Multisource Linear Regression

Víctor Blanco
Universidad de Granada

Desafios de la Matemática Combinatoria
FQM-5849

(joint work with D. Ponce and J. Puerto)

VI International Workshop on Locational Analysis and Related Problems
Barcelona, November 2015

Multiple Linear Regression

Given a set of variables $X_1, \ldots, X_d$ *multiple regression* analyzes the existence of some relationship among them.

Multiple Linear Regression

Given a set of variables $X_1, \ldots, X_d$ *multiple regression* analyzes the existence of some relationship among them.

$$f(X_1, \ldots, X_d) = 0$$

And the function f is estimated based on a sample of data.

Multiple Linear Regression

Given a set of variables $X_1, \dots, X_d$ *multiple regression* analyzes the existence of some relationship among them.

$$f(X_1, \dots, X_d) = 0$$

And the function f is estimated based on a sample of data.
A Linear Regression model appears if we assume that f belongs to the set of linear functions, i.e.:

$$f(X_1, \dots, X_d) = \beta_0 + \sum_{k=1}^{d} \beta_k \, X_k$$

for some $\beta_0, \beta_1, \dots, \beta_d \in \mathbb{R}$.

Residuals

Given a sample of data $\{x_1, \ldots, x_n\} \subseteq \mathbb{R}^{d+1}$ [1] one tries to find the model that minimizes the deviation of the data with respect to the fitting body

$$\mathcal{H}(\hat{\beta}) = \{z \in \mathbb{R}^{d+1} : \sum_{k=0}^{d} \hat{\beta}_k z_k = 0\}.$$

[1] assume that $x_i = (1, x_{i1}, \ldots, x_{id})$

Residuals

Given a sample of data $\{x_1, \ldots, x_n\} \subseteq \mathbb{R}^{d+1}$ [1] one tries to find the model that minimizes the deviation of the data with respect to the fitting body

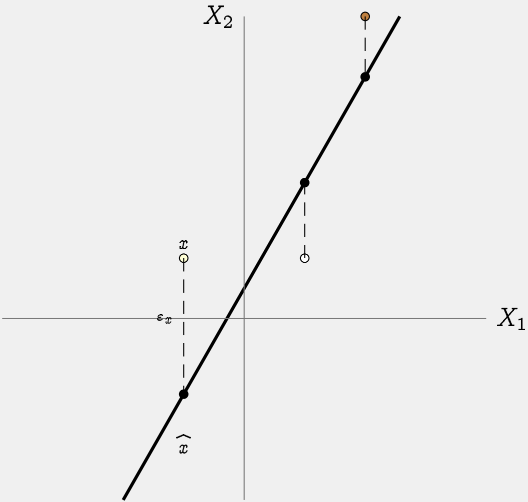
$$\mathcal{H}(\hat{\beta}) = \{z \in \mathbb{R}^{d+1} : \sum_{k=0}^{d} \hat{\beta}_k z_k = 0\}.$$

For an observation x the residual is the error when adjusting a model compared to the sample data.

- Usually: $\varepsilon_x = \left| x_d - \sum_{k=0}^{d-1} \beta_k x_k \right|, \ (\beta_d = 1).$ (Vertical Distance)

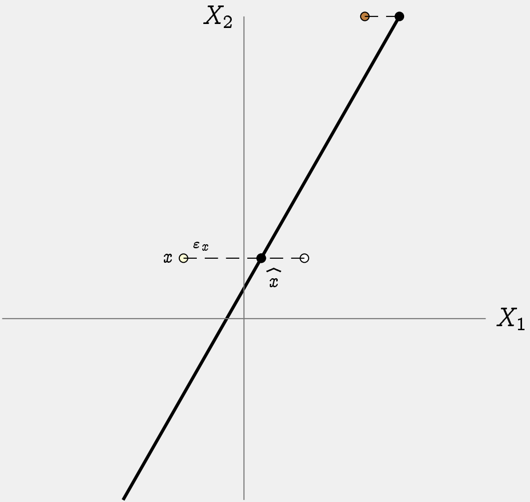
[1] assume that $x_i = (1, x_{i1}, \ldots, x_{id})$

Residuals



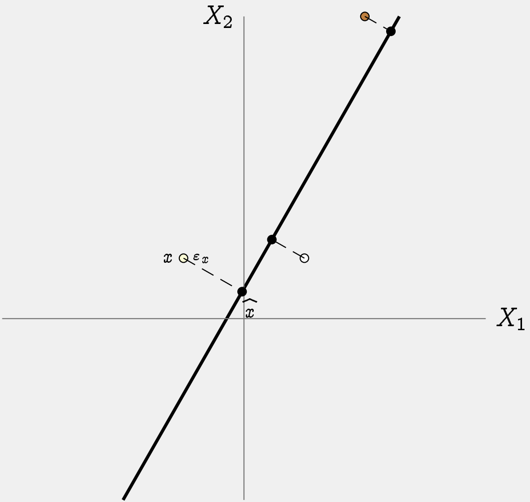
Residuals

Why not to consider different point-to-model error measures?



