COMPUTATIONAL GEOMETRY

Project 2015-2016

Louvet Florian

May 10, 2016

Université Libre de Bruxelles

DIALITY TRANSFORMATION

WHAT IS DUALITY TRANSFORMATION

As seen in class, the duality transformation is a mapping beteen points and lines.

WHAT IS DUALITY TRANSFORMATION

As seen in class, the duality transformation is a mapping beteen points and lines. We define:

- a point p by $p \equiv (x_p, y_p)$ its two coordinates in the cartesian plane
- a line l by the following equation $l \equiv y = mx + b$

As seen in class, the duality transformation is a mapping beteen points and lines. We define:

- a point p by $p \equiv (x_p, y_p)$ its two coordinates in the cartesian plane
- a line l by the following equation $l \equiv y = mx + b$

The dual transformation D of:

- the point p is noted D(p). It is a line defined by $D(p) \equiv y = (-x_p)x + y_p$
- the line l is noted D(l). It is a point defined by $D(l) \equiv (m, b)$

DUALITY TRANSFORMATION PROPERTIES

A duality transformation T can respect a number of properties, for example:

- Inverse : $T(T(l)) = l$ and $T(T(p)) = p$

A duality transformation T can respect a number of properties, for example:

- Inverse : $T(T(l)) = l$ and $T(T(p)) = p$
- Incidence : $I(p, l) \Leftrightarrow I(D(p), D(l))$, with
  $$I(p, l) = \begin{cases} \text{true} & mx_p + b - y_p = 0 \\ \text{false} & \text{otherwise} \end{cases}$$

A duality transformation T can respect a number of properties, for example:

- Inverse : $T(T(l)) = l$ and $T(T(p)) = p$
- Incidence : $I(p, l) \Leftrightarrow I(D(p), D(l))$, with
  $$I(p, l) = \begin{cases} \text{true} & mx_p + b - y_p = 0 \\ \text{false} & \text{otherwise} \end{cases}$$
- Ordering : If p is above l then T(p) is above T(l).

DUALITY TRANSFORMATION PROPERTIES

A duality transformation T can respect a number of properties, for example:

- Inverse (X) : $T(T(l)) = l$ and $T(T(p)) = p$
- Incidence (V) : $I(p, l) \Leftrightarrow I(D(p), D(l))$, with
  $$I(p, l) = \begin{cases} \text{true} & mx_p + b - y_p = 0 \\ \text{false} & \text{otherwise} \end{cases}$$
- Ordering (V) : If p is above l then T(p) is above T(l).

DUALITY TRANSFORMATION – DEMO

Time for a first demo.

REGRESSION DEPTH

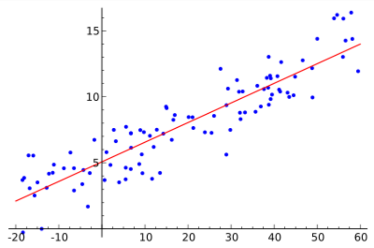
Simple linear regression is a commonly used linear estimation of the relationship between two statistical variables.

Given a dataset $Z_n = \{(x_i, y_i), i = 1, ..., n\} \subset \mathbb{R}^2$, the simple regression line f is given by

$$f \equiv y = \alpha x + \beta$$

where α and β are chosen such that they minimize $LS(\alpha, \beta)$

$$LS(\alpha, \beta) = \sum_{i=1}^{n} (y_i - \beta - \alpha x_i)^2$$



Rousseeuw Peter J. and Mia Hubert described in 1999 in the article ”Regression Depth” a new statistical concept called **regression depth**.

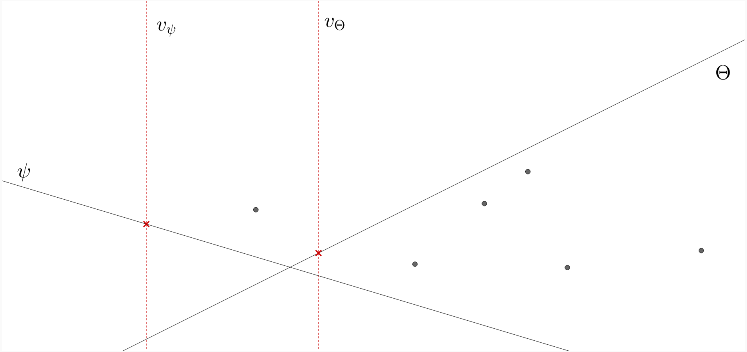
Goal: provide another estimator of relationship for bi-variate statistics.

