Mathematik für Studierende der Biologie und des Lehramtes Chemie

Jens Horn

basierend auf einer Vorlage
von Dominik Schillo

Universität des Saarlandes

10.01.2020

Beispiel

Wir haben folgende Messdaten gegeben:

t	-1	0	1
y	2	1	-3

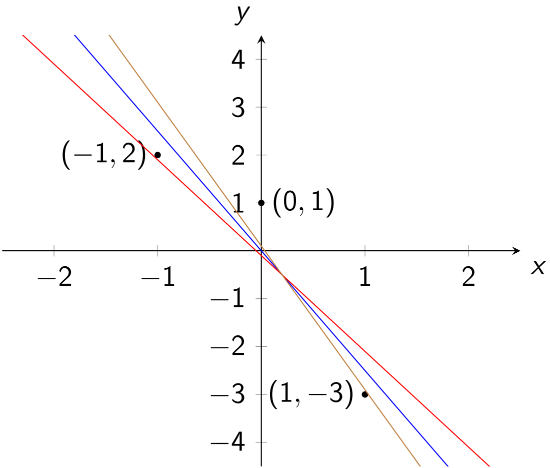
Aus theoretischen Überlegungen vermuten wir, dass sich die Messdaten durch eine Funktion der Form

$$f_{x_1,x_2} : \mathbb{R} \to \mathbb{R}, \ t \mapsto x_1 + x_2 t$$

mit geeigneten Parametern $x_1, x_2 \in \mathbb{R}$ beschreiben lässt.
Da f_{x_1,x_2} von 2 Parametern abhängt und wir 3 Messpunkte haben, werden wir (vermutlich) keine Parameter finden, sodass f_{x_1,x_2} alle Messpunkte interpoliert, d.h. die Gleichungen

$$f_{x_1,x_2}(-1) = 2, \ f_{x_1,x_2}(0) = 1, \ f_{x_1,x_2}(1) = -3$$

werden (i. A.) nicht alle gleichzeitig erfüllt. Können wir Parameter finden, sodass f_{x_1,x_2} die Messdaten „bestmöglich“ approximiert?



y
x
4
3
2
1
−1
−2
−3
−4
−2
−1
1
2
(−1, 2)
(0, 1)
(1, −3)

Um ein Maß für den Fehler der Funktion von der Messung zu definieren, brauchen wir zunächst einen Abstandsbegriff in höheren Dimensionen.

Definition

Für einen Vektor $w = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{R}^n$ nennen wir die positive reelle Zahl

$$|w| = \sqrt{w_1^2 + \ldots + w_n^2}$$

den *Betrag von w*.

Sei $x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$. Wir schauen uns

$$\begin{aligned}
r(x) &= \left| \begin{pmatrix} f_{x_1,x_2}(t_1) \\ f_{x_1,x_2}(t_2) \\ f_{x_1,x_2}(t_3) \end{pmatrix} - \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \right| = \left| \begin{pmatrix} f_{x_1,x_2}(-1) \\ f_{x_1,x_2}(0) \\ f_{x_1,x_2}(1) \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} \right| \\
&= \left| \begin{pmatrix} x_1 - x_2 \\ x_1 \\ x_1 + x_2 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} \right| = \left| \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} \right| \\
&= |Ax - y|
\end{aligned}$$

an, wobei

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}, \ y = \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix}.$$

Wir suchen nun einen Vektor $x^* \in \mathbb{R}^2$, der

$$r(x^*) \leq r(x)$$

für alle $x \in \mathbb{R}^2$ erfüllt, d. h. wir wollen die Funktion

$$r\colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto |Ax - y|$$

minimieren.
Wie finden wir $x^* \in \mathbb{R}^2$?
Wir müssen uns also überlegen, wie wir einen Tiefpunkt einer Funktion mit Definitionsbereich $\mathbb{R}^2$ berechnen können.
Da wir für eine Funktion $f\colon \mathbb{R} \to \mathbb{R}$ die Ableitung berechnen um einen Tiefpunkt zu bestimmen, brauchen wir zunächst einen Begriff der Ableitung für Funktionen mit Definitionsbereich $\mathbb{R}^2$.

