Sommersemester 2020

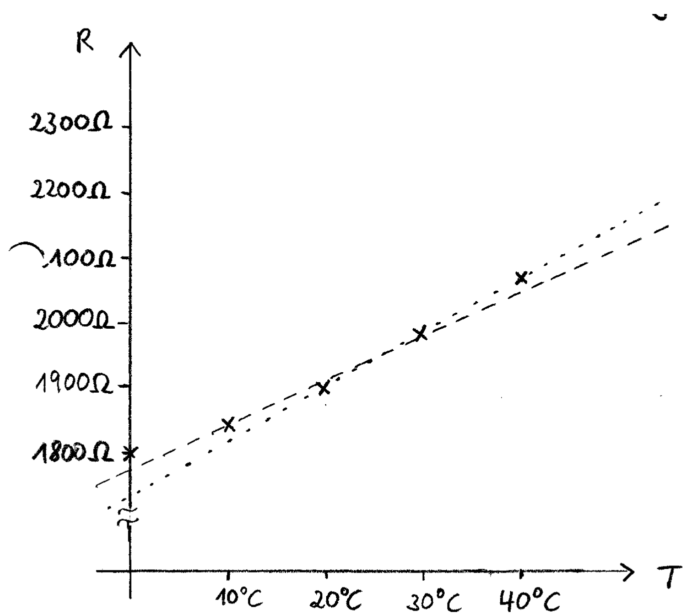
Dr. rer. nat. Frank Morherr

Ergänzungen zum Skript

Herleitung der Formeln zur linearen Regression

Zunächst mittels partiellen Ableitungen Hierzu betrachten wir folgendes Beispiel:

Bei einem Temperatursensor soll die Temperaturabhängigkeit des (ohmschen) Widerstandes über einen eingeschränkten Temperaturbereich, (z.B. $0°C$ - $40°C$) durch eine möglichst einfache, also etwa eine lineare Funktion, ausgedrückt werden. Dazu werden eine Reihe von Messungen gemacht:



Die Messwerte sind durch Kreuze angedeutet. Wie kann man nun eine lineare Funktion, deren Graph die gestrichelte Gerade ist, bestimmen, die möglichst gut zu den Messwerten passt? Bei dieser Linearisierung sind zunächst einmal zwei Effekte zu betrachten:

1. Bei einem realen Sensor wird man meist keinen linearen Zusammenhang zwischen T und R haben, nur über einen kleinen Temperaturbereich wird er näherungsweise linear sein. Deswegen wird die Gerade prinzipiell nicht durch alle Messwerte gehen können.

2. Die Messungen sind natürlich auch immer mit Messfehlern behaftet.

Abstrakt formuliert lautet das Problem so:

Gegeben sind Wertepaare (x_1, y_1) , (x_2, y_2) ,..., (x_n, y_n) , gesucht sind Zahlen a und b , so dass für $f(x) = ax + b$ gilt

$|y_k - f(x_k)|$ ist für **alle** "gleichmäßig" klein.

Das *könnte* man z.B. dadurch auszudrücken, dass man verlangt, dass

$$\sum_{k=1}^{n} |y_k - f(x_k)|$$

möglichst klein wird. Dies empfiehlt sich aber gerade nicht, weil im Beispiel oben auch die gepunktete Gerade möglich wäre, und die ist offensichtlich nicht so gut. Dieses Problem vermeidet man, indem man große Abweichungen $|y_k - f(x_k)|$ "bestraft", also stärker gewichtet. Diese geht z. B. dadurch, dass man verlangt, dass

$$\sum_{k=1}^{n} |y_k - f(x_k)|^2$$

möglichst klein werden soll. Ein weiterer Vorteil ist auch, wie wir unten sehen werden, dass $|y_k - f(x_k)|^2 = (y_k - f(x_k))^2$ in Null differenzierbar ist, während $|y_k - f(x_k)|$ dies nicht ist.