Residuals

Why not to consider different point-to-model error measures?



Residuals

Let $\varepsilon_x : \mathbb{R}^{d+1} \to \mathbb{R}_+$ be a mapping that represents how "far" is the point (observation) $x \in \mathbb{R}^{d+1}$ with respect to the hyperplane $\mathcal{H}(\boldsymbol{\beta}) = \{y \in \mathbb{R}^d : (1, y^t)\boldsymbol{\beta} = 0\}$, as

$$\varepsilon_x(\boldsymbol{\beta}) = \mathrm{D}(x_{-0}, \mathcal{H}(\boldsymbol{\beta})),$$

being D a distance measure in $\mathbb{R}^d$.

(for any $x \in \mathbb{R}^{d+1}$, $x_{-0} = (x_1, \ldots, x_d)$, the vector with the last d coordinates of x excluding the first one)

Aggregation Criteria

The final goal of a regression model is:

Given a set of points $\{x_1, \dots, x_n\} \subseteq \mathbb{R}^d$, find the coefficients minimizing the residuals.

$$\min(\varepsilon_1, \dots, \varepsilon_n)$$

A multiobjective optimization problem (Carrizosa, Conde, Fernández, Muñoz, Puerto; 1995).
It is usual to transform such a multiobjective problem into a scalar problem by *aggregating* residuals.

Aggregation Criteria

The final goal of a regression model is:

Given a set of points $\{x_1, \ldots, x_n\} \subseteq \mathbb{R}^d$, find the coefficients minimizing the residuals.

$$\min(\varepsilon_1, \ldots, \varepsilon_n)$$

A multiobjective optimization problem (Carrizosa, Conde, Fernández, Muñoz, Puerto; 1995).
It is usual to transform such a multiobjective problem into a scalar problem by *aggregating* residuals.

- Sum of Residuals.
- Sum of Squares of residuals.
- Maximum of residuals.

Aggregating Residuals

Let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and let $\varepsilon \in \mathbb{R}^n$ be the residuals.
We consider the aggregation criteria $\Phi : \mathbb{R}^n \to \mathbb{R}_+$ defined as:

$$\Phi(\varepsilon) = \sum_{i=1}^{n} \lambda_i \, \varepsilon_{(i)}^p$$

where $\varepsilon_{(i)} \in \{\varepsilon_1, \ldots, \varepsilon_n\}$ is such that $\varepsilon_{(1)} \leq \cdots \leq \varepsilon_{(n)}$.

Aggregating Residuals

Let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ and let $\varepsilon \in \mathbb{R}^n$ be the residuals.
We consider the aggregation criteria $\Phi : \mathbb{R}^n \to \mathbb{R}_+$ defined as:

$$\Phi(\varepsilon) = \sum_{i=1}^{n} \lambda_i \, \varepsilon_{(i)}^{p}$$

where $\varepsilon_{(i)} \in \{\varepsilon_1, \ldots, \varepsilon_n\}$ is such that $\varepsilon_{(1)} \leq \cdots \leq \varepsilon_{(n)}$.

- SUM ($\lambda_i = 1,\ p = 1$)
- SOS ($\lambda_i = 1,\ p = 2$)
- MAX ($\lambda_n = 1,\ \lambda_i = 0,\ i \neq n$)
- MEDIAN ($\lambda_{\lceil \frac{n}{2} \rceil} = 1,\ \lambda_i = 0,\ i \neq \lceil \frac{n}{2} \rceil$)
- TRIMMED MEAN ($\lambda_i = 0, i = 1, \ldots, \lceil \frac{n}{2} \rceil,\ \lambda_i = 1,\ i > \lceil \frac{n}{2} \rceil$)
- RANGE ($\lambda_n = -\lambda_1 = 1,\ \lambda_i = 0,\ i \neq 1, n$)
- ...

Are the extesions reasonable?

“Least squares regression estimators, has been studied intensively for well over 200 years now, primarily due to its convenient closed form.” (Giloni & Padberg, 2002).

Are the extesions reasonable?