Definition

Seien $n \in \mathbb{N}$, $f\colon \mathbb{R}^n \to \mathbb{R}$ und $a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{R}^n$. Definiere

$$\begin{aligned} g_{a,1}&\colon \mathbb{R} \to \mathbb{R},\ x \mapsto f(x, a_2, \ldots, a_n), \\ g_{a,2}&\colon \mathbb{R} \to \mathbb{R},\ x \mapsto f(a_1, x, a_3, \ldots, a_n), \\ &\vdots \\ g_{a,n}&\colon \mathbb{R} \to \mathbb{R},\ x \mapsto f(a_1, \ldots, a_{n-1}, x). \end{aligned}$$

Für $i \in \{1, \ldots, n\}$ heißt

$$\begin{aligned} \partial_i f(a) = g'_{a,i}(a_i) &= \lim_{h \to 0} \frac{g_{a,i}(a_i + h) - g_{a,i}(a_i)}{h} \\ &= \lim_{h \to 0} \frac{f(a + he_i) - f(a)}{h} \end{aligned}$$

die *i-te partielle Ableitung von f in a*.

Definition

Falls die i-te partielle Ableitung von f in a für alle $i = 1, \ldots, n$ existiert, so nennen wir f *partiell differenzierbar in* a. Ist f in jedem Punkt partiell differenzierbar, so heißt f *partiell differenzierbar*.
Falls f partiell differenzierbar in a ist, nennen wir den Vektor

$$\operatorname{grad}(f)(a) = \begin{pmatrix} \partial_1 f(a) \\ \vdots \\ \partial_n f(a) \end{pmatrix} \in \mathbb{R}^n$$

Gradient von f *in* a.

Bemerkung

Die i-te Ableitung von f wird gebildet indem Sie alle Variablen bis auf die i-te als Konstanten betrachten und nach der i-ten ableiten.

Beispiel

1. Betrachte die Funktion $f\colon \mathbb{R}^2 \to \mathbb{R}$, $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto x_1 \cdot x_2$ und sei $a = \begin{pmatrix} 4 \\ -1 \end{pmatrix} \in \mathbb{R}^2$. Dann gilt

$$g_{a,1}\colon \mathbb{R} \to \mathbb{R},\ x \mapsto x \cdot (-1) = -x,$$
$$g_{a,2}\colon \mathbb{R} \to \mathbb{R},\ x \mapsto 4 \cdot x,$$

sodass

$$g'_{a,1}(x) = -1,$$
$$g'_{a,2}(x) = 4$$

für alle $x \in \mathbb{R}$. Also

$$\partial_1 f(a) = g'_{a,1}(4) = -1,$$
$$\partial_2 f(a) = g'_{a,2}(-1) = 4.$$

Beispiel

2. Seien $f\colon \mathbb{R}^3 \to \mathbb{R}$, $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \mapsto x_1 + \exp(x_2 \cdot x_3)$ und $a = \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$.

   Dann gilt

$$\begin{aligned}
g_{a,1}\colon \mathbb{R} \to \mathbb{R},\ x &\mapsto x + \exp(1 \cdot 2) = x + \exp(2), \\
g_{a,2}\colon \mathbb{R} \to \mathbb{R},\ x &\mapsto 2 + \exp(x \cdot 2) = 2 + \exp(2x), \\
g_{a,3}\colon \mathbb{R} \to \mathbb{R},\ x &\mapsto 2 + \exp(1 \cdot x) = 2 + \exp(x),
\end{aligned}$$

   sodass

$$\begin{aligned}
g'_{a,1}(x) &= 1, \\
g'_{a,2}(x) &= 2\exp(2x), \\
g'_{a,3}(x) &= \exp(x)
\end{aligned}$$

   für alle $x \in \mathbb{R}$.

