Computers and Operations Research 95 (2018) 172–193

Contents lists available at ScienceDirect

Computers and Operations Research

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

Locating hyperplanes to fitting set of points: A general framework

Víctor Blanco[a,*], Justo Puerto[b], Román Salmerón[c]

[a] Dpt. Quant. Methods for Economics & Business and IEMath-GR, Universidad de Granada, Spain
[b] Dpt. Statistics & OR and IMUS, Universidad de Sevilla, Spain
[c] Dpt. Quant. Methods for Economics & Business, Universidad de Granada, Spain

ARTICLE INFO

Article history:
Received 17 November 2016
Revised 21 March 2018
Accepted 22 March 2018
Available online 23 March 2018

2010 MSC:
90B85
90C26
52C35
65D10

Keywords:
Fitting hyperplanes
Mathematical programming
Location of structures
Robust fitting
Linear regression

ABSTRACT

This paper presents a family of methods for locating/fitting hyperplanes with respect to a given set of points. We introduce a general framework for a family of aggregation criteria, based on ordered weighted operators, of different distance-based errors. The most popular methods found in the specialized literature, namely least sum of squares, least absolute deviation, least quantile of squares or least trimmed sum of squares among many others, can be cast within this family as particular choices of the errors and the aggregation criteria. Unified mathematical programming formulations for these methods are provided and some interesting cases are analyzed. The most general setting give rise to mixed integer nonlinear programming problems. For those situations we present inner and outer linear approximations to assess tractable solution procedures. It is also proposed a new goodness of fitting index which extends the classical coefficient of determination and allows one to compare different fitting hyperplanes. A series of illustrative examples and extensive computational experiments implemented in R are provided to show the applicability of the proposed methods.

1. Introduction

The problem of locating hyperplanes with respect to a given set of point is well-known in Location Analysis (LA) Schöbel (1999). This problem is closely related to another common question in Data Analysis (DA): to study the behavior of a given set of data with respect to a fitting body expressed with an equation of the form $f(x)=0$, with $x=(X_1,\ldots,X_d)\in\mathbb{R}^d$. This last problem reduces to the estimation of the 'best' function f that expresses the relationship between the data or, in the jargon of LA, to the location of the surface $f(x)=0$ that minimizes some aggregation function of the distances to these points (see Amaldi et al., 2016; Diaz-Báñez et al., 2004; Drezner et al., 2002). In many cases the family of functions where f belongs to is fixed and then, the parameters defining such an *optimal* function must be determined. The family of linear functions is the most widely used. This implies that the above equation is of the form $f(x)=\beta_0+\sum_{k=1}^d \beta_k X_k=0$ for $\beta_0,\beta_1,\ldots,\beta_d\in\mathbb{R}$.

To perform such a fitting, we are given a set of points $\{x_1,\ldots,x_n\}\subset\mathbb{R}^d$, and the goal is to find the vector $\hat{\boldsymbol{\beta}}=(\hat\beta_0,\hat\beta_1,\ldots,\hat\beta_d)$ that minimizes some measure of the deviation of the data with respect to the hyperplane it induces, $\mathcal{H}(\hat{\boldsymbol{\beta}})=\{z\in\mathbb{R}^d:\hat\beta_0+\sum_{k=1}^d\hat\beta_k z_k=0\}$. For a given point $x\in\mathbb{R}^d$, we define the *residual* with respect to a generic x as a mapping $\varepsilon_x:\mathbb{R}^{d+1}\to\mathbb{R}_+$, that maps any set of coefficients $\boldsymbol{\beta}=(\beta_0,\ldots,\beta_d)\in\mathbb{R}^{d+1}$, into a measure $\varepsilon_x(\boldsymbol{\beta})$ that represents the deviation of the given point x from the hyperplane with those parameters. The problem of locating a hyperplane for a given set of points $\{x_1,\ldots,x_n\}\subseteq\mathbb{R}^d$ consists of finding the coefficients minimizing an aggregation function, $\Phi:\mathbb{R}^n\to\mathbb{R}$, of the residuals of all the points. Different choices for the residuals and the aggregation criteria will give, in general, different optimal values for the parameters and thus different properties for the resulting hyperplanes. This problem is not new and some of these criteria, as the minisum, minimax and some other alternatives, have been widely analyzed from a LA perspective (see Carrizosa and Plastria, 1995; Megiddo and Tamir, 1983; Schöbel, 1996; Schöbel, 1997; Schöbel, 1998; Schöbel, 1999, among other).