> *"Least squares regression estimators, has been studied intensively for well over 200 years now, primarily due to its convenient closed form." (Giloni & Padberg, 2002).*

Under Gaussian distribution of the error terms an impressive statistical apparatus has been created to assess the goodness of fit, the quality of individual and/or subsets of the regression coefficients, as well as other statistical properties of the linear regression model. But:

> *"The ancient solitary reign of the exponential (Gaussian) law of error should come to an end". (Edgeworth, 1920).*

Are the extesions reasonable?

"Least squares regression estimators, has been studied intensively for well over 200 years now, primarily due to its convenient closed form." (Giloni & Padberg, 2002).

Under Gaussian distribution of the error terms an impressive statistical apparatus has been created to assess the goodness of fit, the quality of individual and/or subsets of the regression coefficients, as well as other statistical properties of the linear regression model. But:

"The ancient solitary reign of the exponential (Gaussian) law of error should come to an end". (Edgeworth, 1920).

"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)." (Giloni & Padberg, 2002).

Generalized Linear Regression

Given:

- A sample of data $\{x_1, \ldots, x_n\} \in \mathbb{R}^{d+1}$,
- Residuals $\varepsilon_x : \mathbb{R}^{d+1} \to \mathbb{R}$, and
- Aggregation of residuals criterion $\Phi : \mathbb{R}^n \to \mathbb{R}$.

Generalized Linear Regression

Given:

- A sample of data $\{x_1, \ldots, x_n\} \in \mathbb{R}^{d+1}$,
- Residuals $\varepsilon_x : \mathbb{R}^{d+1} \to \mathbb{R}$, and
- Aggregation of residuals criterion $\Phi : \mathbb{R}^n \to \mathbb{R}$.

Find

$$\hat{\boldsymbol{\beta}} \in \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{d+1}} \Phi(\varepsilon_x(\boldsymbol{\beta})), \qquad (\mathrm{LRP}_{\Phi,\varepsilon})$$

where $\varepsilon_x(\boldsymbol{\beta}) = (\varepsilon_{x_1}(\boldsymbol{\beta}), \ldots, \varepsilon_{x_n}(\boldsymbol{\beta}))^t$ is the vector of residuals.

Regression and Location

Given a set of demand points $A = \{a_1, \ldots, a_n\} \subset \mathbb{R}^{d+1}$ (assuming that $a_{i1} = 1$ for $i = 1, \ldots, n$) endowed with a distance measure between points in $\mathbb{R}^{d+1}$, γ, the goal of continuous location models is to find $\boldsymbol{\beta}^*$

$$\boldsymbol{\beta}^* \in \arg\min_{\beta \in \mathbb{R}^d} \Psi(\gamma(a_1, \boldsymbol{\beta}), \ldots, \gamma(a_n, \boldsymbol{\beta})).$$

For an error measure ε (defined as a norm-based distance $\|\cdot\|$) and an aggregation criterion Φ, solving the linear regression problem to fit the model $\boldsymbol{\beta}^t X = 0$ is nothing but a continuous location problem where the residuals ε_{a_i} are:

$$\varepsilon_{a_i} := \gamma(a_i, \boldsymbol{\beta}) = \mathrm{D}(a_i, \mathcal{H}(\boldsymbol{\beta})) = \frac{|\boldsymbol{\beta}^t a_i|}{\|(\beta_1, \ldots, \beta_d)\|^*}$$

Regression and Location

Given a set of demand points $A = \{a_1, \ldots, a_n\} \subset \mathbb{R}^{d+1}$ (assuming that $a_{i1} = 1$ for $i = 1, \ldots, n$) endowed with a distance measure between points in $\mathbb{R}^{d+1}$, γ, the goal of continuous location models is to find $\boldsymbol{\beta}^*$

$$\boldsymbol{\beta}^* \in \arg\min_{\beta \in \mathbb{R}^d} \Psi(\gamma(a_1, \boldsymbol{\beta}), \ldots, \gamma(a_n, \boldsymbol{\beta})).$$

For an error measure ε (defined as a norm-based distance $\|\cdot\|$) and an aggregation criterion Φ, solving the linear regression problem to fit the model $\boldsymbol{\beta}^t X = 0$ is nothing but a continuous location problem where the residuals ε_{a_i} are:

$$\varepsilon_{a_i} := \gamma(a_i, \boldsymbol{\beta}) = \mathrm{D}(a_i, \mathcal{H}(\boldsymbol{\beta})) = \frac{|\boldsymbol{\beta}^t a_i|}{\|(\beta_1, \ldots, \beta_d)\|^*}$$

This implies that many results already known in the field of Location Analysis can be applicable in solving generalized regression problems.