Beispiel

Also

$$\partial_1 f(a) = g'_{a,1}(2) = 1,$$
$$\partial_2 f(a) = g'_{a,2}(1) = 2\exp(2),$$
$$\partial_3 f(a) = g'_{a,3}(2) = \exp(2).$$

Beispiel

3 Seien $f : \mathbb{R}^2 \to \mathbb{R}$, $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto \exp(x_1)\sin(x_2)$ und $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$. Dann gilt

$$g_{a,1} : \mathbb{R} \to \mathbb{R},\ x \mapsto \exp(x)\sin(a_2),$$
$$g_{a,2} : \mathbb{R} \to \mathbb{R},\ x \mapsto \exp(a_1)\sin(x),$$

sodass

$$g'_{a,1}(x) = \exp(x)\sin(a_2),$$
$$g'_{a,2}(x) = \exp(a_1)\cos(x)$$

für alle $x \in \mathbb{R}$. Also

$$\partial_1 f(a) = g'_{a,1}(a_1) = \exp(a_1)\sin(a_2),$$
$$\partial_2 f(a) = g'_{a,2}(a_2) = \exp(a_1)\cos(a_2).$$

Definition

Seien $a \in \mathbb{R}^n$ und $f\colon \mathbb{R}^n \to \mathbb{R}$ eine in a partiell differenzierbare Funktion. Wir nennen a einen *kritischen Punkt von f*, falls

$$\operatorname{grad}(f)(a) = \begin{pmatrix} \partial_1 f(a) \\ \vdots \\ \partial_n f(a) \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

gilt , d. h. wenn $\partial_i f(a) = 0$ für alle $i = 1, \ldots, n$.

Beispiel

Seien $a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ und $f\colon \mathbb{R}^2 \to \mathbb{R}$, $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \mapsto x_1 \cdot x_2 - x_1$. Dann gilt

$$\begin{aligned} \partial_1 f(a) &= a_2 - 1, \\ \partial_2 f(a) &= a_1, \end{aligned}$$

sodass

$$\operatorname{grad}(f)(a) = \begin{pmatrix} a_2 - 1 \\ a_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Also ist $a = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ der einzige kritische Punkt von f.

Wir wollen nun gerne die kritischen Punkte der Funktion $r^2 \colon \mathbb{R}^2 \to \mathbb{R},\ x \mapsto |Ax - y|^2$ von Folie 5 ausrechnen.

Bemerkung

Wir betrachten nun r^2 statt r, da beide Funktionen die gleichen Minima besitzen und sich mit r^2 leichter rechnen lässt (die Wurzel beim Betrag fällt weg).

Hierfür brauchen wir das Konzept des Transponierens.

Definition

Seien $n, m \in \mathbb{N}$ und $A = (a_{i,j})_{i,j} \in \mathbb{R}^{m\times n}$. Dann heißt

$$A^T = (a_{j,i})_{i,j} = \begin{pmatrix} a_{1,1} & a_{2,1} & \cdots & a_{m,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{m,2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1,n} & a_{2,n} & \cdots & a_{n,m} \end{pmatrix} \in \mathbb{R}^{n\times m}$$

die *transponierte Matrix von A*.

Bemerkung

1. Die transponierte Matrix ergibt sich durch Spiegeln aller Einträge an der Diagonale. In anderen Worten: Die Spalten der Matrix A werden zu den Zeilen der Matrix A^{T}.
2. Es gilt $(A^T)^T = A$, d. h. wenn wir die Matrix zweimal transponieren erhalten wir die ursprüngliche Matrix zurück.

Beispiel

1. Sei $A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \in \mathbb{R}^{3\times 2}$. Dann gilt

$$A^T = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} \in \mathbb{R}^{2\times 3}.$$

2. Sei $A = \begin{pmatrix} 1 & 3 & 0 \\ 4 & -2 & -8 \\ 5 & 2 & 9 \end{pmatrix} \in \mathbb{R}^{3\times 3}$. Dann gilt

$$\begin{pmatrix} 1 & 4 & 5 \\ 3 & -2 & 2 \\ 0 & -8 & 9 \end{pmatrix} \in \mathbb{R}^{3\times 3}.$$

Beispiel

3. Sei $x = \begin{pmatrix} 3 \\ 6 \\ -2 \\ 1 \end{pmatrix} \in \mathbb{R}^4 = \mathbb{R}^{4\times 1}$. Dann gilt

$$x^T = (3, 6, -2, 1) \in \mathbb{R}^{1\times 4}.$$

Bemerkung

1. Wir sehen also, dass aus Spaltenvektoren durch Transponieren Zeilenvektoren entstehen und umgekehrt.
2. Es gilt

$$x^T x = |x|^2$$

für alle $x \in \mathbb{R}^n$.

