Kurvenanpassung durch Regression (3)

Ac 2017

- nichtlineare Regression/Linearisierung -

Für Probleme, die eine nicht lineare (und nicht polynomiale) Anpassungsfunktion nahelegen, ist eine direkte Berechnung nach der Methode der kleinsten Quadrate sehr aufwändig.
Zum Beispiel erhielte man für eine exponentielle Regression mit $f(x) = a \cdot e^{kx}$ folgendes:

$$SQ = \sum_{i=1}^{n} (y_i - a \cdot e^{k \cdot x_i})^2 \quad \Rightarrow$$

$$\frac{dSQ}{da} = \sum_{i=1}^{n} 2 \cdot (y_i - a \cdot e^{k \cdot x_i}) \cdot (-e^{k \cdot x_i}) = 0 \ \wedge \ \frac{dSQ}{dk} = \sum_{i=1}^{n} 2 \cdot (y_i - a \cdot e^{k \cdot x_i}) \cdot (-a \cdot x_i \cdot e^{k \cdot x_i}) = 0$$

Also ein nicht lineares LGS, was nur mit sehr viel Aufwand zu lösen ist !

Diese Probleme kann man in vielen Fällen umgehen, indem man die nicht lineare Regression durch geeignete Umformung auf die lineare Regression zurückführt („Linearisierung“) .

Beispiele für linearisierbare Anpassungen sind:

(1) Exponentielle Regression mit $f(x) = a \cdot e^{kx}$

(2) Logarithmische Regression mit $f(x) = a + c \cdot \ln(x)$

(3) Potenz Regression mit $f(x) = a \cdot x^c$

(4) Logistische Regression mit $f(x) = c / (1 + a \cdot e^{dx})$; $d < 0$

(5) Gebrochen rationale Regression mit $f(x) = 1 / (ax + c)$ bzw. $f(x) = x / (ax + c)$

Anmerkung:
Bekannte Systeme wie GeoGebra oder TI84 oder CASIO Classpad verwenden ebenfalls die Linearisierung für (1) bis (3). (4) wird mit einer anderen Methode gelöst, (5) gar nicht !

Die Vorgehensweise besteht in allen Fällen aus 3 Schritten:

- Linearisierung, d.h. f(x) in eine lineare Funktion umwandeln, z.B. durch Logarithmieren
  Achtung: Auch die y-Daten (evtl. x-Daten) müssen entsprechend umgewandelt werden !
- Lineare Regression auf die angepassten Daten anwenden; Ergebnis: Lineare Funktion
- Linearisierung wieder rückgängig machen

Dies soll an allen oben angeführten Beispielen demonstriert werden.

Die verwendete Formel für die Lineare Regression lautet bekanntlich:

$$y = m \cdot x + b \quad mit \quad m = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \ \wedge \ b = \bar{y} - m \cdot \bar{x}$$

1) Exponentiell: $f(x) = a \cdot e^{kx}$

Durch Logarithmieren der y_i folgt: $\ln[f(x)] = \ln(a \cdot e^{kx}) = \ln(a) + k \cdot x$.
Alle $y_i > 0$ vorausgesetzt !

Datenbeispiel (y_i sind Näherungswerte von $f(x) = 3e^{0,2x}$) :

x	1	2	3	4
y	3,7	4,5	5,4	6,7
ln(y)	1,3..	1,5..	1,6..	1,9..

Die Regression liefert $y = 0{,}196x + 1{,}109$, also $m = 0{,}196$ und $b = 1{,}109$.
In der Regressionsformel muss also <u>k für m</u> und <u>ln(a) für b</u> eingesetzt werden.
Somit gilt : $a = e^b = e^{1,109} = 3{,}03...$

Ergebnis für die Anpassungsfunktion: $\mathbf{f(x) = 3{,}03 \cdot e^{0,2x}}$

2) Logarithmisch: $f(x) = a + c \cdot \ln(x)$

Hier ist es günstig, die x_i zu Logarithmieren und dann die lin. Regression durchzuführen.
Alle $x_i > 0$ vorausgesetzt !
Das Ergebnis $y = cx + d$ ist dann zu interpretieren als $f(x) = d + c \cdot \ln(x)$:

x	1	2	3	4
y	1	3,1	4,3	5,1
ln(x)	0	0,69..	1,09..	1,38..

