

Elementare lineare Regression

Andreas Stahel, Pierre–André Chevalier

Mai 2012

Inhaltsverzeichnis

1	Einführung	2
2	Bestimmen der Koeffizienten der Geraden	2
2.1	Zahlenbeispiel	2
2.2	Allgemeiner Fall	4
3	Statistische Grundlagen	5
3.1	Begriffe der Wahrscheinlichkeitsrechnung	5
3.1.1	Normalverteilung	5
3.1.2	Schätzen von Mittelwert und Standardabweichung aus einer Messreihe	7
3.2	Maximum–Likelihood Methode	8
3.3	Schätzen der Varianz von a und m	9
3.3.1	Bestimmen der Standardabweichung σ der Messwerte y_i	10
3.3.2	Varianz einer Summe von Zufallsgrößen, Varianzen von a und m	10
4	Ergänzungen	11
4.1	Allgemeine lineare Regression, Matrizennotation	11
4.2	Example : Intensity of light of an LED depending on the angle of observation	13
4.2.1	Description of the problem	13
4.2.2	A first, naive solution and its problems	13
4.2.3	How to eliminate the problem	15
4.2.4	Using the command <code>LinearRegression()</code>	17
4.3	Verifikation der Korrektheit der Gleichungen	18
4.3.1	Aufstellen der Gleichungen	18
4.3.2	Matrix– und Vektornotation	18
4.3.3	Effiziente Codierung, QR –Zerlegung und <i>Octave</i>	19
4.4	Messungen mit verschiedener Genauigkeit	21
4.4.1	Regressionsgerade	21
4.4.2	Allgemeine Regression	22
4.5	Exponentialkurven anpassen	22
4.6	Fehler in x und y –Werten	24
4.7	<i>Octave</i> /MATLAB–Befehl <code>LinearRegression()</code>	24

1 Einführung

Gegeben sind einige Punkte (x_i, y_i) , $i = 1 \dots n$ in der Ebene und gesucht ist eine Gerade $y = a + m \cdot x$ die „möglichst gut“ durch diese Punkte geht. Die Werte der Parameter a und m sind geeignet zu bestimmen. Sind nur zwei Punkte gegeben, so ist dies leicht möglich. Bei mehreren Punkten ist das Problem aber etwas schwieriger, da die Punkte üblicherweise **nicht exakt** auf einer Geraden liegen. Sind die Punkte weit gestreut, so ist es sehr schwierig die optimale Gerade graphisch zu bestimmen.

x =	0.0	1.0	2.0	3.5	4.0
y =	-0.5	1.0	2.4	2.0	3.1

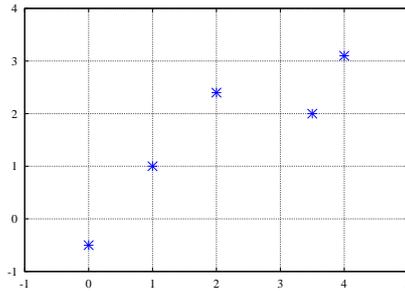


Abbildung 1: Graph einiger Punkte in der Ebene

Ziel ist es eine Gerade der Form

$$f(x) = a + m \cdot x$$

durch diese Punkte zu passen. Eine mögliche “beste“ Variante ist beschrieben durch die folgende Bedingung:

Finde a und m sodass $g := \sum_{i=1}^n (f(x_i) - y_i)^2$ minimal wird.

Die Summe der Quadrate der vertikalen Abstände von der Geraden soll möglichst klein gemacht werden. Deshalb wird dieses Verfahren auch **Methode der kleinsten Quadrate**¹ genannt. Meistens wird der Begriff **lineare Regression** verwendet.

2 Bestimmen der Koeffizienten der Geraden

In diesem Abschnitt wird das Rechenverfahren entwickelt um die beiden gesuchten Parameter a und m zu bestimmen. Zuerst werden die Rechnungen am einleitenden Zahlenbeispiel ausgeführt um anschließend die allgemeinen Formeln mit dem selben Verfahren herzuleiten. Vergleichbare Rechnungen finden Sie in [Pres92, p. 662].

2.1 Zahlenbeispiel

Für das obige Beispiel führt das auf die Funktion

¹In der Literatur wird oft die Notation $\chi^2 = g$ verwendet und man minimiert Chi-Quadrat.

$$\begin{aligned}
g(a, m) &= \sum_{i=1}^5 (a + m \cdot x_i - y_i)^2 \\
&= (a + 0.5)^2 + (a + m - 1)^2 + (a + m \cdot 2 - 2.4)^2 + (a + m \cdot 3.5 - 2.0)^2 + (a + m \cdot 4 - 3.1)^2
\end{aligned}$$

welche bezüglich der zwei Parameter a und m zu minimieren ist. .

Lösung : Wir müssen das Minimum der Funktion $g(a, m)$ suchen. Es ist klar, dass sehr grosse (positive) oder sehr kleine (negative) Werte von a und m zu grossen Funktionswerten $g(a, m)$ führen. Ausserdem ist die Funktion offensichtlich differenzierbar. Bei einem Maximum müssen die Ableitungen bezüglich der Variablen a und m Null sein. Folglich erhalten wir die zwei notwendigen Bedingungen

$$\frac{\partial g}{\partial a} = 0 \quad \text{und} \quad \frac{\partial g}{\partial m} = 0$$

Mit den gegebenen Beispielwerten erhalten wir die zwei folgenden Gleichungen.

$$\begin{aligned}
\frac{\partial g}{\partial a} &= 2 ((a + 0.5) + (a + m - 1) + (a + m \cdot 2 - 2.4) + (a + m \cdot 3.5 - 2.0) + (a + m \cdot 4 - 3.1)) \\
&= 2 a \cdot 5 + 2 m (0 + 1 + 2 + 3.5 + 4) - 2 (-0.5 + 1.0 + 2.4 + 2.0 + 3.1) \\
&= 2 (a \cdot 5 + m \cdot 10.5 - 8) = 0
\end{aligned}$$

$$\begin{aligned}
\frac{\partial g}{\partial m} &= 2 (1 (a + m - 1) + 2 (a + m \cdot 2 - 2.4) + 3.5 (a + m \cdot 3.5 - 2.0) + 4 (a + m \cdot 4 - 3.1)) \\
&\quad \text{verwende } 0 + 1 + 2 + 3.5 + 4 = 10.5 \\
&\quad \text{und } 0^2 + 1^2 + 2^2 + 3.5^2 + 4^2 = 33.25 \\
&= 2 a \cdot 10.5 + 2 m \cdot 33.25 - 2 (-0.5 \cdot 0 + 1.0 \cdot 1 + 2.4 \cdot 2.0 + 2.0 \cdot 3.5 + 3.1 \cdot 4) \\
&= 2 (a \cdot 10.5 + m \cdot 33.25 - 25.2) = 0
\end{aligned}$$

für die zwei Unbekannten a und m . Dies ist ein lineares System von Gleichungen und kann geschrieben werden als

$$\begin{aligned}
a \cdot 5 + m \cdot 10.5 &= 8 \\
a \cdot 10.5 + m \cdot 33.25 &= 25.2
\end{aligned}$$

Die Lösungen sind somit

$$\begin{pmatrix} a \\ m \end{pmatrix} = \begin{bmatrix} 5 & 10.5 \\ 10.5 & 33.25 \end{bmatrix}^{-1} \begin{pmatrix} 8 \\ 25.2 \end{pmatrix} = \begin{pmatrix} 0.025 \\ 0.750 \end{pmatrix}$$