Dies ist die auf Carl Friedrich Gauß zurückgehende *Methode der kleinsten Quadrate*. Sie funktioniert nicht nur bei linearen Zusammenhängen, wie hier mit $f(x) = ax + b$, sondern auch für nichtlineare Zusammenhänge, wie zum Beispiel für

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \text{ (polynomiell } n\text{-ten Grades).}$$

Wir betrachten Sie hier aber nur für lineare Zusammenhänge. Gesucht sind also a, b so, dass

$$F(a,b) = \sum_{k=1}^{n} |y_k - f(x_k)|^2 = \sum_{k=1}^{n} (y_k - (ax_k + b))^2$$

möglichst klein wird, mit anderen Worten suchen wir ein globales Minimum der Funktion $F(a,b)$ von zwei Variablen, welches aber aufgrund der Form von F als lokales Minimum auftritt. Die notwendige Bedingung für ein lokales Minimum in (a,b) ist

$$\operatorname{grad} F(a,b) = \overrightarrow{0} \Longleftrightarrow \begin{pmatrix} \frac{\partial F}{\partial a}(a,b) \\ \frac{\partial F}{\partial b}(a,b) \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} .$$

Nun gilt

$$\begin{aligned} \frac{\partial F}{\partial a}(a,b) &= \frac{\partial}{\partial a}\left(\sum_{k=1}^{n} (y_k - (ax_k + b))^2\right) = \sum_{k=1}^{n} \frac{\partial}{\partial a} (y_k - (ax_k + b))^2 = \\ &= \sum_{k=1}^{n} 2(y_k - (ax_k + b))(-x_k) = -2\sum_{k=1}^{n} (y_k - (ax_k + b))\, x_k , \\ \frac{\partial F}{\partial b}(a,b) &= \frac{\partial}{\partial b}\left(\sum_{k=1}^{n} (y_k - (ax_k + b))^2\right) = \sum_{k=1}^{n} \frac{\partial}{\partial b} (y_k - (ax_k + b))^2 = \\ &= \sum_{k=1}^{n} 2(y_k - (ax_k + b))(-1) = -2\sum_{k=1}^{n} (y_k - (ax_k + b)) . \end{aligned}$$

Diese Ableitung müssen beide gleichzeitig verschwinden. Nun gilt

$$\frac{\partial F}{\partial a}(a,b) = 0 \Longleftrightarrow -2\sum_{k=1}^{n} (y_k - (ax_k + b))\, x_k = 0$$

$$\Longleftrightarrow \sum_{k=1}^{n} (y_k - (ax_k + b))\, x_k = 0 \Longleftrightarrow \sum_{k=1}^{n} x_k y_k - a\sum_{k=1}^{n} x_k^2 - b\sum_{k=1}^{n} x_k = 0 ,$$

sowie

$$\frac{\partial F}{\partial b}(a,b) = 0 \Longleftrightarrow -2\sum_{k=1}^{n} (y_k - (ax_k + b)) = 0$$

$$\Longleftrightarrow \sum_{k=1}^{n} (y_k - (ax_k + b)) = 0 \Longleftrightarrow \sum_{k=1}^{n} y_k - a\sum_{k=1}^{n} x_k - bn = 0 .$$

(Man beachte: $\sum_{k=1}^{n} 1 = n$) Insgesamt ergibt sich somit das Gleichungssystem

$$\begin{pmatrix} \sum_{k=1}^{n} x_k^2 & \sum_{k=1}^{n} x_k \\ \sum_{k=1}^{n} x_k & n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{n} x_k y_k \\ \sum_{k=1}^{n} y_k \end{pmatrix} . \qquad (**)$$

Mit der Cramerschen Regel ergibt sich die Lösung zu:

$$a = \frac{\left|\begin{pmatrix} \sum_{k=1}^n x_k y_k & \sum_{k=1}^n x_k \\ \sum_{k=1}^n y_k & n \end{pmatrix}\right|}{\begin{vmatrix} \sum_{k=1}^n x_k^2 & \sum_{k=1}^n x_k \\ \sum_{k=1}^n x_k & n \end{vmatrix}} = \frac{n \sum_{k=1}^n x_k y_k - \left(\sum_{k=1}^n x_k\right)\left(\sum_{k=1}^n y_k\right)}{n \sum_{k=1}^n x_k^2 - \left(\sum_{k=1}^n x_k\right)^2} \ ,$$

$$b = \frac{\begin{vmatrix} \sum_{k=1}^n x_k^2 & \sum_{k=1}^n x_k y_k \\ \sum_{k=1}^n x_k & \sum_{k=1}^n y_k \end{vmatrix}}{\begin{vmatrix} \sum_{k=1}^n x_k^2 & \sum_{k=1}^n x_k \\ \sum_{k=1}^n x_k & n \end{vmatrix}} = \frac{\left(\sum_{k=1}^n x_k^2\right)\left(\sum_{k=1}^n y_k\right) - \left(\sum_{k=1}^n x_k\right)\left(\sum_{k=1}^n x_k y_k\right)}{n \sum_{k=1}^n x_k^2 - \left(\sum_{k=1}^n x_k\right)^2} \ .$$