Die Regression liefert $y = 2{,}9671 \cdot x + 1{,}0176$
Ergebnis für die Anpassungsfunktion: $\mathbf{f(x) = 1{,}02 + 2{,}97 \cdot \ln(x)}$

3) Potenz: $f(x) = a \cdot x^c$

Hier ist es günstig, sowohl die x_i als auch die y_i zu Logarithmieren und dann die lin. Regression durchzuführen.
Alle y_i , $y_i > 0$ vorausgesetzt !
$\ln[f(x)] = \ln(a \cdot x^c) = \ln(a) + \ln(x^c) = \ln(a) + c \cdot \ln(x)$.
Das Ergebnis $y = g \cdot x + h$ ist dann zu interpretieren als $f(x) = e^h \cdot x^g$:

Datenbeispiel (y_i sind Näherungswerte von $f(x) = 1{,}5 \cdot x^{0,5}$) :

x	1	2	3	4
y	1,5	2,1	2,6	3
ln(x)	0	0,69..	1,09..	1,38..
ln(y)	0,40..	0,74..	0,95..	1,09..

Die Regression liefert $y = 0{,}50115 \cdot x + 0{,}4022$
Ergebnis für die Anpassungsfunktion: $\mathbf{f(x) = 1{,}5 \cdot x^{0,5}}$

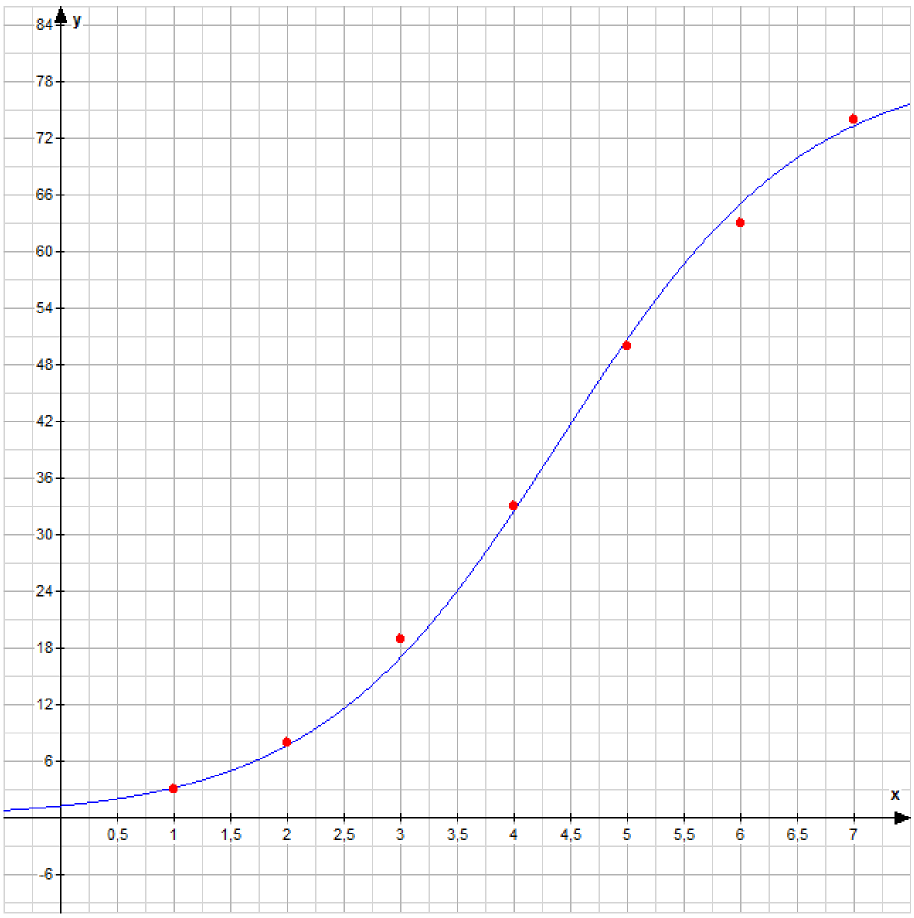
4) <u>Logistisch :</u> **$f(x) = g / (1 + a \cdot e^{d \cdot x})$; $d < 0$**