Responses and Prediction

For norm-based distances (based in Mangasarian, 1999):

For a given observation $z^t = (1, z_1, \ldots, z_d)$ and the linear fitting body $\mathcal{H}(\boldsymbol{\beta})$ the response $\hat{z}$ consistent with the residual $\varepsilon_z = \min_{y \in \mathcal{H}(\boldsymbol{\beta})} ||z_{-0} - y||$ is given by

$$\hat{z} = z_{-0} - \frac{\boldsymbol{\beta}^t z}{||\boldsymbol{\beta}_{-0}||^*} \mathrm{k}(\boldsymbol{\beta}),$$

where $|| \cdot ||^*$ is the dual norm to $|| \cdot ||$ and $\mathrm{k}(\boldsymbol{\beta}) = \arg \max_{||x||=1} \boldsymbol{\beta}^t_{-0} x$.

Moreover,

$$\varepsilon_z = \frac{|\boldsymbol{\beta}^t z|}{||\boldsymbol{\beta}_{-0}||^*}. \tag{1}$$

Marginal variation: $\frac{\partial \hat{z}_d}{\partial z_j} = -\frac{\beta_j}{||\boldsymbol{\beta}_{-0}||^*} \mathrm{k}(\boldsymbol{\beta})_d.$

Responses and Prediction

Let $z = (1, z_1, \ldots, z_d)^t$, then

1. If D is the ℓ_1- distance,
$$\hat{z}_k = \begin{cases} z_k & \text{if } |\beta_k| \neq \max\{|\beta_j| : j = 1, \ldots, d\}, \\ z_k - \frac{\beta^t z}{\|\beta_{-0}\|_\infty} v_k, & \text{if } \beta_k = \max\{|\beta_j| : j = 1, \ldots, d\}, \\ z_k + \frac{\beta^t z}{\|\beta_{-0}\|_\infty} v_k, & \text{if } \beta_k = -\max\{|\beta_j| : j = 1, \ldots, d\}, \end{cases} \quad \text{for}$$
$k = 1, \ldots, d$, and for some $v_1, \ldots, v_d \geq 0$ such that $\sum_j v_j = 1$.

2. If D is the ℓ_∞- distance,
$$\hat{z}_k = \begin{cases} z_k - \frac{\beta^t z}{\|\beta_{-0}\|_1}, & \text{if } \beta_k > 0, \\ z_k + \frac{\beta^t z}{\|\beta_{-0}\|_1}, & \text{if } \beta_k < 0, \end{cases} \qquad k = 1, \ldots, d.$$

3. If D is the ℓ_τ- distance with $1 < \tau < +\infty$ then
$\hat{z}_k = z_k - \frac{\beta^t z}{\|\beta_{-0}\|_\nu} \mathrm{k}_\tau(\beta)_k, \quad k = 1, \ldots, d$ and
$$\mathrm{k}_\tau(\beta)_k = \begin{cases} \frac{\mathrm{sg}(\beta_k)|\beta_k|^{\nu/\tau}}{(\sum_{j=1}^d |\beta_j|^\nu)^{1/\tau}} & \text{if } \beta_k \neq 0 \\ 0 & \text{if } \beta_k = 0, \end{cases} \qquad k = 1, \ldots, d, \ \frac{1}{\tau} + \frac{1}{\nu} = 1.$$

A General Model: non-negative lambdas

$$
\begin{aligned}
\Phi^* = \min \sum_{j=1}^{n} \lambda_j \theta_j & & (\mathrm{LR}_{\Phi,\|\cdot\|})\\
\text{s.t. } \varepsilon_i \geq \frac{|\boldsymbol{\beta}^t x_i|}{\|\boldsymbol{\beta}_{-0}\|^*}, & \quad \forall i = 1, \ldots, n, & (2)\\
z_i^s \geq \varepsilon_i^r, & \quad \forall i = 1, \ldots, n, & (3)\\
z_i \leq \theta_j + M(1 - w_{ij}), & \quad \forall i, j = 1, \ldots, n, & (4)\\
\theta_j \geq \theta_{j-1}, & \quad \forall j = 2, \ldots, n, & (5)\\
\sum_{i=1}^{n} w_{ij} = 1, & \quad \forall j = 1, \ldots, n, & (6)\\
\sum_{j=1}^{n} w_{ij} = 1, & \quad \forall i = 1, \ldots, n, & (7)\\
\boldsymbol{\beta} \in \mathbb{R}^{d+1}, w \in \{0,1\}^{n \times n}, z, \theta \in \mathbb{R}^n_+. & &
\end{aligned}
$$

OWA: (Nickel and Puerto, 2003), (Fernández, Pozo and Puerto, 2015)