Da dies sehr unübersichtlich aussieht, wollen wir die üblichen Abkürzungen einführen:

$$\begin{aligned}
\overline{x} &= \frac{1}{n}\sum_{k=1}^n x_k \text{ (Mittelwert, auch Erwartungswert der } x_k \ , \ k = 1, \dots, n \text{)} \\
\overline{y} &= \frac{1}{n}\sum_{k=1}^n y_k \text{ (Mittelwert, auch Erwartungswert der } y_k \ , \ k = 1, \dots, n \text{)} \\
\overline{x^2} &= \frac{1}{n}\sum_{k=1}^n x_k^2 \ , \\
\overline{xy} &= \frac{1}{n}\sum_{k=1}^n x_k y_k \ .
\end{aligned}$$

Damit ergibt sich nach kürzen durch n^2

$$a = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2} \ , \ b = \frac{\overline{x^2}\cdot\overline{y} - \overline{x}\cdot\overline{xy}}{\overline{x^2} - \left(\overline{x}\right)^2}. \tag{$*$}$$

Tip: Hat man a berechnet, bekommt man b leicht über

$$b = \overline{y} - a\overline{x},$$

denn

$$\begin{aligned}
\overline{y} - a\overline{x} &= \overline{y} - \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2}\cdot\overline{x} = \frac{\left(\overline{x^2} - \left(\overline{x}\right)^2\right)\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2} - \frac{\left(\overline{xy} - \overline{x}\cdot\overline{y}\right)\cdot\overline{x}}{\overline{x^2} - \left(\overline{x}\right)^2} = \\
&= \frac{\overline{x^2}\cdot\overline{y} - \left(\overline{x}\right)^2\overline{y} - \overline{x}\cdot\overline{xy} + \left(\overline{x}\right)^2\overline{y}}{\overline{x^2} - \left(\overline{x}\right)^2} = \frac{\overline{x^2}\cdot\overline{y} - \overline{x}\cdot\overline{xy}}{\overline{x^2} - \left(\overline{x}\right)^2} = b.
\end{aligned}$$

Genau genommen muss man noch feststellen, dass

$$\begin{vmatrix} \sum_{k=1}^n x_k^2 & \sum_{k=1}^n x_k \\ \sum_{k=1}^n x_k & n \end{vmatrix} = n\sum_{k=1}^n x_k^2 - \left(\sum_{k=1}^n x_k\right)^2$$

nicht Null ist, denn wir teilen ja dadurch, und andernfalls haben wir auch keine eindeutige Lösbarkeit des Gleichungssystems. Dies kann man in der Tat nachrechnen: Es gilt nämlich, da mindestens zwei der x_k ungleich sein müssen, sonst gibt es keine Ausgleichsgerade:

$$0 \underset{\substack{\text{Da mindestens zwei der} \\ x_k \text{ ungleich sein müssen}}}{<} \frac{1}{2}\sum_{k=1}^n\left(\sum_{m=1}^n\left(x_k - x_m\right)^2\right) \underset{\text{Binom. Formel}}{=} \frac{1}{2}\sum_{k=1}^n\left(\sum_{m=1}^n\left(x_k^2 - 2x_k x_m + x_m^2\right)\right) =$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{k=1}^{n}\left(x_k^2\sum_{m=1}^{n}1 - 2x_k\sum_{m=1}^{n}x_m + \sum_{m=1}^{n}x_m^2\right) = \frac{1}{2}\sum_{k=1}^{n}\left(x_k^2 n - 2x_k n\overline{x} + n\overline{x^2}\right) = \\
&= \frac{1}{2}\left(n\sum_{k=1}^{n}x_k^2 - 2n\overline{x}\sum_{k=1}^{n}x_k + n\overline{x^2}\sum_{k=1}^{n}1\right) = \frac{1}{2}\left(n^2\overline{x^2} - 2n^2\left(\overline{x}\right)^2 + n^2\overline{x^2}\right) = \\
&= \frac{1}{2}\left(2n^2\overline{x^2} - 2n^2\left(\overline{x}\right)^2\right) = n^2\overline{x^2} - n^2\left(\overline{x}\right)^2 = n\sum_{k=1}^{n}x_k^2 - \left(\sum_{k=1}^{n}x_k\right)^2 ,
\end{aligned}
$$

also ist der Nenner ungleich Null.