A first approach to locate a hyperplane is to consider that residuals, with respect to given points, are individual measures of error and thus, each residual should be minimized independently of the remaining (Carrizosa et al., 1995; Narula and Wellington, 2007). It is clear that this simultaneous minimization will not be possible in most of the cases and then several strategies can be followed: one can try to find the set of Pareto fitting curves (Carrizosa et al.,

* Corresponding author.
E-mail addresses: vblanco@ugr.es (V. Blanco), puerto@us.es (J. Puerto), romansg@ugr.es (R. Salmerón).

https://doi.org/10.1016/j.cor.2018.03.009
0305-0548/

1995) or alternatively, to apply an aggregation function that incorporates the holistic preference of the Decision-Maker on the different residuals (Yager and Beliakov, 2010). This last choice is very difficult and it is usual to apply an approach of *complete uncertainty* (i.e., it is assumed that it is known the set of possible outcomes, but there is no information about the probabilities of those outcomes or about their likelihood ranking) leading to additive aggregations.

The most popular methods to compute the coefficients of an optimal hyperplane consider that the residuals are the differences from one of the coordinates of the space (which are usually known as vertical/horizontal distances). In this paper we present a framework that generalizes previous contributions for optimally locating/fitting hyperplanes to a set of points. It introduces a family of combinations residuals-criteria that allows for a great flexibility to accommodate hyperplanes to a set of points (Marín et al., 2009; Nickel and Puerto, 2005). One of the contributions of our proposal is the use of modern mathematical programming tools to solve the problems which are involved in the computation of the parameters of the fitting models. In addition, it can be combined with some of the mathematical programming techniques for feature selection (Bertsimas and Mazumder, 2014), with classification schemes (Bertsimas and Shioda, 2007), or with constraints on the coefficients of the linear manifold. This unified framework is also able to accommodate general forms of regularization, as upper bound on the ℓ_2-norm of the coefficients (Hoerl and Kennard, 1988), since it would only mean to add additional constraints to the mathematical programming formulations proposed in the paper, at the price of increasing the computational complexity needed for solving the problems. Many of the formulations described in this paper have been implemented in R in order to be available for data analysts.

In our framework, errors are measured as shortest distances, based on a norm, between the given points and the fitting surface. This makes the location problem geometrically invariant which is an interesting advance with respect to vertical/horizontal residuals. We observe that this framework subsumes as particular cases the standard location methods that consider residuals based on vertical distances (commonly used in Statistics); as well as most of the particular cases of fitting linear bodies using vertical distances but different aggregation criteria described in the literature, as ℓ_p fitting (ℓ_p-norm criterion), least quantile of squares (Bertsimas and Mazumder, 2014; Rousseeuw, 1984), least trimmed sum of squares (Atkinson and Cheng, 1999; Rousseeuw, 1983), etc. The use of non-standard residuals is common in the area of LA and other areas of Operations Research. However, it is not that usual in the field of regression analysis although orthogonal (ℓ_2) residuals have been already used, see, e.g., Euclidean Fitting (Bargiela and Hartley, 1993; Cavalier and Melloy, 1991; Pinson et al., 2008) or Total Least Squares (Van Huffel and Vanderwalle, 1991), mainly applied to bidimensional data; and the more general geodesic distance residuals are applied in geodesic regression (Fletcher, 2013). Quoting the reasons for that fact given by Giloni and Padberg (2002): "we have left out a summary of linear regression models using the more general ℓ_τ-norms with $\tau \notin \{1, 2, \infty\}$ for which the computational requirements are considerably more burdensome than in the linear programming case (as they generally require methods from convex programming where machine computations are far more limited today)."

In order to compare the *goodness of the fitting* for the different models, we have developed a new generalized measure of fit. This proposal is based on a generalization of the classical coefficient of determination for least squares fitting, that will allow one to measure how good is an optimal hyperplane with respect to the best constant model, $X_d = \beta_0$.