A General Model

Each constraint $z^s \geq \varepsilon^r$ can be equivalently written as a set of $O(\lfloor \log_2(r) \rfloor)$ second order cone constraints with $\lfloor \log_2(r) \rfloor$ additional nonnegative variables. (B., Puerto, ElHaj; 2014)

A General Model

Each constraint $z^s \geq \varepsilon^r$ can be equivalently written as a set of $O(\lfloor \log_2(r) \rfloor)$ second order cone constraints with $\lfloor \log_2(r) \rfloor$ additional nonnegative variables. (B., Puerto, ElHaj; 2014)

For $0 \leq \lambda_1 \leq \cdots \leq \lambda_n$:

$$\begin{aligned}
\Phi^* = \min \quad & \sum_{j=1}^{n} v_j + \sum_{i=1}^{n} w_i \\
\text{s.t.} \quad & \varepsilon_i \geq \frac{|\boldsymbol{\beta}^t x_i|}{\|\boldsymbol{\beta}_{-0}\|^*}, \forall i = 1, \ldots, n, \\
& z_i^s \geq \varepsilon_i^r, \forall i = 1, \ldots, n, \\
& v_j + w_i \geq \lambda_i z_j, \forall i, j = 1, \ldots, n, \\
& \boldsymbol{\beta} \in \mathbb{R}^{d+1}, z \in \mathbb{R}^n_+, v, w \in \mathbb{R}^n
\end{aligned}$$

(B., Puerto, Salmerón, Arxiv2015): SOCP for block-norm residuals and inner-outer approx. for ℓ_τ. Lots of Experiments...

GCoD

$$\mathrm{GCoD}_{\Phi,\varepsilon} = 1 - \frac{\Phi^*}{\Phi_0^*}$$

Φ_0^*: optimal value when the model is required to be constant: $X_d = \beta_0$.

$\mathrm{GCoD}_{\Phi,\varepsilon}$ measures the improvement of the model that considers all the independent variables with respect to the one that omits all of them.

GCoD

$$\mathrm{GCoD}_{\Phi,\varepsilon} = 1 - \frac{\Phi^*}{\Phi_0^*}$$

Φ_0^*: optimal value when the model is required to be constant: $X_d = \beta_0$. $\mathrm{GCoD}_{\Phi,\varepsilon}$ measures the improvement of the model that considers all the independent variables with respect to the one that omits all of them.

$$\Phi_0^* = \frac{1}{\max\limits_{z \in \mathbb{R}^d : ||z|| \leq 1} z_d} \min_{\beta_0 \in \mathbb{R}} \Phi(|x_{1d} - \beta_0|, \ldots, |x_{nd} - \beta_0|),$$

GCoD

$$\mathrm{GCoD}_{\Phi,\varepsilon} = 1 - \frac{\Phi^*}{\Phi_0^*}$$

Φ_0^*: optimal value when the model is required to be constant: $X_d = \beta_0$.
$\mathrm{GCoD}_{\Phi,\varepsilon}$ measures the improvement of the model that considers all the independent variables with respect to the one that omits all of them.

$$\Phi_0^* = \frac{1}{\max\limits_{z \in \mathbb{R}^d : ||z|| \leq 1} z_d} \min_{\beta_0 \in \mathbb{R}} \Phi(|x_{1d} - \beta_0|, \ldots, |x_{nd} - \beta_0|),$$

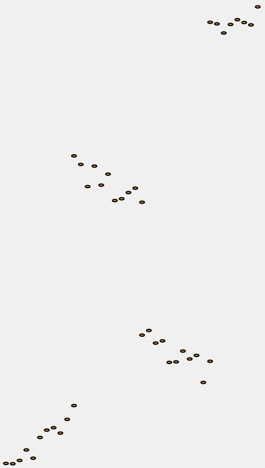
- Φ_0^* can be computed in $O(n^2)$ by a simple exploration.

Multisource regression

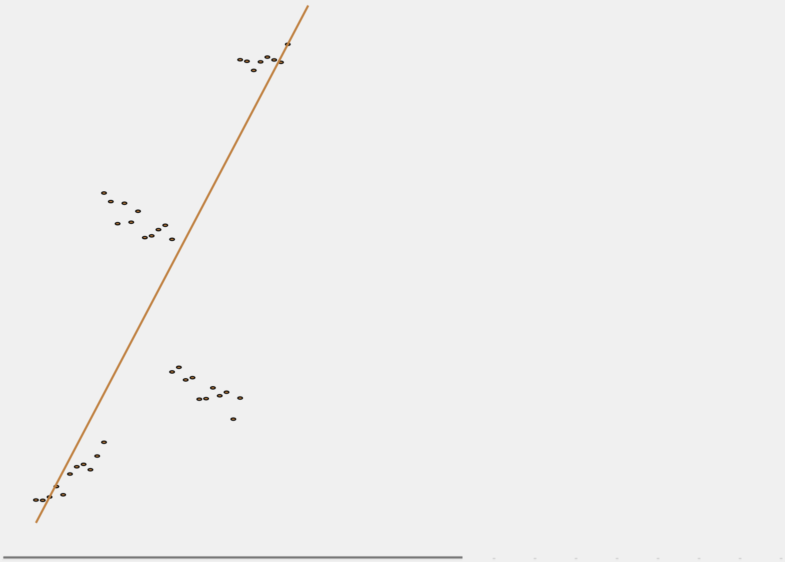
It may occur that a single linear model is not adequate for the data because there are subgroups of the sample with significantly different behavior with respect to the others.