Somit ist die Gerade gegeben durch

$$f(x) = a + m x = 0.025 + 0.75 x$$

Das Resultat der obigen Rechnungen finden Sie in [Abbildung 2](#) in graphischer Form. □

Das obige Problem kann auch mittels weniger Zeilen MATLAB/Octave-Code gelöst werden.

```

Octave
x = [0; 1; 2; 3.5; 4]; y = [-0.5; 1; 2.4; 2.0; 3.1];
F = [ones(length(x),1) x];
[p,y_v,r,p_v] = LinearRegression(F,y)

```

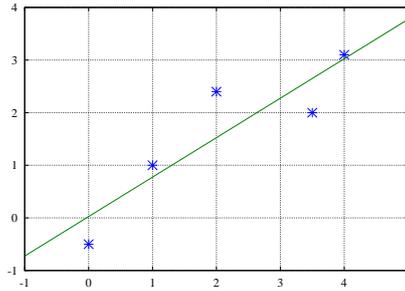


Abbildung 2: Datenpunkte mit Regressionsgerade

2.2 Allgemeiner Fall

Für das obige Beispiel führt das auf die Funktion

$$g(a, m) = \sum_{i=1}^n (a + m \cdot x_i - y_i)^2$$

welche bezüglich der zwei Parameter a und m zu minimieren ist.

Lösung : Wir müssen das Minimum der Funktion $g(a, m) = \chi^2(a, m)$ suchen. Dazu sind die Nullstellen der beiden Ableitungen zu finden.

$$\frac{\partial g}{\partial a} = 0 \quad \text{und} \quad \frac{\partial g}{\partial m} = 0$$

Das führt auf die zwei Gleichungen

$$\begin{aligned} \frac{\partial g}{\partial a} &= 2 \sum_{i=1}^n (a + m x_i - y_i) = 0 \\ \frac{\partial g}{\partial m} &= 2 \sum_{i=1}^n (a + m x_i - y_i) x_i = 0 \end{aligned}$$

für die zwei Unbekannten a und m . Dies ist ein lineares System von Gleichungen und kann geschrieben werden als

$$\begin{aligned} a \cdot n + m \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i \\ a \sum_{i=1}^n x_i + m \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n (y_i \cdot x_i) \end{aligned}$$

Die Koeffizienten des Gleichungssystems sind als Summen von x und y -Werten gegeben. Man kann die folgenden Notationen verwenden:

$$\begin{aligned} S_x &= \sum_{i=1}^n x_i & S_y &= \sum_{i=1}^n y_i \\ S_{xx} &= \sum_{i=1}^n x_i^2 & S_{xy} &= \sum_{i=1}^n x_i y_i \end{aligned}$$

Die vier Summen lassen sich leicht berechnen, falls die Datenpunkte (x_i, y_i) der Reihe nach eingegeben werden. Die meisten Taschenrechner haben diese Befehle fest eingebaut. Dann sind a und m bestimmt durch

$$\begin{bmatrix} n & S_x \\ S_x & S_{xx} \end{bmatrix} \begin{pmatrix} a \\ m \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}$$

Diese 2×2 -Matrix kann invertiert werden und wir erhalten

$$\begin{bmatrix} n & S_x \\ S_x & S_{xx} \end{bmatrix}^{-1} = \frac{1}{n S_{xx} - S_x^2} \begin{bmatrix} S_{xx} & -S_x \\ -S_x & n \end{bmatrix}$$

und

$$\begin{pmatrix} a \\ m \end{pmatrix} = \begin{bmatrix} n & S_x \\ S_x & S_{xx} \end{bmatrix}^{-1} \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}$$

Die Lösungen a und m sind somit gegeben durch

$\begin{aligned} \Delta &= n \cdot S_{xx} - S_x^2 \\ a &= \frac{1}{\Delta} (S_{xx} S_y - S_x S_{xy}) \\ m &= \frac{1}{\Delta} (n \cdot S_{xy} - S_x S_y) \end{aligned}$

Nun müssen nur noch die gegebenen Werte in die Formel eingesetzt werden um das Resultat zu erhalten. Es ist offensichtlich, dass sich die obigen Rechnungen ohne grossen Aufwand programmieren lassen. □

3 Statistische Grundlagen

In diesem Abschnitt werden die statistischen Grundlagen für die Rechnungen im vorangehenden Abschnitt kurz besprochen. Eine sorgfältige Diskussion ist aus Zeit- und Platz-Gründen im gegebenen Rahmen nicht möglich.

3.1 Begriffe der Wahrscheinlichkeitsrechnung

3.1.1 Normalverteilung

Die Funktion

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(\frac{-1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right)$$

beschreibt eine **Normalverteilung** mit **Mittelwert** μ und **Standardabweichung** σ . Der Ausdruck $V = \sigma^2$ heisst auch **Varianz**. Ist x eine Zufallsvariable mit der obigen Wahrscheinlichkeitsverteilung, so können wir die Wahrscheinlichkeit bestimmen mit der eine Messung von x in ein gegebenes Intervall $a \leq x \leq b$ fällt. Dazu ist eine Reihe von Messungen aufzunehmen. Die Wahrscheinlichkeit ist gegeben durch

$$\begin{aligned} P(a \leq x \leq b) &= \frac{\text{Anzahl Messwerte zwischen } a \text{ und } b}{\text{Totale Anzahl der Messungen}} \\ &= \int_a^b f(x) dx = \frac{1}{\sigma \sqrt{2\pi}} \int_a^b \exp\left(\frac{-1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) dx \end{aligned}$$

In **Abbildung 3** sehen Sie eine Normalverteilung mit Mittelwert $\mu = 3.5$ und Standardabweichung $\sigma = 1$. Der Wert von $\mu = 3.5$ können Sie aus der Lage des Maximums ablesen und die Breite dieser **Gauss'schen Glockenkurve** ist durch die Standardabweichung σ gegeben. Wegen

$$f(\mu) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2} \left(\frac{\mu - \mu}{\sigma}\right)^2\right) = \frac{1}{\sigma\sqrt{2\pi}}$$

$$f(\mu \pm \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(\frac{-1}{2} \left(\frac{\mu \pm \sigma - \mu}{\sigma}\right)^2\right) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{\sqrt{e}}$$

gilt

$$\frac{f(\mu \pm \sigma)}{f(\mu)} = \frac{1}{\sqrt{e}} \approx 0.6065$$

fällt der Wert von $f(x)$ auf ca. 60% der Maximalwertes ab falls man sich um eine Standardabweichung vom Maximum entfernt.

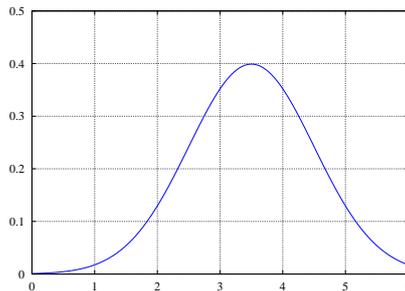


Abbildung 3: Eine Normalverteilung mit Mittelwert $\mu = 3.5$ und Standardabweichung $\sigma = 1$

1 Beispiel : Wir untersuchen diese Verteilung anhand einiger Beispielrechnungen. Als Musterkurve nehmen wir den Graphen in [Abbildung 3](#).