Rechenregeln

Seien A, B Matrizen (mit jeweils passenden Größen) und $\alpha \in \mathbb{R}$. Dann gilt:

1. $(A+B)^T = A^T + B^T$,
2. $(\alpha A)^T = \alpha A^T$,
3. $(AB)^T = B^T A^T$,
4. $\det(A^T) = \det(A)$, *falls A quadratisch ist,*
5. *falls A invertierbar ist, so ist auch A^T invertierbar und es gilt*
$$(A^T)^{-1} = (A^{-1})^T.$$

Wir können also nun

$$
\begin{aligned}
r(x)^2 = |Ax - y|^2 = (Ax-y)^T(Ax-y) &= ((Ax)^T - y^T)(Ax - y) \\
&= (Ax)^T Ax - (Ax)^T y - y^T Ax + y^T y \\
&= x^T A^T Ax - ((Ax)^T y + y^T Ax) + |y|^2 \\
&= x^T A^T Ax - (x^T A^T y + (x^T A^T y)^T) + |y|^2 \\
&= x^T (A^T A) x - 2x^T (A^T y) + |y|^2
\end{aligned}
$$

für alle $x \in \mathbb{R}^2$ schreiben.
Der Gradient von r^2 berechnet sich zu

$$\mathrm{grad}(r^2)(x) = 2A^T Ax - 2A^T y = 2(A^T Ax - A^T y)$$

für alle $x \in \mathbb{R}^2$.

Damit folgt

$$\operatorname{grad}(r^2)(x^*) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad \Leftrightarrow \quad A^T A x^* = A^T y.$$

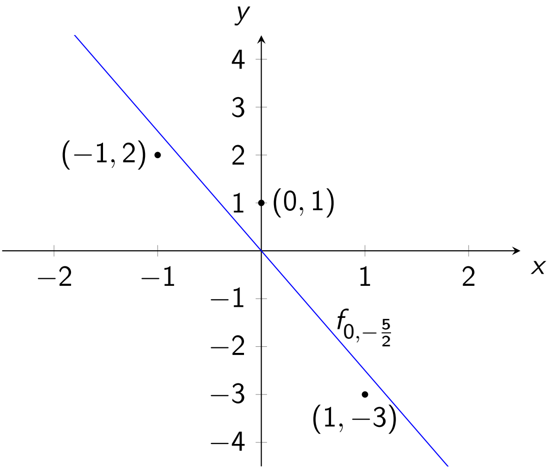
Da

$$A^T A = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix} \rightsquigarrow (A^T A)^{-1} = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{pmatrix},$$

$$A^T y = \begin{pmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} = \begin{pmatrix} 0 \\ -5 \end{pmatrix}$$

gelten, folgt

$$x^* = (A^T A)^{-1} A^T y = \begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 0 \\ -5 \end{pmatrix} = \begin{pmatrix} 0 \\ -\frac{5}{2} \end{pmatrix}.$$


y
4
3
2
1
−1
−2
−3
−4
−2
−1
1
2
x
(−1, 2)
(0, 1)
(1, −3)
$f_{0,-\frac{5}{2}}$

Im Allgemeinen sind wir also an folgender Problemstellung interessiert:
Zu gegebenen Messdaten $(t_1, y_1), \dots, (t_m, y_m) \in \mathbb{R}^2$ und einer Funktion $f_{x_1,\dots,x_n} : \mathbb{R} \to \mathbb{R}$, die von den Parametern $x_1, \dots, x_n \in \mathbb{R}$ abhängt, suchen wir die Parameter $x_1^*, \dots, x_n^* \in \mathbb{R}$, die das *Residuum*

$$r : \mathbb{R}^n \to \mathbb{R}, \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \mapsto \left| \begin{pmatrix} f_{x_1,\dots,x_n}(t_1) \\ \vdots \\ f_{x_1,\dots,x_n}(t_m) \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \right|$$

minimiert.

Im Folgenden betrachten wir nur Funktionen der Form

$$f_{x_1,\dots,x_n} : \mathbb{R} \to \mathbb{R}, \;\; t \mapsto x_1 h_1(t) + x_2 h_2(t) + \cdots + x_n h_n(t),$$

wobei $h_1, \dots, h_n : \mathbb{R} \to \mathbb{R}$ geeignete Funktionen sind. In diesem Fall sprechen wir von einer *linearen Ausgleichsrechnung*.

