Überbestimmte lineare Gleichungssysteme und lineare Regression

Steffen Solyga*

15. - 19. Juni 2006

Vorgelegt seien m reelle Zahlen y_i und mn reelle Zahlen a_{ij}, $i = 1, 2, \dots, m$, $j = 1, 2, \dots, n$. Gesucht sind n reelle Zahlen x_j, so daß gilt

$$\begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \tag{1}$$

oder kurz

$$\mathbf{Ax} = \mathbf{y}. \tag{2}$$

Die lineare Algebra lehrt, daß (2) genau dann lösbar ist, wenn der Rang der Matrix $\mathbf{A}$ kleiner oder gleich n ist. Im ersten Falle, $\mathrm{r}(\mathbf{A}) < n$, existieren unendlich viele Lösungen $\mathbf{x}$, und im zweiten Falle, $\mathrm{r}(\mathbf{A}) = n$, ist die Lösung $\mathbf{x}$ eindeutig bestimmt. Ist jedoch $\mathrm{r}(\mathbf{A}) > n$, was $m > n$ impliziert, ist das System nicht lösbar; man nennt das Gleichungssystem oder $\mathbf{x}$ dann auch überbestimmt.

In diesem Falle kann man aber vielleicht ein $\mathbf{x}$ finden, das (2) wenigstens näherungsweise löst, d.h. den Vektor $\mathbf{y} - \mathbf{Ax}$ möglichst „klein“ macht. Als Maß für die „Größe“ eines m-zeiligen Vektors $\mathbf{v}$ bietet sich das Produkt mit sichselbst an:

$$\mathbf{v}^{\mathrm{T}}\mathbf{v} = v_1^2 + v_2^2 + \cdots + v_m^2. \tag{3}$$

Dies ist offenbar ein quadratisches Maß, weshalb man von einer Näherung nach der *Methode der kleinsten Quadrate* spricht. Wir verlangen also, daß

$$f(\mathbf{x}) = (\mathbf{y} - \mathbf{Ax})^{\mathrm{T}}(\mathbf{y} - \mathbf{Ax}) = \mathbf{y}^{\mathrm{T}}\mathbf{y} - 2\mathbf{y}^{\mathrm{T}}\mathbf{Ax} + \mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{Ax} \tag{4}$$

– eine reellwertige Funktion des variablen Vektors $\mathbf{x}$ – minimal wird. Dann müssen notwendigerweise alle partiellen Ableitungen von f verschwinden

$$\mathbf{grad}\, f(\mathbf{x}) = -2\mathbf{y}^{\mathrm{T}}\mathbf{A} + 2\mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{0}^{\mathrm{T}}, \tag{5}$$

und durch Transponieren erhält man daraus das quadratische System

$$\mathbf{A}^{\mathrm{T}}\mathbf{Ax} = \mathbf{A}^{\mathrm{T}}\mathbf{y}. \tag{6}$$

*solyga@gmx.de

Jede geeignete Näherungslösung $\mathbf{x}$ von (2) muß also der Gleichung (6) genügen. Hinreichend für ein Minimum von f ist, daß außerdem die Matrix der zweiten Ableitungen

$$\begin{pmatrix} f_{x_1x_1} & \cdots & f_{x_1x_n} \\ \vdots & \ddots & \vdots \\ f_{x_nx_1} & \cdots & f_{x_nx_n} \end{pmatrix} = \mathbf{A}^{\mathrm{T}}\mathbf{A} \tag{7}$$

positiv definit ist. Nun ist $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ sicherlich positiv semidefinit, denn $Q(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{x} = (\mathbf{A}\mathbf{x})^{\mathrm{T}}(\mathbf{A}\mathbf{x})$ kann als Summe von Quadraten für kein $\mathbf{x}$ negativ werden. Eine positiv semidefinite Matrix ist aber genau dann positiv definit, wenn sie regulär ist.[1] D.h. $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ ist genau dann positiv definit, wenn (6) eindeutig lösbar ist, weshalb gilt:

Satz 1 *Besitzt (6) eine eindeutig bestimmte Lösung* $\mathbf{x}$*, so handelt es sich dabei gerade um die Näherungslösung von (2) im Sinne der Methode der kleinsten Quadrate.*