- Die Wahrscheinlichkeit, dass eine Messung von x grösser als $\mu = 3.5$ ist ergibt sich als

$$\begin{aligned} P(\mu \leq x) &= \int_{\mu}^{\infty} f(x) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu}^{\infty} \exp\left(\frac{-1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{3.5}^{\infty} \exp\left(\frac{-1}{2} (x - 3.5)^2\right) dx \\ &= 0.5 \end{aligned}$$

Somit sollten 50% der Messwerte grösser als 3.5 sein.

- Die Wahrscheinlichkeit, dass eine Messung von x zwischen $\mu - \sigma$ und $\mu + \sigma$ liegt berechnet sich durch

$$\begin{aligned} P(\mu - \sigma \leq x \leq \mu + \sigma) &= \int_{\mu - \sigma}^{\mu + \sigma} f(x) dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mu - \sigma}^{\mu + \sigma} \exp\left(\frac{-1}{2} \left(\frac{x - \mu}{\sigma}\right)^2\right) dx \\ &\approx 0.682689 \end{aligned}$$

Somit sollten ca. 68% der Messwerte um weniger als σ von μ abweichen.

- Wegen

$$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\mu-2\sigma}^{\mu+2\sigma} \exp\left(\frac{-1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right) dx \approx 0.9545$$

sollten ca. 95% der Messwerte um weniger als 2σ vom Mittelwert μ abweichen.

- Das Integral

$$P(-\infty < x < \infty) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right) dx = 1$$

ergibt immer eins. Dies entspricht der offensichtlichen Tatsache, dass die Messung eine Zahl zwischen $-\infty$ und ∞ ergeben muss. Dies muss der Fall sein, tatsächlich wird die Konstante vor dem Integral durch diese Bedingung bestimmt.

◇

Für die Normalverteilung kann man nachrechnen, dass

$$\begin{aligned} \int_{-\infty}^{\infty} x f(x) dx &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x \exp\left(\frac{-1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right) dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (s+\mu) \exp\left(\frac{-1}{2} \left(\frac{s}{\sigma}\right)^2\right) ds \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} s \exp\left(\frac{-1}{2} \left(\frac{s}{\sigma}\right)^2\right) ds + \frac{\mu}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-1}{2} \left(\frac{s}{\sigma}\right)^2\right) ds \\ &= \mu \end{aligned}$$

Für die letzte Umformung wurden Symmetrie und $P(-\infty < s < \infty) = 1$ verwendet. Man kann auch nachrechnen, dass

$$V = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu)^2 \exp\left(\frac{-1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right) dx$$

Somit kann μ als gewichteter Mittelwert von x und die Varianz $V = \sigma^2$ als gewichteter Mittelwert der quadratischen Abweichung $(x-\mu)^2$ aufgefasst werden. Die Gewichtung ist gegeben durch die Funktion $f(x)$ der Normalverteilung.

3.1.2 Schätzen von Mittelwert und Standardabweichung aus einer Messreihe

Von einem Experiment ist bekannt, dass die Messwerte normalverteilt sind, aber weder μ noch σ sind bekannt. Man führt eine Reihe von Messungen durch und erhält n Resultate x_1, x_2, \dots, x_n . Aus diesen sollen der Mittelwert μ und die Standardabweichung σ geschätzt werden. Die obigen Überlegungen führen sofort auf den Ansatz für den Mittelwert

$$\mu = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{S_x}{n}$$

Die Varianz wird geschätzt durch

$$V = \sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Der eventuell überraschende Faktor $(n - 1)$ (statt n) kann durch weitergehende statistische Argumente gerechtfertigt werden. Die Summe kann noch umgeschrieben werden zu

$$\begin{aligned} \sum_{i=1}^n (x_i - \bar{x})^2 &= \sum_{i=1}^n (x_i^2 - 2\bar{x}x_i + \bar{x}^2) \\ &= \sum_{i=1}^n x_i^2 - 2\bar{x} \sum_{i=1}^n x_i + n\bar{x}^2 \\ &= \sum_{i=1}^n x_i^2 - n\bar{x}^2 = S_{xx} - \frac{S_x^2}{n} \\ &= \frac{n S_{xx} - S_x^2}{n} \end{aligned}$$

Für die letzte Umformung wurde

$$\bar{x}^2 = \left(\frac{1}{n} \sum_{j=1}^n x_j \right)^2 = \frac{S_x^2}{n^2}$$

verwendet. Somit kann man auch die folgende Formulierung verwenden

$$\begin{aligned} \mu &= \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{S_x}{n} \\ V &= \frac{n S_{xx} - S_x^2}{n(n-1)} \\ \sigma &= \sqrt{\frac{n S_{xx} - S_x^2}{n(n-1)}} \end{aligned}$$

Bei sehr grossen Anzahlen n von Messungen sollte nicht die obigen Summenformeln verwendet werden, Rundungsfehler könnten einen grossen Einfluss haben. Verwenden Sie stattdessen

$$V = \frac{1}{n-1} \left(\sum_{i=1}^n (x_i - \bar{x})^2 - \frac{1}{n} \left(\sum_{i=1}^n (x_i - \bar{x}) \right)^2 \right)$$

Falls \bar{x} exakt bestimmt ist muss der zweite Term Null sein, sonst vermindert er üblicherweise den Einfluss von Rundungsfehlern ([Pres92, p. 613]).

3.2 Maximum-Likelihood Methode

Für eine Reihe von Messwerten (x_i, y_i) konstruieren wir mit Hilfe der intuitiv motivierten Bedingung

$$\text{Finde } a \text{ und } m \text{ sodass } \sum_{i=1}^n (f(x_i) - y_i)^2 \text{ minimal wird.}$$

eine Gerade. Diese Bedingung liefert die Werte der Parameter a und m und somit eine „passende“ Gerade. Es wird gezeigt in welchem Sinne diese Gerade die richtige Gerade ist. Dieser Abschnitt wurde aus [Pres92, p. 657] entnommen.

Es gibt nur eine „richtige“ Gerade, alle anderen sind „falsch“, resp. „weniger gut“. Die entscheidende Frage ist: *Falls die wahren Werte a und m bekannt sind, mit welcher Wahrscheinlichkeit können die gemessenen Werte (x_i, y_i) tatsächlich auftreten?* Die richtigen Werte von a und m sind dadurch charakterisiert, dass diese Wahrscheinlichkeitsdichte maximal ist.

Wir arbeiten mit den untenstehenden Annahmen. Bei konkreten Messungen müssen diese Annahmen gerechtfertigt werden.

1. Die Werte x_i werden exakt gemessen.
2. Untersuchen Sie gewählte Werte von a und m .
3. Die Messwerte von y_i sind normalverteilt um die Werte $a + m x_i$ mit einer von x_i unabhängigen Standardabweichung.
4. Nun wird die Wahrscheinlichkeit (präziser: Wahrscheinlichkeitsdichte) bestimmt, mit der die effektiv gemessenen Datenwerte y_i auftreten können, für die angenommenen Werte von a und m .
5. Für eine gegebene Messreihe suchen wir diejenigen Werte von a und m für welche die obige Wahrscheinlichkeitsdichte **maximal** ist.

Statt die Wahrscheinlichkeit auszurechnen mit der die Werte a und m „richtig“ sind, bestimmt man die Wahrscheinlichkeit des Eintreffens der (bereits gemessenen) Messdaten. Diese soll maximal werden. Man spricht von der **Maximum-Likelihood-Methode** (test du rapport des vraisemblances). Die Lösung dieser Aufgabe ist gegeben durch die Lösung des oben vorgestellten Minimalproblems.