Damit gilt

$$\begin{pmatrix} f_{x_1,\dots,x_n}(t_1) \\ \vdots \\ f_{x_1,\dots,x_n}(t_m) \end{pmatrix} = \begin{pmatrix} x_1 h_1(t_1) + \dots + x_n h_n(t_1) \\ \vdots \\ x_1 h_1(t_m) + \dots + x_n h_n(t_m) \end{pmatrix}$$
$$= \underbrace{\begin{pmatrix} h_1(t_1) & \cdots & h_n(t_1) \\ \vdots & \ddots & \vdots \\ h_1(t_m) & \cdots & h_n(t_m) \end{pmatrix}}_{A \in \mathbb{R}^{m \times n}} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

sodass

$$r(x) = |Ax - y|$$

für alle $x \in \mathbb{R}^n$, wobei $y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$.

Bemerkung

Im eindimensionalen Fall lassen sich hinreichende Bedingungen formulieren, die sicherstellen, dass ein kritischer Punkt ein Tiefpunkt ($f''(x^*) > 0$) oder ein Hochpunkt ($f''(x^*) < 0$) ist. In höheren Dimensionen kann solche Bedingungen ebenfalls formulieren (dies führt auf die *Definitheit* der sog. *Hessematrix*) ; dies führt aber an dieser Stelle zu weit.
Für

$$r^2 \colon \mathbb{R}^n \to \mathbb{R}, \; x \mapsto |Ax - y|^2$$

lässt sich zeigen, dass, wenn $m \geq n$ (d. h. wir haben mehr Messdaten als Parameter) und $A^T A$ invertierbar ist, die Bedingung für einen Tiefpunkt immer erfüllt ist.

Satz

Falls $m \geq n$ und $A^T A$ ist invertierbar, so ist

$$x^* = (A^T A)^{-1} A^T y$$

die eindeutige Lösung des Minimierungsproblems.

Bemerkung

Gilt $m = n$ und

$$h_1(t) = 1,\ h_2(t) = t,\ h_3(t) = t^2, \ldots,\ h_n(t) = t^{n-1}$$

für alle $t \in \mathbb{R}$, d. h.

$$f_{x_1,\ldots,x_n}(t) = x_1 + x_2 t + x_3 t^2 + \ldots + x_n t^{n-1}$$

für alle $t \in \mathbb{R}$, so erhalten wir die Polynominterpolation zurück.

Beispiel

1. Betrachte die Messdaten

t	-1	0	1
y	e^2	e	e^{-3}

und die Funktion

$$f_{x_1,x_2} \colon \mathbb{R} \to \mathbb{R},\ t \mapsto x_1 \exp(x_2 t)$$

mit Parametern $x_1, x_2 \in \mathbb{R}$. Dies ist a priori keine lineare Ausgleichsrechnung, da x_2 in der Exponentialfunktion steht. Betrachte deswegen

$$\begin{aligned} g_{\tilde{x}_1,\tilde{x}_2} = \ln(f_{x_1,x_2}) \colon \mathbb{R} \to \mathbb{R},\ t \mapsto \ln(x_1 \exp(x_2 t)) &= \ln(x_1) + x_2 t \\ &= \tilde{x}_1 + \tilde{x}_2 t \end{aligned}$$

mit $\tilde{x}_1 = \ln(x_1)$ und $\tilde{x}_2 = x_2$.

Beispiel

Die neuen Messdaten, die durch das Logarithmieren entstehen, lauten

t	-1	0	1
$\tilde{y}$	2	1	-3

Dies ist aber unser ursprüngliches Beispiel, sodass wir

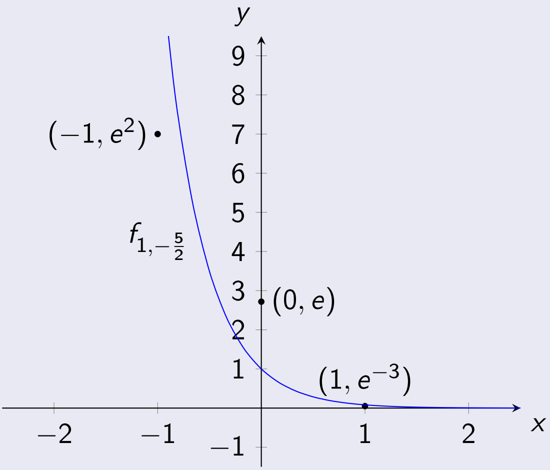
$$0 = \tilde{x}_1 = \ln(x_1) \quad \Leftrightarrow \quad x_1 = 1$$
$$-\frac{5}{2} = \tilde{x}_2 = x_2$$