One of the solutions to this problem is to consider simultaneously the two-side problem of classifying and fit the data to several linear models with an unified framework. This approach is called Clusterwise regression Jiang et al. (2013) or Segmented Regression Chen et al. (2012).

Multisource Regression

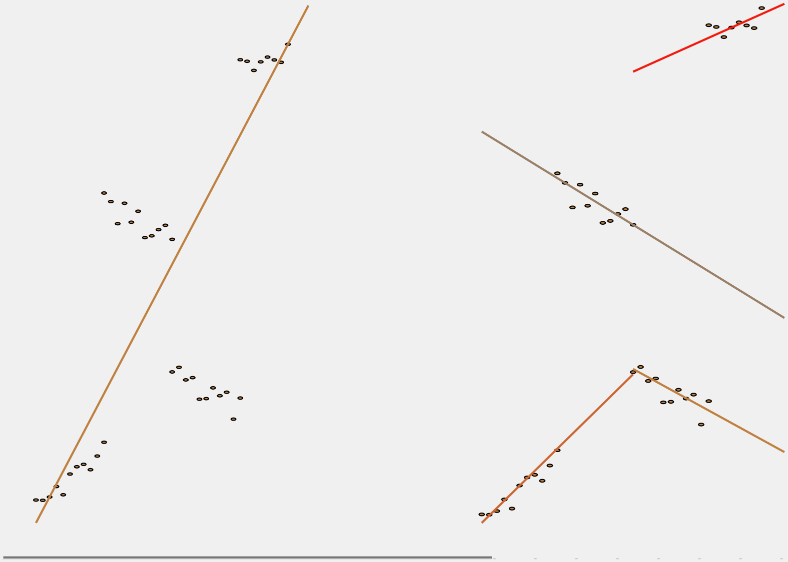


Multisource Regression

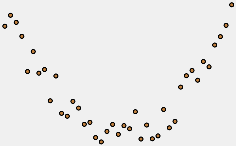


Multisource Regression

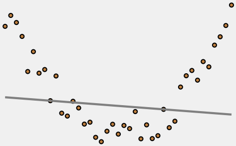
Multisource Regression



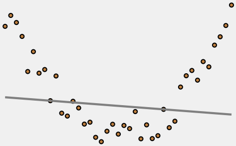
Multisource regression

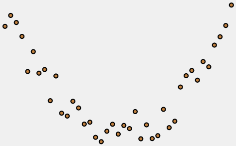


Multisource regression

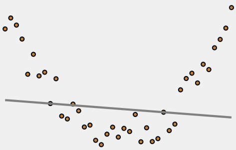


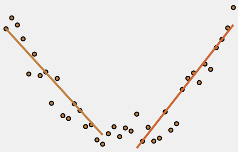
Multisource regression





Multisource regression

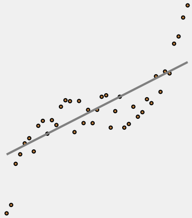




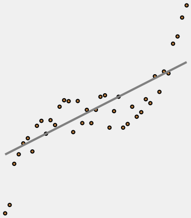
Multisource Regression

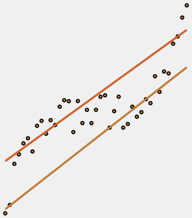
Multisource Regression

Multisource Regression

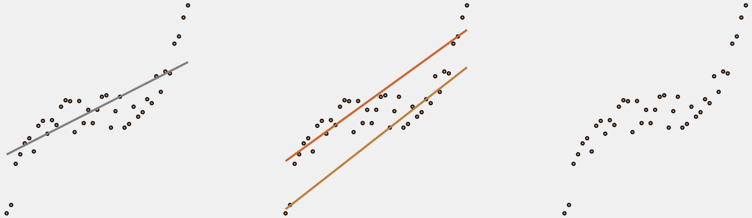


Multisource Regression

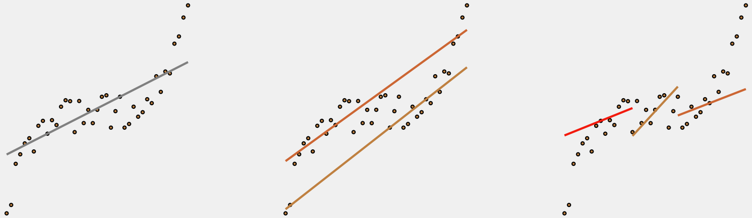




Multisource Regression



Multisource Regression



Multisource Regression

A sample of n observations about d quantitative measures, $\{x_1, \ldots, x_n\} \subset \mathbb{R}^{d+1}$

Our goal is to compute p linear models to fit the data as well as the allocations of each point to the best model (in terms of the residuals). We compute a set of p hyperplanes of the following general shape:

$$\mathcal{H}(\boldsymbol{\beta}_j) = \{y \in \mathbb{R}^d : (1, y^t)\boldsymbol{\beta}_j = 0\}, \quad j = 1, \ldots, p.$$

Multisource Regression

A sample of n observations about d quantitative measures, $\{x_1, \ldots, x_n\} \subset \mathbb{R}^{d+1}$

Our goal is to compute p linear models to fit the data as well as the allocations of each point to the best model (in terms of the residuals). We compute a set of p hyperplanes of the following general shape:

$$\mathcal{H}(\boldsymbol{\beta}_j) = \{y \in \mathbb{R}^d : (1, y^t)\boldsymbol{\beta}_j = 0\}, \quad j = 1, \ldots, p.$$

Residuals: $\varepsilon_i = \min_{j \in \{1,\ldots p\}} \varepsilon_{ij}$, with ε_{ij} the residual of *allocating* observation x_i to model $\mathcal{H}(\boldsymbol{\beta}_j)$, i.e., $\varepsilon_{ij} = d(x_i, \mathcal{H}(\boldsymbol{\beta}_j))$.

Aggregation Criterion: $\Phi(\varepsilon_1, \ldots, \varepsilon_n)$.

Multisource Regression

A sample of n observations about d quantitative measures, $\{x_1, \dots, x_n\} \subset \mathbb{R}^{d+1}$

Our goal is to compute p linear models to fit the data as well as the allocations of each point to the best model (in terms of the residuals). We compute a set of p hyperplanes of the following general shape:

$$\mathcal{H}(\boldsymbol{\beta}_j) = \{y \in \mathbb{R}^d : (1, y^t)\boldsymbol{\beta}_j = 0\}, \quad j = 1, \dots, p.$$

Residuals: $\varepsilon_i = \min_{j \in \{1,\dots p\}} \varepsilon_{ij}$, with ε_{ij} the residual of *allocating* observation x_i to model $\mathcal{H}(\boldsymbol{\beta}_j)$, i.e., $\varepsilon_{ij} = d(x_i, \mathcal{H}(\boldsymbol{\beta}_j))$.
Aggregation Criterion: $\Phi(\varepsilon_1, \dots, \varepsilon_n)$.

Cluster & Regression (CRIO): First group, then estimate (Bertsimas and Shioda, 2007).

Multisource regression

$$
\begin{aligned}
&\min \quad \Phi(e_1, \dots, e_n) \\
s.t. & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (8)\\
& e_i \geq \varepsilon_{ij} \, z_{ij} \, , \\
& \texttt{representation\_of\_residuals}(\varepsilon), \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (9)\\
& \sum_{j=1}^{p} z_{ij} = 1, \forall i = 1, \dots, n, \\
& z_{ij} \in \{0, 1\}, \forall i = 1, \dots, n, j = 1, \dots, p, \\
& e_i \in \mathbb{R}_+, \forall i = 1, \dots, n, \\
& \boldsymbol{\beta}_{jk} \in \mathbb{R}, \forall j = 1, \dots, p, k = 0, \dots, d-1.
\end{aligned}
$$

where

$$
z_{ij} = \begin{cases} 1 & \text{if the } i\text{th observation is assigned to } \mathcal{H}(\boldsymbol{\beta}_j), \\ 0 & \text{otherwise,} \end{cases}
$$

