAC, 202 E. Border St. Ste. 323 Arlington, TX 76010, USA

[b] Conifer Health Solutions, 2401 Internet Boulevard Frisco, TX 75034, USA

[c] Department of Industrial and Manufacturing Systems Engineering, The University of Texas at Arlington, Campus Box 19017 Arlington, TX 76019, USA

[d] School of Industrial Management Engineering, Korea University, Seoul, Republic of Korea

HIGHLIGHTS

- Convex-MARS enables a convex approximation without degrading the quality of fit.
- Convex-MARS is appropriate for approximations in convex optimization problems.
- The threshold version of Convex-MARS provides stronger convexity and better accuracy.

ARTICLE INFO

Article history:
Received 23 May 2012
Received in revised form 24 July 2014
Accepted 24 July 2014
Available online 4 August 2014

Keywords:
Regression splines
Convexity

ABSTRACT

Multivariate adaptive regression splines (MARS) provide a flexible statistical modeling method that employs forward and backward search algorithms to identify the combination of basis functions that best fits the data and simultaneously conduct variable selection. In optimization, MARS has been used successfully to estimate the unknown functions in stochastic dynamic programming (SDP), stochastic programming, and a Markov decision process, and MARS could be potentially useful in many real world optimization problems where objective (or other) functions need to be estimated from data, such as in surrogate optimization. Many optimization methods depend on convexity, but a non-convex MARS approximation is inherently possible because interaction terms are products of univariate terms. In this paper a convex MARS modeling algorithm is described. In order to ensure MARS convexity, two major modifications are made: (1) coefficients are constrained, such that pairs of basis functions are guaranteed to jointly form convex functions and (2) the form of interaction terms is altered to eliminate the inherent non-convexity. Finally, MARS convexity can be achieved by the fact that the sum of convex functions is convex. Convex-MARS is applied to inventory forecasting SDP problems with four and nine dimensions and to an air quality ground-level ozone problem.

1. Introduction

Computer modeling is having a profound effect on scientific research. Many processes are so complex that physical experimentation is too time-consuming, too expensive or simply impossible. As a result, experiments have increasingly turned to mathematical models to simulate these complex systems. Advances in computational power have allowed both greater complexity and more extensive use of such models. The purpose of design and analysis of computer experiments (DACE, Sacks

* Corresponding author.
E-mail address: vchen@uta.edu (V.C.P. Chen).

http://dx.doi.org/10.1016/j.csda.2014.07.015

et al., 1989; Kleijnen, 2008; Chen et al., 2006) is to provide methods for conducting computer experiments to build a metamodel that can be efficiently employed to improve the performance of a complex system. In DACE, the computer experiment replaces the physical experiment by organizing computer model runs and observing the model output of performance. A common DACE objective is to obtain a computationally-efficient response surface approximation (a.k.a., metamodel) of the output. This metamodel may then be used to study and potentially "optimize" the performance of the system. The effectiveness of an optimization method in using a metamodel to improve system performance depends on the convexity of the objective function (Luenberger, 2004). A non-convex metamodel requires a global optimization method, and in practice these typically cannot guarantee optimality. Consequently, if the true underlying performance objective function is known to be convex, it is highly desirable for the approximating metamodel to share this critical property.