Wir rechnen jetzt noch die Formeln in die Darstellung $(*)$ aus der Vorlesung um mit den dort verwendeten Abkürzungen. In der Vorlesung haben wir verwendet

$$
\begin{aligned}
SS_{xy} &= \sum_{k=1}^{n}x_k y_k - \frac{1}{n}\sum_{k=1}^{n}x_k\sum_{k=1}^{n}y_k = n\frac{1}{n}\sum_{k=1}^{n}x_k y_k - n\left(\frac{1}{n}\sum_{k=1}^{n}x_k\right)\left(\frac{1}{n}\sum_{k=1}^{n}y_k\right) \\
&= n\overline{xy} - n\overline{x}\cdot\overline{y} = n\left(\overline{xy} - \overline{x}\cdot\overline{y}\right),
\end{aligned}
$$

sowie

$$
\begin{aligned}
SS_{xx} &= \sum_{k=1}^{n}x_k^2 - \frac{1}{n}\sum_{k=1}^{n}x_k\sum_{k=1}^{n}x_k = n\frac{1}{n}\sum_{k=1}^{n}x_k^2 - n\left(\frac{1}{n}\sum_{k=1}^{n}x_k\right)\left(\frac{1}{n}\sum_{k=1}^{n}y_k\right) \\
&= n\left(\overline{x^2} - \overline{x}\cdot\overline{x}\right) = n\left(\overline{x^2} - \left(\overline{x}\right)^2\right).
\end{aligned}
$$

Damit folgt mit $(*)$und indem man n kürzt

$$a = \frac{SS_{xy}}{SS_{xx}},$$

sowie

$$b = \overline{y} - a\overline{x} = \overline{y} - \frac{SS_{xy}}{SS_{xx}}\overline{x} = \frac{SS_{xx}\overline{y} - SS_{xy}\overline{x}}{SS_{xx}}.$$

Die sogenannte *Ausgleichsgerade* $y = ax + b$ ist also eindeutig bestimmt, und nach obigen Formeln leicht berechenbar. Besonders praktisch ist, dass man bei nachträglicher Hinzunahme eines Wertepaares (x_{n+1}, y_{n+1}) nicht die ganze Rechnung komplett neu durchführen muss, sondern die Summen einfach um den zusatzlichen Term ergänzen kann.

Example 1 *Es liege folgende Tabelle von Messwerten vor*

x_k	0	10	20	30	40
y_k	180	184	190	198	207

.

Hier ergibt sich

$$\sum_{k=1}^{n}x_k = 100\ ,\ \sum_{k=1}^{n}x_k^2 = 3000\ ,\ \sum_{k=1}^{n}y_k = 959\ \textit{und}\ \sum_{k=1}^{n}x_k y_k = 19860\ .$$