3.3 Schätzen der Varianz von a und m

Eine neue Messreihe würde zu einer neuen Schätzung von a und m führen. Somit sollten man auch irgendwie abschätzen können wie zuverlässig die Werte von a und m sind. Dazu ist es **nicht angebracht** mehrere Messreihen aufzunehmen, sondern man solle mit einer einzigen Messreihe arbeiten können. Beachten Sie hierzu Abbildung 4. Nur die erste Messreihe wird aufgenommen, alle anderen sind hypothetische Messreihen. Es wird sich zeigen, dass die Werte von a und m auch wieder normalverteilt sind. Somit ist die Varianz (Quadrat der Standardabweichung) ein Mass für die zu erwartenden Fluktuationen.

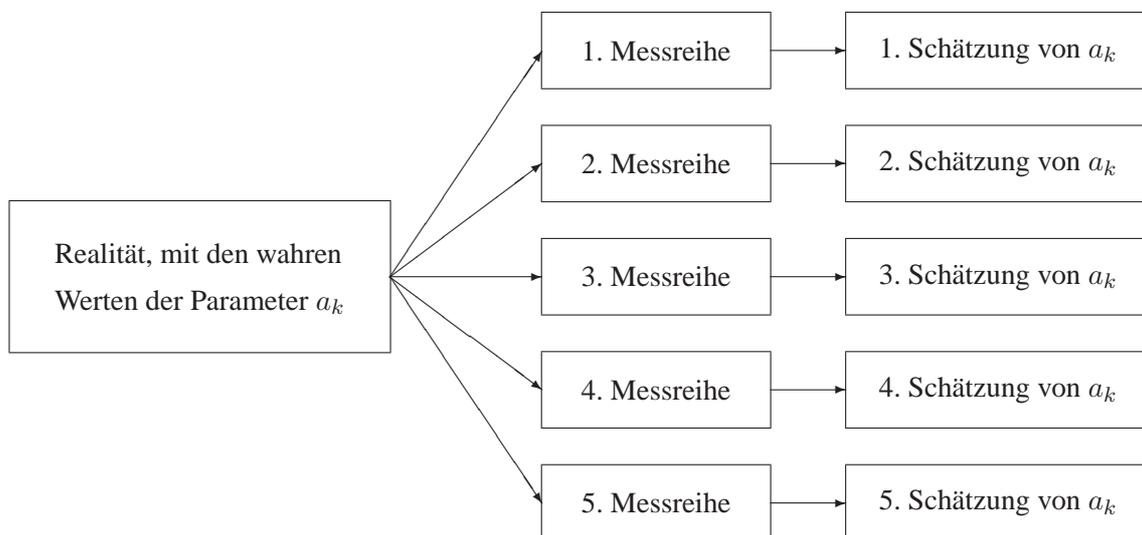


Abbildung 4: Verschiedene Messreihen führen zu verschiedenen Parametern

3.3.1 Bestimmen der Standardabweichung σ der Messwerte y_i

Manchmal kann die Standardabweichung σ der Messwerte aufgrund von Information über die Messanordnung geschätzt werden, ohne die Daten zu analysieren. Falls dies nicht möglich ist kann man versuchen mit Hilfe der Messwerte y_i eine Schätzung zu erhalten.

Wir gehen davon aus, dass die beiden Parameter a und m der Geraden bereits bestimmt wurden. Die Messwerte y_i sollten normalverteilt sein um die „richtigen“ Werte $a + m x_i$, d.h. die Differenzen $y_i - a - m x_i$ sollten normalverteilt sein mit Mittelwert Null. Das führt auf die Schätzung

$$\sigma^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - a - m x_i)^2$$

Der Nenner ist $(n-2)$, da bereits zwei Parameter bestimmt wurden.

3.3.2 Varianz einer Summe von Zufallsgrößen, Varianzen von a und m

Sind y_i **unabhängige**, normalverteilte Zufallsgrößen mit Varianzen $V(y_i)$, so ist auch eine Linearkombination dieser y_i eine normalverteilte Zufallsgröße und für die Varianzen gilt

$$\begin{aligned} V(y_1 + y_2) &= V(y_1) + V(y_2) \\ V(\alpha_1 y_1) &= \alpha_1^2 V(y_1) \\ V\left(\sum \alpha_i y_i\right) &= \sum \alpha_i^2 V(y_i) \end{aligned}$$

Diese Grundregeln für das Rechnen mit Varianzen können mit Hilfe von Integralen verifiziert werden. Somit können wir die folgenden Resultate verifizieren. Die Werte von x_i seien exakt bekannt.

$$S_y = \sum_{i=1}^n y_i \implies V(S_y) = \sum_{i=1}^n V(y_i) \quad \text{und} \quad V(S_x S_y) = S_x^2 \sum_{i=1}^n V(y_i)$$

Nun untersuchen wir die in den Formeln für a und m auftretenden Summen

$$\begin{aligned} S_{xx} S_y - S_x S_{xy} &= \sum_{i=1}^n (S_{xx} - S_x x_i) y_i \implies V(S_{xx} S_y - S_x S_{xy}) = \sum_{i=1}^n (S_{xx} - S_x x_i)^2 V(y_i) \\ n S_{xy} - S_x S_y &= \sum_{i=1}^n (n x_i - S_x) y_i \implies V(n S_{xy} - S_x S_y) = \sum_{i=1}^n (n x_i - S_x)^2 V(y_i) \end{aligned}$$

Wir nehmen an, dass die Standardabweichungen aller Werte y_i durch σ gegeben sind, d.h. $V(y_i) = \sigma^2$. Das ergibt

$$\begin{aligned} V(S_{xx} S_y - S_x S_{xy}) &= \sum_{i=1}^n (S_{xx}^2 - 2 S_{xx} S_x x_i + S_x^2 x_i^2) \sigma^2 \\ &= (n S_{xx}^2 - 2 S_{xx} S_x^2 + S_x^2 S_{xx}) \sigma^2 = (n S_{xx}^2 - S_{xx} S_x^2) \sigma^2 \\ V(n S_{xy} - S_x S_y) &= \sum_{i=1}^n (n^2 x_i^2 - 2 n x_i S_x + S_x^2) \sigma^2 \\ &= (n^2 S_{xx} - 2 n S_x S_x + n S_x^2) \sigma^2 = (n^2 S_{xx} - n S_x^2) \sigma^2 \end{aligned}$$

Dann folgt aus den Grundformeln

$$a = \frac{1}{\Delta} (S_{xx} S_y - S_x S_{xy}) \quad , \quad m = \frac{1}{\Delta} (n \cdot S_{xy} - S_x S_y) \quad \text{wobei} \quad \Delta = n \cdot S_{xx} - S_x^2$$

sofort

$$\sigma_a^2 = V(a) = \frac{1}{\Delta^2} (n S_{xx}^2 - S_{xx} S_x^2) \sigma^2 = \frac{S_{xx}}{n \cdot S_{xx} - S_x^2} \sigma^2$$

$$\sigma_m^2 = V(m) = \frac{1}{\Delta^2} (n^2 S_{xx} - n S_x^2) \sigma^2 = \frac{n}{n \cdot S_{xx} - S_x^2} \sigma^2$$

Somit können wir die Standardabweichungen der Geraden-Parameter a und m schätzen, mit Hilfe der gemessenen Werte (x_i, y_i) und der Standardabweichung σ der y -Werte von der Geraden.

Teile dieser Überlegungen wurden [Bevi69, p. 113] entnommen. ähnliche Überlegungen und Formeln finden Sie in [Pres92].