erhalten, also

$$f_{1,-\frac{5}{2}}(t) = \exp\left(-\frac{5}{2}t\right)$$

für alle $t \in \mathbb{R}$.

Beispiel



Beispiel

2. Betrachte die Messdaten

t	0	$\frac{\pi}{2}$	π	2π
y	$\frac{1}{2}$	1	0	$\frac{1}{8}$

und die Funktion

$$f_{x_1,x_2} \colon \mathbb{R} \to \mathbb{R},\ \ t \mapsto x_1 \sin(x) + x_2 \cos(x)$$

mit Parametern $x_1, x_2 \in \mathbb{R}$. Also gilt

$$\begin{aligned} \frac{1}{2} &= f_{x_1,x_2}(0) &&= x_2, \\ 1 &= f_{x_1,x_2}\left(\frac{\pi}{2}\right) &&= x_1, \\ 0 &= f_{x_1,x_2}(\pi) &&= -x_2, \\ \frac{1}{8} &= f_{x_1,x_2}(2\pi) &&= x_2. \end{aligned}$$

Beispiel

Damit erhalten wir

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix} \quad \text{und} \quad y = \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ \frac{1}{8} \end{pmatrix},$$

sodass

$$A^T A = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 0 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix},$$

$$(A^T A)^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{pmatrix}, \; A^T y = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} \\ 1 \\ 0 \\ \frac{1}{8} \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{5}{8} \end{pmatrix}.$$

Beispiel

Also

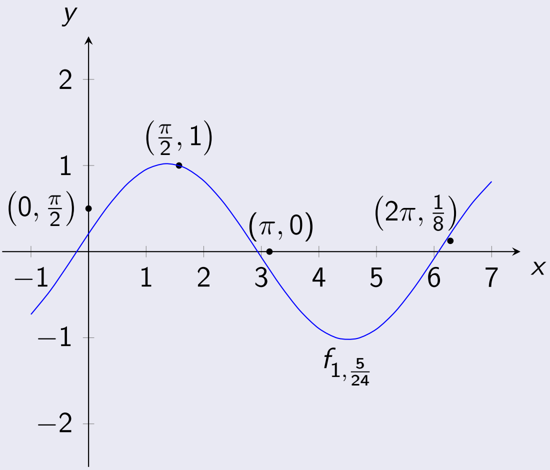
$$x^* = (A^T A)^{-1} A^T y = \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 \\ \frac{5}{8} \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{5}{24} \end{pmatrix},$$

d. h.

$$f_{1,\frac{5}{24}} : \mathbb{R} \to \mathbb{R}, \ t \mapsto \sin(x) + \frac{5}{24} \cos(t)$$

ist die gesuchte Funktion.

Beispiel


y
2
1
−1
−2
x
−1
1
2
3
4
5
6
7
$\left(\frac{\pi}{2}, 1\right)$
$\left(0, \frac{\pi}{2}\right)$
$(\pi, 0)$
$\left(2\pi, \frac{1}{8}\right)$
$f_{1,\frac{5}{24}}$

Es seien Messdaten $(t_1, y_1), \ldots, (t_m, y_m) \in \mathbb{R}^2$ gegeben. Wir haben versucht immer den gleichen Wert zu messen, sodass wir als Modellfunktion eine Konstante annehmen:

$$f_{x_1} : \mathbb{R} \to \mathbb{R}, t \mapsto x_1.$$

Wie sieht der optimale Parameter x_1^* aus? Dazu berechnen wir

$$A = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^{m \times 1} \quad \text{und} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m,$$

sodass

$$A^T A = m \quad \text{und} \quad A^T y = y_1 + \cdots + y_m.$$

Damit erhalten wir

$$x_1^* = (A^T A)^{-1} A^T y = \frac{1}{m}(y_1 + \cdots + y_m) =: \overline{y},$$

wobei $\overline{y}$ der sogenannte *Mittelwert* ist.