Wegen $n = 5$ *ergibt sich also*

$$a = \frac{5\cdot 19860 - 100\cdot 959}{5\cdot 3000 - 100^2} = 0,68$$

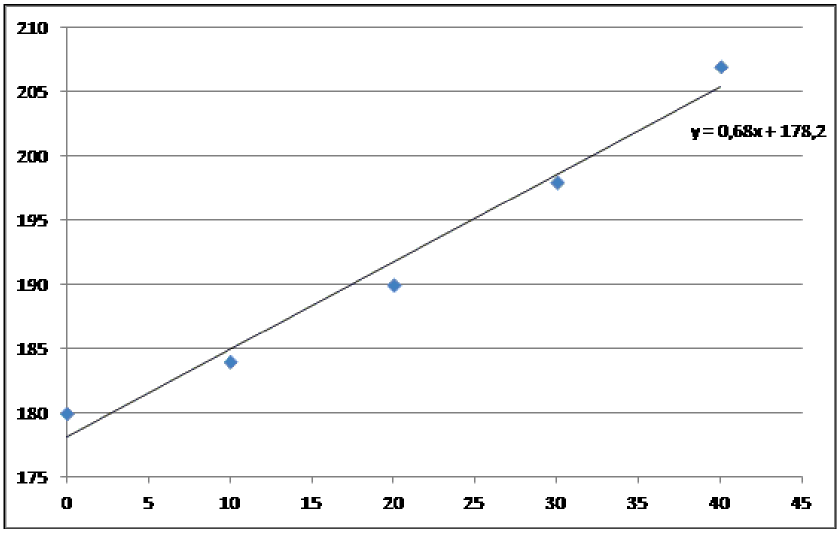
und

$$b = \frac{3000\cdot 959 - 100\cdot 19860}{5\cdot 3000 - 100^2} = 178,2\ .$$

Die Ausgleichsgerade ist demnach

$$y = 0,68x + 178,2\ .$$

Durchführen der Rechnung mit EXCEL ergibt folgendes Bild mit der von EXCEL berechneten Ausgleichsgerade:



Berechnung über die Normalengleichung Das was wir oben über partielle Ableitungen gemacht haben, kann man auch ganz formal herleiten. Dazu stellen wir das folgende Gleichungssystem auf. Wieder beschränken wir uns nur auf lineare Funktionen. Die Verallgemeinerung ist aber offensichtlich. Setzt man die gegebenen Punkte in die lineare Funktion ein, ergeben sich die Gleichungen

$$\begin{aligned} ax_1 + b &= y_1 \\ ax_2 + b &= y_2 \end{aligned}$$

$$\vdots$$

$$ax_n + b = y_n$$

Dieskann man in Matrixform-Vektor-Darstellung schreiben. Dazu betrachten wir die Matrix und die Vektoren (wir lassen die Pfeile weg, der Übersichtlichkeit halber)

$$A = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix}, \; x = \begin{pmatrix} a \\ b \end{pmatrix} \text{ und } b = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_4 \end{pmatrix}.$$

Dann ist

$$Ax = b.$$

das überbestimmte Gleichungssystem (welches in der Regel keine Lösung hat), und deren Residuen $Ax-b=\varepsilon$ zu minimieren sind. Man betrachtet jetzt das Betragsquadrat$\|\cdot\|^2$, auch euklidische Länge des Vektors, welche über das Skalarprodukt $x \cdot y = x^\intercal y$ mittels $\|x\|^2 = x \cdot x = x^\intercal x$ definiert ist. Diese gilt es für $Ax - b$ zu minimieren. Mit den Rechnenregeln des Skalarprodukts, bzw. den Matrizenrechenregeln, rechnen wir

$$\begin{aligned} \|A(x+h) - b\|^2 &= (A(x+h) - b)^\intercal (A(x+h) - b) = (Ax - b + Ah)^\intercal (Ax - b + Ah) \\ &= ((Ax-b)^\intercal + (Ah)^\intercal)(Ax - b + Ah) \\ &= (Ax-b)^\intercal (Ax-b) + (Ax-b)^\intercal Ah + (Ah)^\intercal (Ax-b) + (Ah)^\intercal Ah \\ &= \|Ax-b\|^2 + (Ah)^\intercal (Ax-b) + (Ah)^\intercal (Ax-b) + \|Ah\|^2 \\ &= \|Ax-b\|^2 + \|Ah\|^2 + 2(Ah)^\intercal (Ax-b) \\ &= \|Ax-b\|^2 + \|Ah\|^2 + 2h^\intercal A^\intercal (Ax-b) \\ &= \|Ax-b\|^2 + \|Ah\|^2 + 2h^\intercal (A^\intercal Ax - A^\intercal b) \\ &= \|Ax-b\|^2 + \|Ah\|^2 + 2(A^\intercal Ax - A^\intercal b)^\intercal h. \end{aligned}$$

