REKURSIONEN ZUR METHODE DER KLEINSTEN QUADRATE

von Ian Grant, South Australian Institute of Technology
Originaltitel in "Teaching Statistics" Vol.9 (1987) Nr.1:
Recursive Least Squares
Übertragung: Bernd Wollring, Universität Münster

Es gibt heutzutage viele und weitgehende Kritik zu dem schädigenden Einfluß, den Computer auf die arithmetischen Fähigkeiten von Lernenden haben. Natürlich müssen Computer in einem Bereich wie der Statistik eine wichtige Rolle spielen, aber etwa der Aufsatz von SEARLE (1983) erinnert daran, daß sorgfältige Vorplanung im Bereich der Arithmetik oft die Genauigkeit von Schätzvorgängen steigern kann, die vom Computer ausgeführt werden.

Es gibt im Bereich der technischen Umwelt eine Menge von Problemen - das Steuern von Supertankern und ökonomische Planungen seien als Beispiele genannt -, die die Lernenden gerne diskutieren und wo sie schnell wahrnehmen, daß die Struktur und Qualität eines beschreibenden mathematischen Modells regelmäßig auf den neuesten Stand zu bringen sind, sobald neue Daten verfügbar sind. Wenn die Struktur des Modells gut bekannt ist und neue Daten in regelmäßigen Abständen zu Verfügung stehen, so sind rekursive oder gar "on-line"-Prozeduren besonders nützlich. Viele rekursive Algorithmen erfordern weitgehende arithmetische Manipulationen, um sie auf einem Computer einfach verfügbar zu machen. Hat man aber diese Manipulationen ein für allemal durchgeführt, so kann man das Modell sehr zügig auf den neuesten Stand bringen und braucht minimalen Platz zum Abspeichern von Daten.

Anpassen einer Geraden

Wir betrachten das Problem, mit Hilfe der Methode der kleinsten Quadrate eine Gerade an eine Punktserie $(x_1;y_1),\ldots,(x_n;y_n)$ anzupassen. Wir können die Gleichungen

der angepaßten Geraden für die jeweils ersten n Punkte ($n \geqslant 2$) wie folgt notieren:

$$y_i = A_n + B_n(x_i - M_n) \qquad (i = 1, 2, \ldots, n) \tag{1}$$

Dabei ist

$$M_n = (x_1 + x_2 + \cdots + x_n)/n = \sum_{i=1}^{n} x_i/n$$

Aufgrund der Methode der kleinsten Quadrate erhält man für A_n und B_n die bekannten Formeln

$$nA_n = \sum_{i=1}^{n} y_i \tag{2}$$

und

$$\sum_{i=1}^{n} (x_i - M_n)^2 B_n = \sum_{i=1}^{n} y_i(x_i - M_n) \tag{3}$$

Mit der Notierung von SEARLE, nämlich

$$S_n^2 = \textstyle\sum_{i=1}^{n} (x_i - M_n)^2$$

können wir (3) in folgender Form schreiben:

$$S_n^2 B_n = \sum_{i=1}^{n} y_i(x_i - M_n) \tag{4}$$

Einbinden zusätzlicher Daten

Angenommen sei nun, wir erhalten einen weiteren Datenpunkt $(x_{n+1};y_{n+1})$ und wollen unsere Werte für den Achsenabschnitt A und Steigung B auf den neuesten Stand bringen, der auf allen n+1 Punkten basiert. Um A_n zu korrigieren, kehren wir zu (2) zurück zu schreiben:

$$(n+1)A_{n+1} = \sum_{i=1}^{n+1} y_i = \sum_{i=1}^{n} y_i + y_{n+1}$$

oder

$$(n+1)A_{n+1} = nA_n + y_{n+1}$$

und erhalten

$$A_{n+1} = A_n + (y_{n+1} - A_n)/(n+1) \tag{5}$$

Um B_n zu korrigieren, kehren wir zu (4) zurück und schreiben:

$$\begin{aligned} S_{n+1}^2 B_{n+1} &= \sum_{i=1}^{n+1} y_i(x_i - M_{n+1}) \\ &= \sum_{i=1}^{n} y_i(x_i - M_{n+1}) + y_{n+1}(x_{n+1} - M_{n+1}) \end{aligned} \tag{6}$$