4 Ergänzungen

4.1 Allgemeine lineare Regression, Matrizennotation

Für eine Messreihe (x_i, y_i) sind wir auf das Gleichungssystem

$$\begin{bmatrix} n & S_x \\ S_x & S_{xx} \end{bmatrix} \begin{pmatrix} a \\ m \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}$$

gestossen. Diese Gleichung werden wir nun elementar umschreiben mit Hilfe einer Matrix \mathbf{F} und eines Vektors \vec{y} , welche die Daten enthalten.

$$\mathbf{F} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \text{und} \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix}$$

Es gilt

$$\begin{bmatrix} n & S_x \\ S_x & S_{xx} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} = \mathbf{F}^T \cdot \mathbf{F}$$

und

$$\begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ x_1 & x_2 & x_3 & \cdots & x_n \end{bmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{F}^T \cdot \vec{y}$$

Somit kann das obige System auch geschrieben werden als

$$\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y}$$

Für den Fall einer Regressionsgerade bringt diese Notationsänderung keine Vorteile. Aber mit derselben Notation können auch komplexere Probleme beschrieben werden.

Die obigen Ideen können nicht nur für das Einpassen einer Geraden in eine Menge von Datenpunkten verwendet werden. Man kann auch mehrere unabhängige Variablen betrachten und muss auch nicht unbedingt Geraden (oder Ebenen) verwenden. Die grossen zusätzlichen Schwierigkeiten liegen vor allem in der Notation und etwas längeren Rechnungen, die aber ein Computer gerne übernimmt.

Wir betrachten also mehrere Werte der unabhängigen Variablen x_1, x_2, \dots, x_n und Funktionen der Form

$$\Phi(x) = \sum_{j=1}^k p_j \cdot f_j(x)$$

wobei die Funktionen $f_j(x)$ vorgegeben sind. Die Parameter p_1, \dots, p_k sind zu berechnen. Für die gegebenen Datenpunkte (x_i, y_i) soll der Ausdruck

$$\chi^2 := \sum_{i=1}^n (\Phi(x_i) - y_i)^2 = \sum_{i=1}^n \left(\left(\sum_{j=1}^k p_j \cdot f_j(x_i) \right) - y_i \right)^2$$

minimiert werden.

Die Bedingung, dass alle partiellen Ableitungen bezüglich der p_j Null sein müssen führt wieder auf ein System von k linearen Gleichungen für die k Unbekannten p_j . Details werden in Abschnitt 4.3 untersucht.

2 Beispiel : Mit den Ausdrücken

$$k = 2 \quad p_1 = a \quad p_2 = m \quad f_1(x) = 1 \quad f_2(x) = x \quad \Phi(x) = a + mx$$

erhalten wir die Situation des vorangehenden Abschnitts: eine Gerade soll an Messdaten angepasst werden. \diamond

3 Beispiel : Mit den Parametern

$$\begin{array}{lll} p_1 = a & p_2 = b & p_3 = c \\ f_1(x) = x^2 & f_2(x) = x & f_3(x) = 1 \end{array}$$

wird $\Phi(x) = ax^2 + bx + c$ und wir suchen wir eine Parabel durch die gegebenen Punkte.

In Abschnitt 4.3 werden wir sehen, dass in diesem Fall die Matrix \mathbf{F} gegeben ist durch

$$\mathbf{F} = \begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ x_3^2 & x_3 & 1 \\ \vdots & \vdots & \vdots \\ x_n^2 & x_n & 1 \end{bmatrix}$$

und die im System $\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y}$ zu untersuchende Matrix $\mathbf{M} = \mathbf{F}^T \cdot \mathbf{F}$ ist eine 3×3 Matrix. \diamond

Es muss unbedingt herausgestrichen werden, dass die Funktionen $f_i(x)$ durchaus nichtlinear sein können. Die Linearität bezieht sich auf die Parameter p_j . In Tabelle 1 finden Sie einige Beispiele.

Funktionstyp	Parameter	lineare / nichtlineare Regression
$y = a + m x$	a, m	linear
$y = a x^2 + b x + c$	a, b, c	linear
$y = a e^{c x}$	a, c	nichtlinear
$y = a e^{c x}$	a	linear
$y = a \sin(\omega t + \delta)$	a, ω, δ	nichtlinear
$y = a \cos(\omega t) + b \sin(\omega t)$	a, b	linear

Tabelle 1: Beispiele von linearer und nichtlinearer Regression

4.2 Example : Intensity of light of an LED depending on the angle of observation

4.2.1 Description of the problem

The intensity of the light emitted by an LED will depend on the angle α of observation. The data sheets of the supplier should show this information. A sample of real data is stored in the file `LEDdata.txt` and the results is shown in Figure 5.

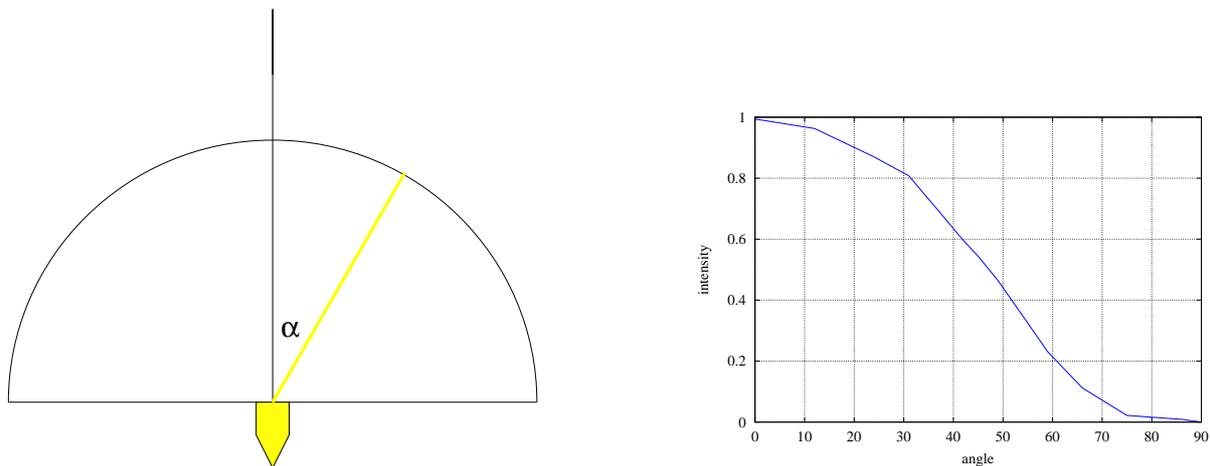


Figure 5: Intensity of light as function of the angle

4.2.2 A first, naive solution and its problems

To do further analysis it can be useful to have a formula for the intensity as function of the angle and linear regression is one of the options on how to obtain such a formula. We try to fit a polynomial of degree 4 through the given points, i.e. we seek parameters p_j such that

$$\Phi(x) = p_1 1 + p_2 x + p_3 x^2 + p_4 x^3 + p_5 x^4$$

In this case the matrix \mathbf{F} is given by

$$\mathbf{F} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 \\ 1 & x_3 & x_3^2 & x_3^3 & x_3^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & x_n^4 \end{bmatrix}$$

The following code will fit a polynomial of degree 4 through those points and then display the result in Figure 6. The algorithm is performing the following steps:

- Read the data.
- Construct the matrix \mathbf{F} with the correct columns
- Solve the linear system $\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y}$ of 5 linear equations for the 5 unknown components of the parameter vector \vec{p}

```

Octave
angle = [ -0 12 24 31 37 42 45 49 54 59 66 75 86 90]';
intensity = [ 0.994 0.963 0.87 0.808 0.693 0.596...
             0.544 0.464 0.348 0.231 0.112 0.022 0.009 0]';

F = ones(length(angle),5);
for k = 1:5
    F(:,k) = angle.^(k-1);
end

M = F'*F;
p = M\(F'*intensity)
```

The resulting parameters point towards an intensity function

$$\text{Intensity}(\alpha) = 124.28 + 0.1111\alpha - 4.0576 \cdot 10^{-3}\alpha^2 + 5.0299 \cdot 10^{-5}\alpha^3 - 2.1087 \cdot 10^{-7}\alpha^4$$

The code below generates Figure 6.

```

Octave
al = (0:1:90)'; % consider angles from 0 to 90 degree

Inew = polyval(flipud(p),al);
plot(angle,intensity,al,Inew);
grid on
xlabel('angle');ylabel('intensity')
```

The result in Figure 6 is useless since the estimated variances of the parameters are of the same order of magnitude as the values². We need to find the reason for the failure and how to eliminate the problems.

The previous implementation of the linear regression algorithm has to solve a system of equations with the matrix $\mathbf{F}' \cdot \mathbf{F}$.

With the help of

```

Octave
```

²Previous releases of MATLAB and Octave generated numbers which were obviously wrong. This was very useful for didactical purposes, but less so for real world problems.

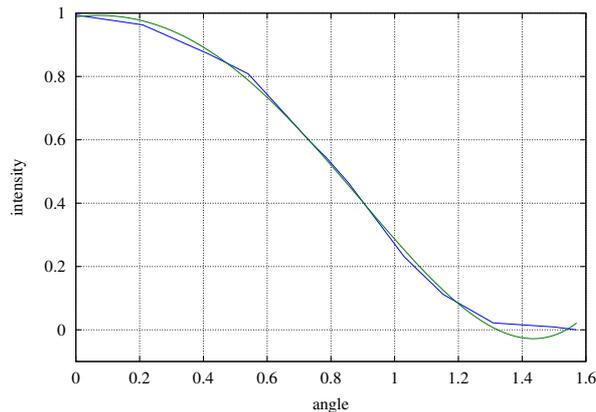


Figure 6: Intensity of light as function of the angle and a first regression curve

```
F' * F
ans =
 1.4000e+01    6.7000e+02    4.1114e+04    2.8161e+06    2.0767e+08
 6.7000e+02    4.1114e+04    2.8161e+06    2.0767e+08    1.6113e+10
 4.1114e+04    2.8161e+06    2.0767e+08    1.6113e+10    1.2949e+12
 2.8161e+06    2.0767e+08    1.6113e+10    1.2949e+12    1.0666e+14
 2.0767e+08    1.6113e+10    1.2949e+12    1.0666e+14    8.9414e+15
```

we see that the matrix contains numbers of the order 1 and of the order 10^{16} and one should not be surprised by trouble when solving such a system of equations. Mathematically speaking we have a **condition number** that is very large (10^{16}) and thus we will lose many digits of precision. Entries of vastly different sizes are an indication for large condition numbers, but other effects also matter and you will have to consult specialized literature or a mathematician to obtain more information.

4.2.3 How to eliminate the problem

There are different measures to be taken to avoid the problem. For a reliable solution they should all be taken into account.

1. Rescaling

For a polynomial of degree 4 and angles of 90° the matrix \mathbf{F} will contain numbers of the size 1 and 90^4 . Thus $\mathbf{F}' \cdot \mathbf{F}$ will contain numbers of the size $90^8 \approx 100^8 = 10^{16}$. If we switch to radians instead of degrees this will be reduced to $(\frac{\pi}{2})^8 \approx 40$ and thus should avoid the problem. The code below will generate a reasonable solution.

Octave

```
load 'LEDdata.txt'
scalefactor = 180/pi;
newangle = angle/scalefactor;

F = ones(length(newangle),5);
for k = 1:5
    F(:,k) = newangle.^(k-1);
end

M = F' * F;
p = M \ (F' * intensity)

al = ((0:1:90)')/scalefactor; % consider angles from 0 to 90 degree
Inew = polyval(flipud(p),al);
```

```

plot(newangle*scalefactor,intensity,al*scalefactor,Inew);
grid on
xlabel('angle');ylabel('intensity')

```

2. Better choice of basis functions

Since the intensity function $I(\alpha)$ has to be symmetric with respect to α , i.e. $I(-\alpha) = I(\alpha)$, there can be no contributions of the form α or α^3 . Thus we seek a good fit for a function of the type

$$I(\alpha) = p_1 + p_2 \alpha^2 + p_3 \alpha^4$$

The code below leads to the result in Figure 7. The condition number of $\mathbf{F}' \cdot \mathbf{F}$ is approximately 200 and thus poses no problem. The result in Figure 7 is now useful for further investigations and the computations indicate that the intensity is approximated by

$$I(\alpha) = 1.02951 - 0.95635 \alpha^2 + 0.21890 \alpha^4$$

The code is a slight modification of the previous code.

```

Octave
angle      = [ -0 12 24 31 37 42 45 49 54 59 66 75 86 90]';
intensity  = [ 0.994 0.963 0.87 0.808 0.693 0.596...
              0.544 0.464 0.348 0.231 0.112 0.022 0.009 0]';
scalefactor = 180/pi;
newangle   = angle/scalefactor;

F = ones(length(angle),3);
F(:,1) = ones(length(angle),1);
F(:,2) = newangle.^2;
F(:,3) = newangle.^4;

M = F'*F;
p = M\(F'*intensity)

al = ((0:1:90)')/scalefactor; % consider angles from 0 to 90 degree
Inew = polyval([p(3) 0 p(2) 0 p(1)],al);
plot(newangle*scalefactor,intensity,al*scalefactor,Inew);
grid on
xlabel('angle');ylabel('intensity')

```

This point is by far the most important aspect to consider when using the linear regression method.

Choose your basis function for linear regression very carefully,
based on information about the system to be examined.

There are many software packages (*Mathematica*, *MATLAB*, *Octave*, *Excel*, ...) that allow to perform linear regression with polynomials of high degrees. This author is not aware of **one single problem** where a polynomial of high degree lead to useful information. All software behave according to the **GIGO**³ principle.

3. QR decomposition instead of $\mathbf{F}' \cdot \mathbf{F}$ matrix

Idea and consequences of this change in algorithm is given in the next section. Any serious implementation of a linear regression method should use this modification. The provided *Octave*-code `LinearRegression()` does take this point into account.