Ferner verschwindet $Q(\mathbf{x})$ dann und nur dann, wenn $\mathbf{A}\mathbf{x} = \mathbf{0}$ ist. Folglich ist $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ genau dann regulär, wenn $\mathrm{r}(\mathbf{A}) \geq n$ gilt. Damit ergibt sich für die ggf. näherungsweise Lösung von (2) folgendes Bild:

1. Ist $\mathrm{r}(\mathbf{A}) < n$, so besitzt (2) beliebig viele (exakte) Lösungen $\mathbf{x}$. Diese lösen dann natürlich auch (6), was man durch Vormultiplikation von (2) mit $\mathbf{A}^{\mathrm{T}}$ sofort einsieht. Die Lösungsmengen müssen aber nicht identisch sein.

2. Ist $\mathrm{r}(\mathbf{A}) = n$, so besitzen (2) und (6) dieselbe (exakte) Lösung $\mathbf{x}$.

3. Ist $\mathrm{r}(\mathbf{A}) > n$, so ist (2) zwar nicht exakt, aber im Sinne der Methode der kleinsten Quadrate näherungsweise lösbar. Die Näherungslösung $\mathbf{x}$ ist als Lösung von (6) eindeutig bestimmt.

Ein Spezialfall der eben besprochenen Ausgleichsrechnung ist die lineare Regression: Gegeben seien zwei n-zeilige Vektoren $\mathbf{x}$ und $\mathbf{y}$. Gesucht sind reelle Zahlen p und q, so daß

$$f(p,q) = \sum_{i=1}^{n}[y_i - (px_i + q)]^2 \tag{8}$$

minimal wird. Dazu betrachten wir das aus n Gleichungen bestehende System

$$\begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}\begin{pmatrix} p \\ q \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \tag{9}$$

oder kurz

$$\mathbf{A}\mathbf{p} = \mathbf{y}. \tag{10}$$

Im Falle $n = 1$ existieren unendlich viele (exakte) Lösungen, und im Falle $n = 2$ existiert eine eindeutig bestimmte Lösung, sofern $x_1 \neq x_2$. Die Ausgleichsrechnung greift im Falle $\mathrm{r}(\mathbf{A}) \geq 2$, und zwar erhalten wir die beste Näherungslösung von (10) als Lösung von

$$\mathbf{A}^{\mathrm{T}}\mathbf{A}\mathbf{p} = \mathbf{A}^{\mathrm{T}}\mathbf{y}. \tag{11}$$

[1] Beweis: Eine Matrix ist genau dann positiv definit (semidefinit), wenn sie nur positive (nichtnegative) Eigenwerte besitzt (Satz). Andererseits ist eine Matrix genau dann regulär, wenn keiner ihrer Eigenwerte verschwindet (Satz).

Wegen

$$\mathbf{A}^\mathrm{T}\mathbf{A} = \begin{pmatrix} x_1 & \dots & x_n \\ 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} x_1 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} x_i^2 & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & n \end{pmatrix} \tag{12}$$

und

$$\mathbf{A}^\mathrm{T}\mathbf{y} = \begin{pmatrix} x_1 & \dots & x_n \\ 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^{n} x_i y_i \\ \sum_{i=1}^{n} y_i \end{pmatrix} \tag{13}$$

folgt nach der CRAMERschen Regel

$$\mathbf{p} = \begin{pmatrix} p \\ q \end{pmatrix} = \frac{1}{n \sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i \right)^2} \begin{pmatrix} n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i \end{pmatrix}. \tag{14}$$

Mit $n\overline{x} := \sum_{i=1}^{n} x_i$ und $n\overline{y} := \sum_{i=1}^{n} y_i$ erhält man daraus die bekannten Formeln

$$p = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{xy}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}, \tag{15}$$

$$q = \frac{\overline{y} \sum_{i=1}^{n} x_i^2 - \overline{x} \sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}. \tag{16}$$

Literatur

[1] Karl Brammer; Gerhard Siffling:
Kalman-Bucy-Filter, Deterministische Beobachtung und stochastische Filterung.
R. Oldenbourg Verlag München/Wien 1975
ISBN 3-486-34661-X