Man bildet **$\ln(g / f(x) - 1)$** $= \ln(a \cdot e^{dx}) = \ln(a) + \ln(e^{dx}) =$ **$\ln(a) + d \cdot x$**
g = „Sättigungsgrenze (Grenzwert der logist.Fkt.)“ muss bekannt sein !
Anmerkung: Bei unbekanntem g kann man eine Probiermethode anwenden.
Man führt das Verfahren für eine Reihe von g-Werten durch und wählt dasjenige Ergebnis, für das die Summe der quadratischen Abweichungen minimal ist. $\text{sum}((y_i - f(x_i))^2$ minimal !
Alle $y_i \neq 0$ sowie $g / y_i > 1$ vorausgesetzt !

Im folgenden Beispiel sei die Sättigungsgrenze $g = 80$.
In der Tabelle wird daher auch $\ln(80 / y_i - 1)$ eingetragen !

x	1	2	3	4	5	6	7
y	3	8	19	33	50	63	74
ln(80/y - 1)	3,24..	2,19..	1,16..	0,35..	-0,51..	-1,30..	-2,51..

Die lin. Regression liefert: $y = -0{,}927857 \cdot x + 4{,}0848$
Es gilt dann: $a = e^{4{,}088571} = 59{,}6$ $d = -0{,}927287$
Lösung: **$f(x) = 80 / (1 + 59{,}43 \cdot e^{-0{,}927287 \cdot x})$**



<u>Hinweise:</u>

Obige „Probiermethode“ (ohne Kenntnis der Sättigungsgrenze) liefert:
c = 80,887 a = 57,784 d = -0,91046

Geogebra / TI84 / CASIO Classpad liefern die beste Approximation (ohne Kenntnis der Sättigungsgrenze). Allerdings wird hierbei nicht die Linearisierungsmethode verwendet.
c = 81,813 a = 45,147 d = -0,850736

5.1) Gebrochen rational(1) : $f(x) = 1 / (a \cdot x + c)$

Linearisierung durch Kehrwertbildung der yi ! Alle $y_i \neq 0$ vorausgesetzt !
$1 / f(x) = ax + c$; man erhält also sofort eine lineare Funktion.

Datenbeispiel (y_i sind Näherungswerte von $f(x) = 1 / (0{,}5x + 2)$) :

x	1	2	3	4
y	0,4	0,33	0,29	0,25
1/y	2,5	3,0..	3,4..	4

Die Regression liefert $y = 0{,}4918 \cdot x + 2{,}01515$
Ergebnis für die Anpassungsfunktion: **$f(x) = 1 / (0{,}5 \cdot x + 2)$**

5.2) Gebrochen rational(2) : $f(x) = x / (a \cdot x + c)$

Linearisierung durch Bildung von xi / yi ! Alle x_i , $y_i \neq 0$ vorausgesetzt !
$x / f(x) = a \cdot x + c$, eine lineare Funktion

