Journal of Computer and System Sciences 91 (2018) 69–81

Contents lists available at ScienceDirect

Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Weighted L_∞ isotonic regression

Quentin F. Stout

Computer Science and Engineering, University of Michigan, Ann Arbor, MI 48109-2121, United States

ARTICLE INFO

Article history:
Received 24 March 2012
Received in revised form 16 July 2013
Accepted 23 June 2017
Available online 14 September 2017

Keywords:
Isotonic regression
Monotonic
Unimodal
L_∞
Tree
Linear order
Dag

ABSTRACT

Algorithms are given for determining weighted L_∞ isotonic regressions satisfying order constraints given by a directed acyclic graph with n vertices and m edges. An $\Theta(m \log n)$ algorithm is given, but it uses parametric search, so a practical approach is introduced, based on calculating prefix solutions. For linear and tree orderings it yields isotonic and unimodal regressions in $\Theta(n \log n)$ time. Practical algorithms are given for when the values are constrained to a specified set, and when the number of different weights, or different values, is $\ll n$. We also give a simple randomized algorithm taking $\Theta(m \log n)$ expected time. L_∞ isotonic regressions are not unique, so we examine properties of the regressions an algorithm produces. In this regard the prefix approach is superior to algorithms, such as parametric search and the randomized algorithm, which are based on feasibility tests.

1. Introduction

A directed acyclic graph (dag) $G = (V, E)$ with n vertices and m edges defines a partial order over the vertices, where $u \prec v$ if and only if there is a path from u to v. A function g on V is *isotonic* iff it is a weakly order-preserving mapping into the real numbers, i.e., iff for all $u, v \in V$, if $u \prec v$ then $g(u) \le g(v)$. By *weighted data* on G we mean a pair of real-valued functions (f, w) on V, where w, the *weights*, is non-negative, and f, the *values*, is arbitrary. Given weighted data (f, w) on G, an L_p *isotonic regression* is an isotonic function g on G that minimizes

$$\begin{cases} \left(\sum_{v \in V} w(v) \cdot |f(v) - g(v)|^p\right)^{1/p} & 1 \le p < \infty \\ \max_{v \in V} w(v) \cdot |f(v) - g(v)| & p = \infty \end{cases}$$

among all isotonic functions. The L_p *regression error* is the value of this expression. For $1 < p < \infty$ the regression values are unique, but for L_1 and L_∞ they may not be. For example, on the vertices $\{1, 2, 3\}$, if $f = 3, 1, 2.5$ and $w = 2, 2, 1$: for $1 < p < \infty$ the L_p isotonic regression is $g(1) = g(2) = 2$, $g(3) = 2.5$, L_1 isotonic regressions are of the form $g(3) = 2.5$ and $g(1) = g(2) = x$ for $x \in [1, 2.5]$, and L_∞ isotonic regressions are of the form $g(1) = g(2) = 2$ and $g(3) \in [2, 4.5]$. In the example in Fig. 1 note that the regression values form level sets and that the regression is undefined in some regions.

Isotonic regression has long been applied to a wide range of problems in statistics [5,21,38,49] and classification [14,16,44], with numerous recent applications to machine learning and data mining [27,28,34,35,37,50]. Routines for it are available in several machine-learning systems [39,40]. It is also used for some optimization problems [6,29]. It is of increasing importance as researchers reduce their assumptions, replacing parametric requirements with weaker assumptions about an underlying direction. For example, to shrink tumors via radiation and chemotherapy, it might be assumed that at any

E-mail address: qstout@umich.edu.

http://dx.doi.org/10.1016/j.jcss.2017.09.001
0022-0000/

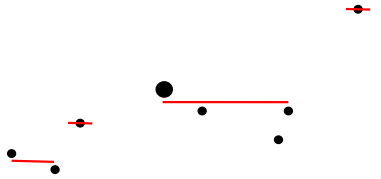

Fig. 1. An isotonic regression.

given radiation level the shrinkage increases with dose, and at any given dose the shrinkage increases with radiation. However, there may be no assumptions about a low dose and high radiation combination vs. a high dose and low radiation. Independent variables need not have a metric, merely an ordering, such as S < M < L < XL sizes.