Daraus folgt mit Hilfe von (2) und (4):

$$\begin{aligned} \sum_{i=1}^{n} y_i(x_i - M_{n+1}) &= \sum_{i=1}^{n} y_i(x_i - M_n + M_n - M_{n+1}) \\ &= \sum_{i=1}^{n} y_i(x_i - M_n) + (M_n - M_{n+1}) \sum_{i=1}^{n} y_i \\ &= S_n^2 B_n + n(M_n - M_{n+1})A_n \end{aligned}$$

Wir finden ferner - etwa mit Hilfe von SEARLE:

$$x_{n+1} = (n+1)M_{n+1} - nM_n \quad ,$$

so daß der Term

$$y_{n+1}(x_{n+1} - M_{n+1})$$

aus (6) folgendermaßen geschrieben werden kann:

$$y_{n+1}[(n+1)M_{n+1} - nM_n - M_{n+1}] = n(M_{n+1} - M_n)y_{n+1}$$

Sammeln wir nun alle Terme, so erhalten wir Gleichung (6) in folgender neuer Form:

$$\begin{aligned} S_{n+1}^2 B_{n+1} &= S_n^2 B_n + n(M_n - M_{n+1})A_n + n(M_{n+1} - M_n)y_{n+1} \\ &= S_n^2 B_n + n(M_n - M_{n+1})(A_n - y_{n+1}) \end{aligned} \tag{7}$$

Die Rekursionsformeln (5) und (7) können mit n=2 gestartet werden. Sie ergeben dann eine Gerade durch $(x_1;y_1)$ und $(x_2;y_2)$. Auf den neuesten Stand mit jeweils mehr Punkten

bringt man sie dann so, wie es das folgende vollständig durchgeführte numerische Beispiel andeutet.

Ein Beispiel

i	x_i	y_i
1	1	3
2	5	11
3	6	12
.	.	.
.	.	.

1) Die Gerade mit $y_i = 1 + 2x_i = 7 + 2(x_i - 3)$ führt genau durch die beiden Punkte (1;3) und (5;11), es gilt:

$$A_2 = 7 \quad \text{and} \quad B_2 = 2$$

Zusätzlich benötigen wir

$$M_2 = (1+5)/2 = 3$$

und

$$S_2^2 = (1-3)^2 + (5-3)^2 = 8$$

2) Nun bringen wir M, S, A und B auf den neuen Stand, bei dem wir den Punkt (6;12) hinzunehmen. Nach SEARLE gilt

$$M_{n+1} = M_n + (x_{n+1} - M_n)/(n+1) \tag{8}$$

Also hier:

$$M_3 = 3 + (6-3)/3 = 4$$

Ferner gilt:

$$S_{n+1}^2 = S_n^2 + n(n+1)(M_{n+1} - M_n)^2 \tag{9}$$

Also hier:

$$S_3^2 = 8 + 6(4-3)^2 = 14$$

Aus (5) folgt:

$$A_3 = (2A_2 + y_3)/3 = (14+12)/3 = 26/3$$

Aus (7) folgt:

$$S_3^2 B_3 = S_2^2 B_2 + 2(M_2 - M_3)(A_2 - y_3)$$
$$14 B_3 = 16 + 2(3-4)(7-12)$$
$$B_3 = 13/7$$

Die Gleichung der neuen Ausgleichsgraden ist:

$$y_i = 26/3 + 13(x_i - 4)/7 = 26/21 + 13x_i/7$$

Man kann dies durch direkte Berechnung leicht bestätigen. Zu beachten ist noch, daß, egal wie umfangreich die "historischen" Daten waren, nur n, M_n, $S^2{}_n$, A_n und B_n, also fünf Werte für das Erneuern zu ändern sind. Es ist (für Studenten) eine leichte Übung, ein Computerprogramm zu diesem Algorithmus zu schreiben.

Zusammenfassung

Bei einfacher linearer Regression kann beim Erhalt neuer beobachteter Punkte die Regressionsgerade mit Hilfe von (5), (7), (8) und (9) leicht auf den neuesten Stand gebracht werden. Man kann dieses Verfahren auf Multiple Lineare Regression ausdehnen. In neueren Texten, etwa HARVEY (1981) wird der Nutzen eines solchen Ansatzes diskutiert. Die zunehmende Bedeutung rekursiver Verfahren für Ingenieure, Ökonomen und andere Wissenschaftler legen es nahe, daß diese Verfahren auch in frühen Statistik-Kursen angesprochen werden.

LITERATUR

(1) HARVEY, A.C. (1981). The Econometric Analysis of Time Series, Philip Allan, Oxford.
(2) SEARLE, S.A. (1983). The Recurrence formulae for means and variances. Teaching Statistics, Vol.5, No.1, 7-10.