³Garbage In, Garbage Out

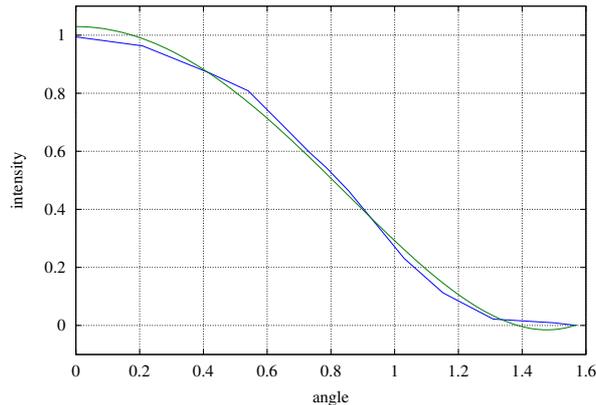


Figure 7: Intensity of light as function of the angle and regression with an even function

4.2.4 Using the command `LinearRegression()`

At the usual place you find the *Octave* script `LinearRegression.m`. Within *Octave* you may type `help LinearRegression` to access the builtin help information. The command will perform a general linear regression and also determine the estimated standard deviation for the parameters, see section 3.3.2 (page 10).

```

Octave
angle      = [ -0 12 24 31 37 42 45 49 54 59 66 75 86 90]';
intensity  = [ 0.994 0.963 0.87 0.808 0.693 0.596...
              0.544 0.464 0.348 0.231 0.112 0.022 0.009 0]';
scalefactor = 180/pi;
newangle   = angle/scalefactor;

F = ones(length(angle),3);
F(:,1) = ones(length(angle),1);
F(:,2) = newangle.^2;
F(:,3) = newangle.^4;

[p,y_var,r,p_var] = LinearRegression(F,intensity);
p
p_stddev = sqrt(p_var)

Inew = F*p;
plot(angle,intensity,angle,Inew)

```

Mit den obigen Notationen und der Realisierung in *Octave* durch den Befehl `LinearRegression()` können verschiedenste Klassen von Funktionen untersucht werden, mit der Absicht sie an die Daten anzupassen. Es muss aber herausgestrichen werden, dass die **Entscheidung über den Typ Funktion unbedingt aufgrund der Anwendung fallen muss.**

Wissen Sie, dass die gesuchte Kurve eine Parabel ist, so können Sie mit linearer Regression die drei Parameter bestimmen. Es ist sehr unüblich Regression mit Polynomen von hohem Grad zu versuchen.

4.3 Verifikation der Korrektheit der Gleichungen

Für gegebene Datenpunkte (x_i, y_i) für $1 \leq i \leq n$ und eine geeignete Wahl der Funktion der Basisfunktion $f_j(x)$ für $1 \leq j \leq k$ suchen wir die optimalen Werte der Parameter p_j , sodass die Funktion

$$\Phi(x) = \sum_{j=1}^k p_j f_j(x)$$

sich möglichst gut an die gegebenen Punkte anpasst. Finden Sie das System von Gleichungen für die Unbekannten p_j . Die Grundidee kann Abschnitt 2 entnommen werden.

Lösung : Diese Rechnungen sind etwas länger. Versuchen Sie trotzdem die Übersicht zu behalten. Die Lösung wird in mehrere Schritte unterteilt.

4.3.1 Aufstellen der Gleichungen

Die zu minimierende Funktion ist

$$\chi^2(\vec{p}) := \sum_{i=1}^n (\Phi(x_i) - y_i)^2 = \sum_{i=1}^n \left(\left(\sum_{j=1}^k p_j \cdot f_j(x_i) \right) - y_i \right)^2$$

bezüglich der Variablen p_l für $l = 1, \dots, k$. Die partiellen Ableitungen sind

$$\frac{\partial \Phi(x_i)}{\partial p_l} = \frac{\partial}{\partial p_l} \sum_{j=1}^k p_j \cdot f_j(x_i) = f_l(x_i) \quad \text{für } l = 1, 2, \dots, k$$

und somit müssen die Gleichungen

$$\frac{\partial \chi^2(\vec{p})}{\partial p_l} = 2 \sum_{i=1}^n \left(\sum_{j=1}^k p_j \cdot f_j(x_i) - y_i \right) \cdot f_l(x_i) = 0 \quad \text{für } l = 1, \dots, k$$

gelöst werden. Das kann umgeschrieben werden zu

$$\sum_{j=1}^k p_j \left(\sum_{i=1}^n f_j(x_i) \cdot f_l(x_i) \right) = \sum_{i=1}^n f_l(x_i) \cdot y_i \quad \text{für } l = 1, \dots, k$$

Es ist durchaus möglich direkt mit diesen Gleichungen zu arbeiten. Allerdings ist die Gefahr sehr gross, dass der Code wegen der vielen Summen und Indices unübersichtlich wird. Dieselben Rechnungen können mit Hilfe von Matrizen übersichtlicher dargestellt werden.

4.3.2 Matrix- und Vektornotation

Um das Folgende verstehen zu können muss man mit der Notation von Vektoren und Matrizen vertraut sein und auch mit deren Multiplikation. Das einzige Ziel ist es die obigen Ausdrücke umzuschreiben. Sei also

$$\vec{p} = \begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_k \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{F} = \begin{bmatrix} f_1(x_1) & f_2(x_1) & \dots & f_k(x_1) \\ f_1(x_2) & f_2(x_2) & \dots & f_k(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_n) & f_2(x_n) & \dots & f_k(x_n) \end{bmatrix}$$

Es ist leicht zu sehen, dass

$$(\mathbf{F} \cdot \vec{p})_i = \sum_{j=1}^k f_j(x_i) \cdot y_j \quad \text{für } i = 1, 2, \dots, n$$

Die obigen Rechnungen können auf mit dieser „neuen“ Notation geschrieben werden und wir erhalten für $\chi^2(\vec{p})$

$$\begin{aligned} \chi^2(\vec{p}) &= |\mathbf{F} \cdot \vec{p} - \vec{y}|^2 = (\mathbf{F} \cdot \vec{p} - \vec{y})^T \cdot (\mathbf{F} \cdot \vec{p} - \vec{y}) \\ &= (\vec{p}^T \cdot \mathbf{F}^T - \vec{y}^T) \cdot (\mathbf{F} \cdot \vec{p} - \vec{y}) \\ &= \vec{p}^T \mathbf{F}^T \mathbf{F} \vec{p} - y^T \mathbf{F} \vec{p} - \vec{p}^T \mathbf{F}^T \vec{y} + \vec{y}^T \vec{y} \end{aligned}$$

Aufgrund von $\vec{a}^T \vec{b} = \langle \vec{a}, \vec{b} \rangle$ (d.h. das Produkt eines Zeilenvektors mit einem Spaltenvektor ist gleich dem üblichen Skalarprodukt zweier Vektoren) erhalten wir

$$\chi^2(\vec{p}) = \langle \mathbf{F} \cdot \vec{p} - \vec{y}, \mathbf{F} \cdot \vec{p} - \vec{y} \rangle = \langle \mathbf{F} \vec{p}, \mathbf{F} \vec{p} \rangle - 2 \langle \mathbf{F} \vec{p}, \vec{y} \rangle + \langle \vec{y}, \vec{y} \rangle$$

Mit den obigen Notationen gilt

$$\begin{aligned} (\mathbf{F}^T \cdot \vec{y})_l &= \sum_{i=1}^n f_l(\vec{x}_i) \cdot y_i \\ (\mathbf{F}^T \cdot \mathbf{F})_{l,j} &= \sum_{i=1}^n f_l(\vec{x}_i) \cdot f_j(\vec{x}_i) \\ (\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p})_l &= \sum_{j=1}^m (\mathbf{F}^T \cdot \mathbf{F})_{l,j} p_j \end{aligned}$$

und somit kann das System von Gleichungen

$$\sum_{j=1}^m p_j \left(\sum_{i=1}^n f_j(\vec{x}_i) \cdot f_l(\vec{x}_i) \right) = \sum_{i=1}^n f_l(\vec{x}_i) \cdot y_i \quad l = 1, \dots, m$$

auch geschrieben werden als

$$\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y} \quad .$$

In diesem Ausdruck sind der Vektor \vec{y} und die Matrix \mathbf{F} gegeben durch die gegebenen Datenpunkte (x_i, y_i) und die Funktionen f_j . Zu bestimmen ist der Vektor \vec{p} .

