Robuste Regression mit der Methode der kleinsten absoluten Abweichungen

Robert Schieweck 27. April 2010

Abstract

In der Statistik ist die Regression ein wichtiges Hilfsmittel zur Bestimmung von linearen Zusammenhängen zwischen verschiedenen Größen. Bei der einfachen linearen Regression in zwei Dimensionen folgt das statistische Modell der Gleichung

$$Y_i = \beta_0 + \beta_1 x_i + \epsilon_i \qquad i = 1, \ldots, n$$

mit zufälligen Fehlern $\epsilon_1, \ldots, \epsilon_n$ mit verschwindendem Erwartungswert $\mathrm{E}(\epsilon_i) = 0$, also $\mathrm{E}(Y_i) = \beta_0 + \beta_1 x_i$. Die wahren Parameter β_0 und β_1 sind unbekannt und müssen geschätzt werden. Für Realisationen $(x_1, y_1), \ldots, (x_n, y_n)$ geschieht dies häufig durch den Kleinste-Quadrate-Schätzer (engl. least squares estimator) $\hat{\beta}_0^{\mathrm{LSE}}$, $\hat{\beta}_1^{\mathrm{LSE}}$, der die Summe der quadratischen Abweichungen

$$\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

minimiert. Er hat viele angenehme Eigenschaften und ist einfach zu berechnen, ist allerdings nicht sehr robust. Eine Alternative ist der Schätzer der kleinsten absoluten Abweichungen (engl. least absolute deviations) $\hat{\beta}_0^{\mathrm{LAD}}$, $\hat{\beta}_1^{\mathrm{LAD}}$, der die Summe der absoluten Abweichungen

$$\sum_{i=1}^{n} |y_i - \beta_0 - \beta_1 x_i|$$

minimiert. Er ist robuster als der LSE und besitzt ebenfalls viele angenehme Eigenschaften. Allerdings ist er nicht so einfach zu berechnen wie der LSE, insbesondere nicht geschlossen darstellbar.

Man kann die Robustheit durch sogenanntes *Trimmen* noch steigern. Trimmen bedeutet, eine gewisse Anzahl der Beobachtungen $(x_1, y_1), \ldots, (x_n, y_n)$, die besonders schlecht zum Modell passen (*Ausreißer* genannt), zu ignorieren. Der TLAD-Schätzer (trimmed least absolute deviations) soll auch für Regressionsmodelle mit sogenannten *latenten Klassen* untersucht werden. Dabei folgen die Beobachtungen nicht einem linearen Zusammenhang, sondern zerfallen in a priori unbekannte Klassen, die alle ihre eigene Regressionsgerade haben.

Die Bestimmung der entsprechenden Schätzer ist also in der Statistik von Interesse. Man kann dies aber auch als Problem der Standortplanung auffassen: Platziere K Geraden, um n Punkte gut anzunähern. Ignoriere aber alle bis auf m der gegeben Punkte. Verwende als Abstandsmaß die absolute Abweichung. Ziel der Arbeit ist es, dieses Vorgehen statistisch zu rechtfertigen und Lösungsvorschläge aus Sicht der Standortplanung zu finden.

Einige Definitionen

nicht-vertikale Gerade $\ell(\beta_0, \beta_1) = \{(x, y) \in \mathbb{R}^2 : y = \beta_0 + \beta_1 x\}$

vertikales Residuum $r_{\mathrm{v}}(z, l) = y - \beta_0 - \beta_1 x$ für einen Punkt $z = (x, y) \in \mathbb{R}^2$ und eine Gerade $l = \ell(\beta_0, \beta_1)$

vertikaler Abstand $d_{\mathrm{v}}(z, l) = |r_{\mathrm{v}}(z, l)|$

LAD-Problem für Punkte $z_1, \ldots, z_n \in \mathbb{R}^2$:

$$\text{(LAD)} \qquad \min_{(\beta_0,\beta_1)\in\mathbb{R}^2} \quad \sum_{i=1}^{n} d_{\mathrm{v}}(z_i, \ell(\beta_0,\beta_1))$$