Some definitions:

- observations : $Z_n = \{(x_i, y_i), i = 1, ..., n\} \subset \mathbb{R}^2$
- candidate fit l : $l \equiv y = mx + b$ or $l(m, b)$
- residuals: $r_i(l) = y_i - mx_i - b$ for every point $(x_i, y_i) \in Z_n$

A candidate fit $l = (m, b)$ to Z_n is called a **non-fit** iff there exists a real number v that does not coincide with any x_i and such that $\phi_i(l)$ is satisfied.

$$\phi_i(l) = ((\forall x_i < v, r_i(l) < 0) \wedge (\forall x_i > v, r_i(l) > 0)) \\ \vee((\forall x_i < v, r_i(l) > 0) \wedge (\forall x_i > v, r_i(l) < 0))$$



Regression depth: Is the the minimal number of observations that need to be removed from Z_n in order for a fit l to become a non-fit with respect to Z_n.

GEOMETRY

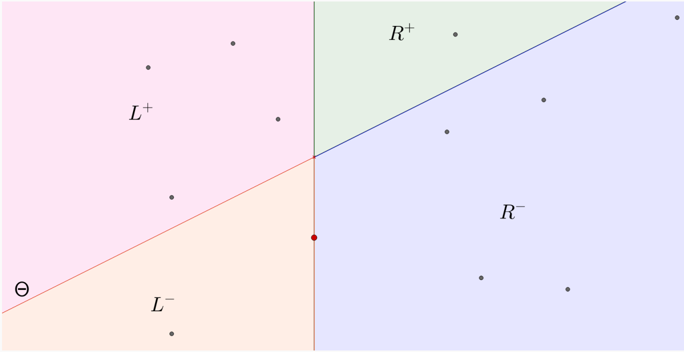
The problem defined by the dual transformation of the regression depth plot is a geometry problem.

The regression depth $\mathrm{rdepth}(\eta, L_n)$ of any candidate fit η relative to a set of n lines L_n is the minimal number of lines from L_n that should be removed so that η becomes a non-fit.

In the dual space, a candidate fit is a non-fit if there exists a direction u (such that $\| u \|= 1$) and such that the halfline $[\eta, \eta + u >$ does not intersect any line from the set L_n.

A FIRST ALGORITHM

FIRST ALGORITHM



- Sort all element of Z_n by their x coordinate such that $x_1 < x_2 < \ldots < x_n$
- Compute $\mathrm{rdepth}(l, Z_n) = \min_{1 \leq i \leq n}(\min\{L^+(x_i) + R^-(x_i), L^-(x_i) + R^+(x_i)\})$

REGRESSION DEPTH - DEMO

Time for a second demo.

BOUNDS

- **Lower bound**:

$$\max_{l} \text{rdepth}(l, Z_n) \geq \left\lceil \frac{n}{3} \right\rceil$$

- **Upper bound**: if all observations of Z_n are in general position

$$\max_{l} \text{rdepth}(l, Z_n) \leq \left\lfloor \frac{n+2}{2} \right\rfloor$$

DEEPSET REGRESSION LINE

AN SIMPLE ALGORITHM

As an estimator we take the fit that has the deepest (greatest) regression depth regarding a dataset Z_n. A simple algorithm is the following:

- Sort all element of Z_n by their x coordinate such that $x_1 < x_2 < ... < x_n$
- For each pair ij such that $0 < i < j \leq n$:
  - Find the line l_{ij} that intersects with observations i and j
  - Use the algorithm defined previously $r_{ij} = \text{rdepth}(l_{ij}, Z_n)$ (without the sorting wich is already done)
  - If r_{ij} is greater than the regression depth of the temporary solution, keep l_{ij} as a new temporary solution.
- Return the l_{ij} that has the greatest regression depth, it is the deepest regression line.

DEEPESET REGRESSION DEPTH - DEMO

Time for a last demo.

CONCLUSION

Better algorithms have been described since the publication of Rousseeuw and Hubert in 1999. In our ptoject we also described but did not implement an algorithm presented by Langerman Stefan and Steiger William in ”The Complexity of Hyperplane Depth in the Plane”. They describe an algorithm performing in $\mathcal{O}(n \log n)$.

Regression depth is defined for higher dimensions as Hyperplane depth. Algorithms presented in this presentation can be easily adapted to compute the of hyperplanes.

In this project, we presented a problem that originated from statistical resarch but created a new computational geometry problem by taking it's dual. The regression depth problem illustrates well that the interest of multiple research area led to the discovery of better algorithms for the problem and it's dual.

Thank you for your attention