Datenbeispiel (y_i sind Näherungswerte von $f(x) = x / (0{,}5x + 2)$) :

x	1	2	3	4
y	0,4	0,67	0,86	1
x/y	2,5	2,9..	3,48..	4

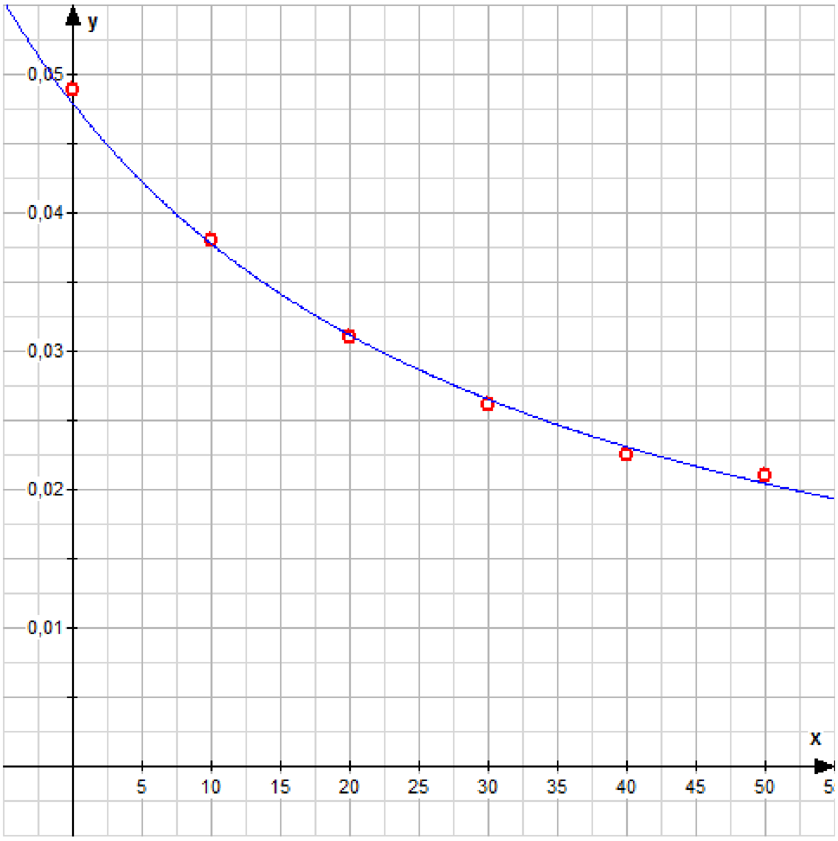
Die Regression liefert $y = 0{,}5003 \cdot x + 1{,}9925$
Ergebnis für die Anpassungsfunktion: **$f(x) = = x / (0{,}5 \cdot x + 2)$**

Weitere Beispiele für das Modell „gebrochen rational (1)“ :

Bsp 1:
Die Lösbarkeit von Sauerstoff in Wasser hängt unter anderem von der Wassertemperatur ab. 1 Liter Wasser absorbiert die in der Tabelle angegebenen Sauerstoffmengen.
Gesucht ist eine gebrochen rationale Modellierungsfunktion f .

Temperatur in °C	0	10	20	30	40	50
Sauerstoff in cm³	0,0489	0,038	0,031	0,0261	0,0225	0,021
1 / y						

Lösung: f(x) = 1 / (0,56082 * x + 20,87968)



Bsp 2: Zu den folgenden Daten ist eine „gebrochen rationale“ Anpassung gesucht:

x	0	1	2	3	4	5	6	7	8	9
y	5	2,4	1,8	1,2	1,1	0,9	0,7	0,4	0,4	0,4

f(x) = 1 / (ax + c) und Kehrwertbildung der yi :

x	0	1	2	3	4	5	6	7	8	9
y	5	2,4	1,8	1,2	1,1	0,9	0,7	0,4	0,4	0,4
1/y ≈	0,2	0,417	0,556	0,833	0,909	1,111	1,429	2	2,5	2,5

Die Lineare Regression zwischen x und 1/y liefert y ≈0,2697x + 0,032, was durch erneute Kehrwertbildung zu f(x) = 1 / (0,2697x + 0,032) führt.
Dieses Ergebnis ist (insbesondere für die Anfangswerte) eher schlecht !