The paper is organized as follows. In Section 2 we introduce the new framework for fitting hyperplanes as well as some results that allow us to interpret the obtained solutions for practical purposes. Next, in Section 3, a residual-aggregation dependent goodness of fitting index is defined and an efficient approach for its computation is presented. Two types of residuals are analyzed in more detail, namely those induced by polyhedral-and-ℓ_τ norms for rational $\tau \geq 1$. In Section 4, we present new methods for the location of hyperplanes assuming that the residuals are measured as the shortest norm-based distance between the given points (data set) and the linear fitting body using polyhedral norms. The results of this section are instrumental. They constitute the basis to address the more general problems in Section 5, since they will permit to develop inner and outer linear approximations for more general Mixed Integer Non Linear Programming (MINLP) problems that result in the general case. Section 5 analyzes the location of hyperplanes using ℓ_τ norms. Since in this case non convex problems are derived, we also present outer and inner linear approximations that reduce, the corresponding MINLP problems with ℓ_τ-norms residuals, to problems with polyhedral norm residuals. Section 6 is devoted to the computational experiments. We report results for synthetic data and for the classical data set given in Durbin and Watson (1951). In addition, we include an illustrative example of the scalability of the methodology with several thousands of points. The paper finishes with some concluding remarks and future research.

2. A flexible methodology for the location of hyperplanes

Given is a set of n observations or demand points (depending that we use the *jargon* of data analysis or location analysis, respectively) in a $(d+1)$-dimensional space, $\{x_1, \ldots, x_n\} \subset \{1\} \times \mathbb{R}^d$ (we will assume, for a clearer description of the models, that the first, the 0th, component of x_i is the one that account for the intercept, being $x_{10} = \cdots = x_{n0} = 1$). Next, we analyze the problem of locating a linear form (hyperplane) to fit these points minimizing different forms of measuring the residuals and their aggregation. For any $y \in \mathbb{R}^{d+1}$, we shall denote $y_{-0} = (y_1, \ldots, y_d)$, i.e., the vector with the last d coordinates of y excluding the first one. First, we assume that the point-to-hyperplane deviation is modeled by a residual mapping $\varepsilon_x : \mathbb{R}^{d+1} \to \mathbb{R}_+,\ \varepsilon_x(\boldsymbol{\beta}) = \mathrm{D}(x_{-0}, \mathcal{H}(\boldsymbol{\beta}))$, being D a distance measure in $\mathbb{R}^d$. This residual represents how "far" is the point (observation) $x \in \mathbb{R}^{d+1}$ with respect to the hyperplane $\mathcal{H}(\boldsymbol{\beta}) = \{y \in \mathbb{R}^d : (1, y^t)\boldsymbol{\beta} = 0\}$. At times, for the sake of brevity, we will write the hyperplane as $\boldsymbol{\beta}^t X = 0$, with $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_d)^t \in \mathbb{R}^{d+1}$. In addition, to simplify the presentation, we will refer, whenever no possible confusion occurs, to the residual with respect to the point x_i as ε_i.

An overall measure of the deviations of the whole data set with respect to the hyperplane induced by $\boldsymbol{\beta}$ is obtained by using an aggregation function of the residuals, $\Phi : \mathbb{R}^n \to \mathbb{R}$. With this setting, one tries to minimize such an aggregation function and the *Fitting Hyperplane Problem* (FHP) consists of finding $\hat{\boldsymbol{\beta}} \in \mathbb{R}^{d+1}$ such that:

$$\hat{\boldsymbol{\beta}} \in \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{d+1}} \Phi(\boldsymbol{\varepsilon}(\boldsymbol{\beta})), \tag{1}$$

where $\boldsymbol{\varepsilon}(\boldsymbol{\beta}) = (\varepsilon_1(\boldsymbol{\beta}), \ldots, \varepsilon_n(\boldsymbol{\beta}))^t$ is the vector of residuals.

Note that the difficulty of solving Problem (1) depends on both the expressions for the residuals and the aggregation criterion Φ. If Φ and ε_x are linear, the above problem becomes a linear programming problem. In this paper, we consider a general family of aggregation criteria that includes as particular cases most of the classical ones used in the literature (Bertsimas and Mazumder, 2014; Giloni and Padberg, 2002; Rousseeuw and Leroy, 2003; Yager and Beliakov, 2010).

Let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and let $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ be the vector of residuals of all of the points in the given data set. We consider aggregation criteria