Gibt es einfache Darstellung oder eine Interpretation der optimalen Parameter, wenn wir als Modellfunktion

$$f_{x_1,x_2} \colon \mathbb{R} \to \mathbb{R}, \ t \mapsto x_1 + x_2 t$$

nehmen? Dazu berechnen wir

$$A = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_m \end{pmatrix} \in \mathbb{R}^{m\times 2} \quad \text{und} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} \in \mathbb{R}^m,$$

sodass

$$A^T A = \begin{pmatrix} m & t_1 + \cdots + t_m \\ t_1 + \cdots + t_m & t_1^2 + \cdots + t_m^2 \end{pmatrix}, \ A^T y = \begin{pmatrix} y_1 + \cdots + y_m \\ t_1 y_1 + \cdots + t_m y_m \end{pmatrix}.$$

Die Lösung ergibt sich mit $\overline{t} = \frac{1}{m}(t_1 + \cdots + t_m)$ und $\overline{y} = \frac{1}{m}(y_1 + \cdots + y_m)$ zu

$$x_1^* = \overline{y} - \frac{s_{t,y}}{s_{t,t}}\overline{t} \quad \text{und} \quad x_2^* = \frac{s_{t,y}}{s_{t,t}},$$

wobei

$$s_{t,y} = \frac{1}{m-1}\sum_{i=1}^{m}(t_i - \overline{t})(y_i - \overline{y}) = \frac{1}{m-1}\left(\sum_{i=1}^{m} t_i y_i - m\overline{t}\,\overline{y}\right)$$

die *(korrigierte) Stichprobenkovarianz* und

$$s_{t,t} = \frac{1}{m-1}\sum_{i=1}^{m}(t_i - \overline{t})^2 = \frac{1}{m-1}\left(\sum_{i=1}^{m} t_i^2 - m\overline{t}^2\right)$$

die *(korrigierte) Stichprobenvarianz* ist.

Bemerkung

1. Der Koeffizient $\frac{1}{m-1}$ wird nicht einheitlich verwendet. Es sollte also immer die Definition im jeweiligen Kontext nachschlagen.
2. Gilt $s_{t,y} = 0$, so heißen $t = (t_1, \ldots, t_n)^T$ und y *unkorreliert*.
3. Sind t und y unkorreliert, so gilt $x_2^* = 0$, sodass wir
$$x_1^* = \overline{y}$$
erhalten, d.h. wir befinden uns im Setting von Folie 2.

Beispiel

Betrachte die folgenden Messdaten

t	-1	0	1	2
y	2	1	-3	-4

und die Funktion

$$f_{x_1,x_2} : \mathbb{R} \to \mathbb{R}, \; t \mapsto x_1 + x_2 t$$

mit Parametern $x_1, x_2 \in \mathbb{R}$. Die Mittelwerte ergeben sich zu

$$\overline{t} = \frac{1}{4}(-1+0+1+2) = \frac{1}{2} \quad \text{und} \quad \overline{y} = \frac{1}{4}(2+1-3-4) = -1.$$

Beispiel

Für die Stichprobenkovarianz erhalten wir

$$s_{t,y} = \frac{1}{4-1}\left(((-1)\cdot 2 + 0\cdot 1 + 1\cdot(-1) + 2\cdot(-4)) - 4\frac{1}{2}\cdot(-1)\right)$$
$$= -\frac{11}{3}$$

und für die Stichprobenvarianz

$$s_{t,t} = \frac{1}{4-1}\left(((-1)^2 + 0^2 + 1^2 + 2^2) - 4\cdot\left(\frac{1}{2}\right)^2\right)$$
$$= \frac{5}{3}.$$

Also

$$x_1^* = -1 - \frac{\frac{-11}{3}}{\frac{5}{3}}\cdot\frac{1}{2} = \frac{1}{10} \quad \text{und} \quad x_2^* = \frac{\frac{-11}{3}}{\frac{5}{3}} = -\frac{11}{5}.$$

Beispiel