Multisource regression

$$
\begin{aligned}
& \min \quad \Phi(e_1, \ldots, e_n) \\
s.t. & \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (8) \\
& e_i \geq \varepsilon_{ij} - M(1 - z_{ij}), \forall i, j, \\
& \texttt{representation\_of\_residuals}(\varepsilon), \qquad\qquad\qquad (9) \\
& \sum_{j=1}^{p} z_{ij} = 1, \forall i = 1, \ldots, n, \\
& z_{ij} \in \{0, 1\}, \forall i = 1, \ldots, n, j = 1, \ldots, p, \\
& e_i \in \mathbb{R}_+, \forall i = 1, \ldots, n, \\
& \boldsymbol{\beta}_{jk} \in \mathbb{R}, \forall j = 1, \ldots, p, k = 0, \ldots, d - 1.
\end{aligned}
$$

where

$$
z_{ij} = \begin{cases} 1 & \text{if the } i\text{th observation is assigned to } \mathcal{H}(\boldsymbol{\beta}_j), \\ 0 & \text{otherwise,} \end{cases}
$$

Set partitioning formulation

- Let $I = \{1, \dots, n\}$ denote the entire set of observations.
- Let S be a cluster of observations $S \subseteq I$.
- Let c_S denote the cost of cluster S, i.e. the overall aggregation of the residuals of data in S.

$$y_S = \begin{cases} 1 & \text{if cluster } S \text{ is selected} \\ 0 & \text{otherwise.} \end{cases}$$

Set partitioning formulation

- Let $I = \{1, \ldots, n\}$ denote the entire set of observations.
- Let S be a cluster of observations $S \subseteq I$.
- Let c_S denote the cost of cluster S, i.e. the overall aggregation of the residuals of data in S.

$$y_S = \begin{cases} 1 & \text{if cluster } S \text{ is selected} \\ 0 & \text{otherwise.} \end{cases}$$

The set partition formulation is:

$$\min \sum_S c_S y_S \tag{10}$$

$$\sum_S y_S = p$$

$$\sum_{S \ni i} y_S = 1 \qquad \forall\, i = 1, \ldots, n$$

$$y_S \in \{0, 1\},\ S \subset \{1, \ldots, n\}. \tag{11}$$

Pricing problem

Let u be the dual variable for constraint $(\sum_S y_S = p)$ and v_i the dual variables for constraints $(\sum_{S \ni i} y_S = 1)$. The reduced cost for variable y_S is $\bar{c}_S = c_S - u - \sum_{i \in S} v_i$.
For instance, the pricing problem for the vertical distance residual:

$$\min_S \sum_{i \in S} e_i^2 - u - \sum_{i \in S} v_i$$

Pricing problem

Let u be the dual variable for constraint ($\sum_S y_S = p$) and v_i the dual variables for constraints ($\sum_{S \ni i} y_S = 1$). The reduced cost for variable y_S is $\bar{c}_S = c_S - u - \sum_{i \in S} v_i$.
For instance, the pricing problem for the vertical distance residual:

$$\min_S \sum_{i \in S} e_i^2 - u - \sum_{i \in S} v_i$$

Clearly, this pricing problem can be formulated as a Mixed Integer Non Linear Programmming Problem similar to the *single*-source regression models.

Pricing as a mixed integer quadratic

$$
\begin{aligned}
\min \sum_{i=1}^{n} t_i - \sum_{i=1}^{n} v_i h_i & \\
\text{s.t. } e_i \geq |y - \boldsymbol{\beta}^t x_i| - M(1 - h_i), \forall i & \qquad (12)\\
t_i \geq e_i^2, \forall i = 1, \ldots, n, & \qquad (13)\\
h_i \in \{0, 1\}, \forall i = 1, \ldots, n, & \\
e_i \in \mathbb{R}_+, \forall i = 1, \ldots, n, & \\
\boldsymbol{\beta}_k \in \mathbb{R}, k = 0, \ldots, d - 1. &
\end{aligned}
$$

where $h_i = 1$ iff $i \in S$.

Pricing as a mixed integer quadratic

$$\min \sum_{i=1}^{n} t_i - \sum_{i=1}^{n} v_i h_i$$

$$\text{s.t.}\, e_i \geq |y - \boldsymbol{\beta}^t x_i| - M(1-h_i), \forall i \tag{12}$$

$$t_i \geq e_i^2, \forall i = 1, \ldots, n, \tag{13}$$

$$h_i \in \{0,1\}, \forall i = 1, \ldots, n,$$

$$e_i \in \mathbb{R}_+, \forall i = 1, \ldots, n,$$

$$\boldsymbol{\beta}_k \in \mathbb{R}, k = 0, \ldots, d-1.$$

where $h_i = 1$ iff $i \in S$.

COLUMN GENERATION...

To be continued...

- Behavior of CG...
- Use of norm-based residuals.
- Notion of MultiSource GCoD... Computation?
- Adapt to study *Structural Changes* in Time Series.
- ...

Thank you!

vblanco@ugr.es