Isotonic functions are also known as monotonic, monotonic increasing, or order-preserving (as a function from G to $\Re$), and the L_∞ metric is also known as the supremum norm, Chebyshev distance, uniform metric, minimax optimization, or bottleneck criterion.

We develop algorithms for finding L_∞ isotonic regressions for weighted data. In Section 3 we give an algorithm for general dags taking $\Theta(m \log n)$ time. While conceptually simple, based on using parametric search as a black box procedure, it is of only theoretical interest because parametric search is very complex and infeasible. Therefore we give practical alternatives for several important special cases. E.g., when the number of different weights or different values is $\ll n$ a practical search can be used (Section 4.1).

L_∞ isotonic regressions are not unique, raising questions as to whether some are better than others. Previous algorithms have been compared only in terms of their running time, not in their mathematical properties when viewed as a mapping from data to isotonic functions. In Section 2 we examine properties such as minimizing the number of vertices with large regression errors, looking beyond just minimizing the worst error. By these measures the parametric search approach produces the worst regressions, lacking several properties that are shared by all L_p regressions for $1 < p < \infty$. In contrast, Strict, introduced in Section 2.1.3, has many of these properties. This topic is pursued further in [43].

Section 2.1.1 introduces Prefix, which is used in Section 6 to find isotonic regression for linear and tree orders and river isotonic regression on trees, where the regression value of a node is the same as the largest of its children. Unimodal regression on linear and tree orders is also examined, where the objective is to determine an optimal root and an isotonic regression towards the root. This arises in optimization problems such as competing failure modes [17,23,24] and multi-arm bandits [25,51].

Section 4.2 considers isotonic regression where the regression values are restricted to a specific set, such as the integers, using algorithms that are quite different from those in the other sections. An $\Theta(n \log n)$ algorithm is given for finding an integer-valued isotonic regression for a linear order, improving upon the $\Theta(n^2)$ algorithm of Liu and Ubhaya [33]. Section 7 contains concluding remarks.

Finally, Appendix A gives a very practical algorithm for arbitrary dags, replacing parametric search with a simple randomized procedure. It takes $\Theta(m \log n)$ expected time, not the worst-case time considered elsewhere in the paper, achieving this with high probability.

2. Background

Throughout we assume that G is a single connected component, and hence $m \geq n-1$. If it has more than one component then isotonic regressions can be found for each component separately.

The complexity of determining an isotonic regression depends on the regression metric and the partially ordered set. For example, for a linear order it is well-known that a simple left-right scanning approach using pair adjacent violators, PAV, can be used to determine the L_2 isotonic regression in $\Theta(n)$ time, L_1 in $\Theta(n \log n)$ time [3,42], and L_∞ on unweighted data in $\Theta(n)$ time. In Section 6 an algorithm taking $\Theta(n \log n)$ time is given for L_∞ isotonic regression on weighted data.

L_∞ isotonic regression on an arbitrary dag (poset) can easily be done in $\Theta(m)$ time for unweighted data (see Section 2.1.1), but for weighted data the fastest previously known algorithm is $\Theta(n \log^2 n + m \log n)$, due to Kaufman and Tamir [29]. Theorem 1 reduces this to $\Theta(m \log n)$.

Given weighted data (f, w) on dag $G = (V, E)$, for vertex $v \in V$, the *error of using r as the regression value at v*, $\mathrm{err}(v, r)$, is $w(v) \cdot |f(v) - r|$. Given vertices u and v, the weighted mean of their values, $\mathrm{mean}(f, w : u, v)$, is $[w(u)f(u) + w(v)f(v)] / [w(u) + w(v)]$. Note that $\mathrm{mean}(f, w : u, v)$ has equal error for the two vertices. Let $\mathrm{mean\_err}(f, w : u, v)$ denote this error.

There is a simple geometric interpretation of mean and mean_err: when regression values are plotted horizontally and the error is plotted vertically, the ray with x-intercept $f(u)$ and slope $-w(u)$ gives the error for using a regression value below $f(u)$ at u, and the ray with x-intercept $f(v)$ and slope $w(v)$ gives the error for using a regression value above $f(v)$ at v. The regression value where these lines intersect is $\mathrm{mean}(f, w : u, v)$, and the error of the intersection point is $\mathrm{mean\_err}(f, w : u, v)$. See Fig. 2.