TLAD-Problem für Punkte $z_1, \ldots, z_n \in \mathbb{R}^2$ und eine Zahl $m : 0 < m \leq n$. Ordered-median-Formulierung:

$$\text{(TLAD)} \qquad \min_{(\beta_0,\beta_1)\in\mathbb{R}^2} \quad \sum_{i=1}^{m} d_{\mathrm{v}}(z_{(i)}, \ell(\beta_0,\beta_1))$$

mit einer Permutation $(\cdot)$ mit

$$d_{\mathrm{v}}(z_{(1)}, \ell(\beta_0,\beta_1)) \leq \ldots \leq d_{\mathrm{v}}(z_{(n)}, \ell(\beta_0,\beta_1))$$

Kombinatorische Formulierung:

$$\text{(TLAD)} \qquad \min_{(\beta_0,\beta_1)\in\mathbb{R}^2} \quad \min_{\substack{I\subset\{1,\ldots,n\}\\ |I|=m}} \quad \sum_{i\in I} d_{\mathrm{v}}(z_i, \ell(\beta_0,\beta_1))$$

K-**TLAD-Problem** kombinatorische Formulierung:

$$(K\text{-TLAD}) \qquad \min_{(\beta_{01},\beta_{11}),\ldots,(\beta_{0K},\beta_{1K})\in\mathbb{R}^2} \quad \min_{\substack{I\subset\{1,\ldots,n\}\\ |I|=m}} \quad \sum_{i\in I} \min_{k=1,\ldots,K} d_{\mathrm{v}}(z_i, \ell(\beta_{0k},\beta_{1k}))$$

Duale Transformation * Für Punkte $z = (x, y) \in \mathbb{R}^2$ und Geraden $\ell(\beta_0, \beta_1)$

$$z^* = \ell(-y, x) \qquad \text{und} \qquad l^* = (\beta_1, -\beta_0)$$

Die duale Transformation ist idempotent, bijektiv zwischen Punkten und nicht-vertikalen Geraden und erhält die Residuen: $r_{\mathrm{v}}(z, l) = r_{\mathrm{v}}(l^*, z^*)$.

Duale Parkettierung Eine duale Zelle zu einer Punktmenge $P = \{z_1, \ldots, z_n\} \subset \mathbb{R}^2$ ist gegeben durch

$$C(\sigma) = \{l^* \in \mathbb{R}^2 : \sigma_i \cdot r_{\mathrm{v}}(l^*, z_i^*) \geq 0 \quad \forall\, i = 1, \ldots, n\}$$

falls $|C(\sigma)| > 1$. Die duale Parkettierung ist die Menge der dualen Zellen:

$$DT(P) = \{C(\sigma) : \sigma \in \{0,1\}^n,\ |C(\sigma)| > 1\}$$

Literatur

[1] Bertsimas, D. ; Shioda, R.: Classification and Regression via Integer Optimization. In: *Operations Research* 55 (2007), März–April, Nr. 2, S. 252–271

[2] Lozano, A. J. ; Plastria, F.: *The Ordered Median Euclidean Straight-line Location Problem.* Juni 2007. – working paper

[3] Neykov, N. ; Filzmoser, P. ; Dimova, R. ; Neytchev, P.: Robust Fitting of Mixtures Using the Trimmed Likelihood Estimator. In: *Computational Statistics & Data Analysis* 52 (2007), September, Nr. 1, S. 299–308

[4] Nickel, S. ; Puerto, J.: *Location Theory. A Unified Approach.* 1. Auflage. Berlin : Springer Verlag, Mai 2005

[5] Schöbel, A.: *Locating Lines and Hyperplanes: Theory and Algorithms.* 1. Auflage. Dordrecht, Niederlande : Kluwer Academic Publishers, 1999