Multivariate adaptive regression splines (MARS, Friedman, 1991) modeling has been applied in DACE-based approaches for some large-scale optimization problems, including continuous-state stochastic dynamic programming (SDP, Chen, 1999; Chen et al., 1999; Tsai et al., 2004; Tsai and Chen, 2005; Cervellera et al., 2007; Yang et al., 2007, 2009), Markov decision processes (MDP, Chen et al., 2003; Siddappa et al., 2007, 2008), and two-stage stochastic programming (SP, Pilla et al., 2008, 2012; Shih et al., 2014). The DACE-based SDP and MDP approaches used an experimental design to discretize the continuous (or near-continuous) state space, and then used MARS to approximate the continuous value function over the state space. The MDP application studied an airline revenue management problem with the objective of more accurately estimating the fair market value of a seat over time. The two-stage SP problem studied an airline fleet assignment model that seeks an assignment of aircraft in the first stage, so that swapping of crew-compatible aircraft can be achieved in the second stage to maximize expected revenue. The DACE approach for SP was used to create a MARS approximation of the first-stage expected revenue objective function, so as to speed up the first-stage optimization. MARS has been successful in these applications not only because of the flexibility of its modeling, but also its parsimony. Parsimony is critical in achieving computational-tractability in large-scale complex problems. Shih et al. (2014) added a data mining variable selection phase that reduced the dimension of the airline fleet assignment model from about 1200 to 400 variables prior to executing DACE, so as to reduce the computational effort of DACE from 2.5 days to an estimated 0.5 days.

Under the assumption that an optimization function f is convex, it is desired that the response surface metamodel $\hat{f}$ that estimates f be convex as well. For example, in the above-mentioned SDP, MDP, and SP problems, the underlying function is theoretically convex. Convexity is not a typical assumption of statistical modeling methods, and a specialized approach must be developed. There are several options for DACE metamodeling, including polynomial response surface models (Box and Draper, 1987), spatial correlation models, a.k.a., kriging (Sacks et al., 1989), MARS, regression trees (Breiman et al., 1984; Friedman, 2001), and artificial neural networks (Haykin, 1999). None of these guarantee convexity. Convex-MARS uses the modification of both the MARS basis functions and algorithms to build a sum of convex functions, therefore, the final approximation will be convex. The C code is available from this website: http://www.uta.edu/cosmos/software.php.

2. Multivariate adaptive regression splines (MARS)

Friedman (1991) introduced MARS as a statistical method for high-dimensional modeling with interactions. The MARS model is essentially a linear statistical model with a forward stepwise algorithm to select model terms followed by a backward procedure to prune the model terms. A univariate version (appropriate for additive relationships) was presented by Friedman and Silverman (1989). The MARS approximation bends to model curvature at "knot" locations, and one of the objectives of the forward stepwise algorithm is to simultaneously select variables and appropriate knots. After selection of the basis functions is completed, smoothness to achieve a certain degree of continuity may be applied. MARS is both flexible and easily implemented with the computational effort primarily dependent on the number of basis functions added to the model. The MARS approximation is a linear model:

$$\hat{f}_M(\boldsymbol{x};\beta) = \beta_0 + \sum_{m=1}^{M} \beta_m B_m(\boldsymbol{x}),$$

where $B_m(\boldsymbol{x})$ initially is a basis function of the form described below in Eq. (1) that later can be smoothed, M is the number of linearly independent basis functions, and β_m is the unknown coefficient for the mth basis function. In the forward stepwise algorithm, univariate basis functions are represented in the form of truncated linear functions,

$$b^+(x;k) = [+(x-k)]_+, \qquad b^-(x;k) = [-(x-k)]_+, \tag{1}$$

where $[q]_+ = \max\{0, q\}$ and k is a univariate knot. The set of eligible knots are assigned separately for each input variable dimension and are chosen to coincide with input levels represented in the data. Interaction basis functions are created by multiplying an existing basis function with a truncated linear function involving a new variable. Both the existing "parent" basis function and the newly created interaction basis function are used in the MARS approximation. Thus, the form of the mth basis function is

$$B_m(\boldsymbol{x}) = \prod_{l=1}^{L_m} [s_{l,m} \cdot (x_{v(l,m)} - k_{l,m})]_+,$$

where $x_{v(l,m)}$ is the input variable corresponding to the lth truncated linear function in the mth basis function, $k_{l,m}$ is the knot value corresponding to $x_{v(l,m)}$, and $s_{l,m}$ is $+1$ or -1. L_m is the number of truncated linear functions multiplied in the mth basis