Für die Funktion $f(x) = \|Ax - b\|^2$ bedeutet das

$$f(x+h) = f(x) + 2\left(A^\intercal Ax - A^\intercal b\right)^\intercal h + \|Ah\|^2$$

gilt. Dies ist die Weierstraß-Form der Differenzierbarkeit. f ist minimal, wenn die Ableitung von f verschwindet. Aus obigem sieht man, dass die Ableitung

$$f'(x) = 2\left(A^\intercal Ax - A^\intercal b\right)^\intercal$$

entspricht. Nun gilt

$$f'(x) = 0 \Leftrightarrow 2\left(A^\intercal Ax - A^\intercal b\right)^\intercal = 0 \Leftrightarrow A^\intercal Ax = A^\intercal b.$$

Dies ist gerade die Normalengleichung. Statt das ursprüngliche Gleichungssystem lösen wir also die Normalgleichung

$$A^\intercal Ax = A^\intercal b,$$

von der man zeigen kann, dass sie immer lösbar ist und deren Lösung das Problem löst.

Wir zeigen nun, dass sich damit dieselben Formeln wie mit der ersten Methode ergeben: Es gilt

$$A^\intercal A = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \vdots & \vdots \\ x_n & 1 \end{pmatrix} = \begin{pmatrix} \sum\limits_{k=1}^n x_k^2 & \sum\limits_{k=1}^n x_k \\ \sum\limits_{k=1}^n x_k & n \end{pmatrix},$$

$$A^\intercal b = \begin{pmatrix} x_1 & x_2 & \cdots & x_n \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_4 \end{pmatrix} = \begin{pmatrix} \sum\limits_{k=1}^n x_k y_k \\ \sum\limits_{k=1}^n y_k \end{pmatrix}$$

und somit mit $x = \begin{pmatrix} a \\ b \end{pmatrix}$,

$$A^\intercal Ax = A^\intercal b \Leftrightarrow \begin{pmatrix} \sum\limits_{k=1}^n x_k^2 & \sum\limits_{k=1}^n x_k \\ \sum\limits_{k=1}^n x_k & n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \sum\limits_{k=1}^n x_k y_k \\ \sum\limits_{k=1}^n y_k \end{pmatrix}.$$

Dies ist genau das Gleichungssystem aus $(**)$.

Verallgemeinerung auf nichtlineare Regression Die Verallgemeinerung auf andere Regressionsfunktionen ist offensichtlich, siehe hierzu auch die anderen Folien im Material über Normalengleichung. Als Beispiel zeigen wir hier die Veallgemeinerung auf quadratische Funktionen. Hier sind die Punkte mit der quadratischen Funktion

$$f(x) = ax^2 + bx + c$$

anzunähern. Wir lösen das mit der Normalgleichung: Dazu betrachten wir die Matrix und die Vektoren

$$A = \begin{pmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_4^2 & x_4 & 1 \end{pmatrix}, \; x = \begin{pmatrix} a \\ b \\ c \end{pmatrix} \text{ und } b = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_4 \end{pmatrix}.$$

Dann ist das überbestimmte Gleichungssystem (welches in der Regel keine Lösung hat), und deren Residuen zu minimieren sind, wieder

$$Ax = b.$$

Statt dieses Gleichungssystems lösen wir wieder die Normalgleichung

$$A^\intercal A x = A^\intercal b.$$

Diese bekommt jetzt hier die Gestalt

$$\begin{pmatrix} x_1^2 & x_2^2 & \cdots & x_4^2 \\ x_1 & x_2 & \cdots & x_4 \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \vdots & \vdots & \vdots \\ x_4^2 & x_4 & 1 \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} x_1^2 & x_2^2 & \cdots & x_4^2 \\ x_1 & x_2 & \cdots & x_4 \\ 1 & 1 & \cdots & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_4 \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} \sum_{k=1}^n x_k^4 & \sum_{k=1}^n x_k^3 & \sum_{k=1}^n x_k^2 \\ \sum_{k=1}^n x_k^3 & \sum_{k=1}^n x_k^2 & \sum_{k=1}^n x_k \\ \sum_{k=1}^n x_k^2 & \sum_{k=1}^n x_k & n \end{pmatrix} \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^n x_k^2 y_k \\ \sum_{k=1}^n x_k y_k \\ \sum_{k=1}^n y_k \end{pmatrix}.$$

Mittels des Gaußalgorithmus oder der Cramerschen Regel können wir jetzt hier wieder a, b und c bestimmen.