4.3.3 Effiziente Codierung, QR-Zerlegung und Octave

Für eine Matrix mit mehr Zeilen als Spalten kann die QR-Zerlegung einer Matrix bestimmt werden durch

$$\mathbf{F} = \mathbf{Q} \cdot \mathbf{R}$$

wobei \mathbf{Q} eine orthogonale Matrix ist ($\mathbf{Q}^{-1} = \mathbf{Q}^T$) und \mathbf{R} eine obere Dreiecksmatrix

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_o \\ \mathbf{0} \end{bmatrix} \quad \text{und} \quad \mathbf{Q} = \begin{bmatrix} \mathbf{Q}_l & \mathbf{Q}_r \end{bmatrix}$$

Hierbei sind die oberen (resp. linken) Blöcke \mathbf{R}_o und \mathbf{Q}_l invertierbare Matrizen. Multiplizieren eines Vektors mit einer orthogonalen Matrix \mathbf{Q} entspricht einer Rotation des Vektors, somit ändert die Länge nicht. Zu lösen ist das Problem

$$\begin{aligned} \mathbf{F} \cdot \vec{p} - \vec{y} &= \vec{r} \quad \text{minimale Länge} \\ \mathbf{Q} \cdot \mathbf{R} \cdot \vec{p} - \vec{y} &= \vec{r} \quad \text{minimale Länge} \end{aligned}$$

$$\mathbf{R} \cdot \vec{p} - \mathbf{Q}^T \cdot \vec{y} = \mathbf{Q}^T \cdot \vec{r}$$

$$\begin{bmatrix} \mathbf{R}_o \\ \mathbf{0} \end{bmatrix} \cdot \vec{p} - \begin{bmatrix} \mathbf{Q}_l^T \\ \mathbf{Q}_r^T \end{bmatrix} \cdot \vec{y} = \begin{bmatrix} \mathbf{Q}_l^T \\ \mathbf{Q}_r^T \end{bmatrix} \cdot \vec{r}$$

$$\begin{bmatrix} \mathbf{R}_o \cdot \vec{p} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{Q}_l^T \cdot \vec{y} \\ \mathbf{Q}_r^T \cdot \vec{y} \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_l^T \cdot \vec{r} \\ \mathbf{Q}_r^T \cdot \vec{r} \end{bmatrix}$$

Der Vektor \vec{p} hat keinen Einfluss auf die untere Hälfte dieses Systems. Somit ist die Länge des Vektors auf der rechten Seite minimal, falls $\mathbf{Q}_r^T \cdot \vec{r} = \vec{0}$. In dieser Form können die ersten Gleichungen verwendet werden, um \vec{p} eindeutig zu bestimmen durch

$$\mathbf{R}_o \cdot \vec{p} = \mathbf{Q}_l^T \cdot \vec{y}$$

$$\vec{p} = \mathbf{R}_o^{-1} \cdot \mathbf{Q}_l^T \cdot \vec{y}$$

Der Einsatz der QR-Zerlegung ist dem Lösen des Systems $\mathbf{F}^T \cdot \mathbf{F} \cdot \vec{p} = \mathbf{F}^T \cdot \vec{y}$ vorzuziehen, da die Konditionierungszahl erheblich kleiner ist. Somit sind weniger numerische Probleme zu erwarten.

Für das einführende Beispiel auf den ersten Seiten erhalten wir

$$\vec{x} = \begin{pmatrix} 0 \\ 1 \\ 2 \\ 3.5 \\ 4 \end{pmatrix}, \quad \vec{y} = \begin{pmatrix} -0.5 \\ 1 \\ 2.4 \\ 2.0 \\ 3.1 \end{pmatrix} \quad \text{für} \quad \mathbf{F} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3.5 \\ 1 & 4 \end{bmatrix}$$

Die QR-Zerlegung von \mathbf{F} ist gegeben durch

$$\mathbf{F} = \mathbf{Q} \cdot \mathbf{R} = \begin{bmatrix} -0.447 & -0.627 & -0.421 & -0.350 & -0.326 \\ -0.447 & -0.329 & 0.034 & 0.504 & 0.661 \\ -0.447 & -0.030 & 0.846 & -0.197 & -0.211 \\ -0.447 & 0.418 & -0.219 & 0.562 & -0.512 \\ -0.447 & 0.568 & -0.240 & -0.519 & 0.388 \end{bmatrix} \cdot \begin{bmatrix} -2.236 & -4.696 \\ 0 & 3.347 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Das effektiv zu lösende Problem ist somit

$$\begin{bmatrix} -2.236 & -4.696 \\ 0 & 3.347 \end{bmatrix} \cdot \begin{pmatrix} a \\ m \end{pmatrix} = \begin{bmatrix} -0.447 & -0.447 & -0.447 & -0.447 & -0.447 \\ -0.627 & -0.329 & -0.030 & 0.418 & 0.568 \end{bmatrix} \begin{pmatrix} -0.5 \\ 1 \\ 2.4 \\ 2.0 \\ 3.1 \end{pmatrix}$$

$$= \begin{pmatrix} -3.578 \\ 2.510 \end{pmatrix}$$

Dieses System ist leicht von unten nach oben auflösbar mit der Lösung $(a, m) = (0.025, 0.750)$. Der untenstehende Octave/MATLAB-Code führt die beschriebenen Rechnungen aus.

Octave

```
x = [0;1;2;3.5;4]; y = [-0.5;1;2.4;2.0;3.1];

F = ones(length(x),2); % construct the matrix F
F(:,2) = x; % values of x coordinates as second column

[Q,R] = qr(F) % factorization
p = R(1:2,1:2)\(Q(:,1:2)'*y) % solve the upper block of R*p=Q'*y
```

Falls die vollständigen Matrizen **R** und **Q** nicht gebraucht werden, können die letzten beiden Codezeilen ersetzt werden durch den folgenden Code.

Octave

```
[Q,R] = qr(F,0);
p = R\ (Q'*y)
```

□

4.4 Messungen mit verschiedener Genauigkeit

4.4.1 Regressionsgerade

Für eine Reihe von Messwerten (x_i, y_i) sind die x_i -Werte exakte bekannt und die Standardabweichung der y_i -Werte ist gegeben durch σ_i . In den vorangehenden Abschnitten haben wir immer verwendet, dass die Werte σ_i alle gleich sind. Dies muss nicht der Fall sein. Ist σ_i klein, so heisst dass, das der zugehörige Wert y_i sehr genau bekannt ist. Diesem Messpunkt sollte also ein grösseres Gewicht zugeordnet werden als einem Messwert mit grosser Standardabweichung. Die wird realisiert durch die zu minimierende Funktion

$$\chi^2 := \sum_{i=1}^n \left(\frac{a + m x_i - y_i}{\sigma_i} \right)^2$$

Damit diese Funktion minimal wird müssen die Ableitungen bezüglich a und m verschwinden. Das führt auch hier auf das Gleichungssystem

$$\begin{bmatrix} S & S_x \\ S_x & S_{xx} \end{bmatrix} \begin{pmatrix} a \\ m \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}$$

wobei

$$\begin{aligned} S &= \sum_{i=1}^n \frac{1}{\sigma_i^2} & S_x &= \sum_{i=1}^n \frac{x_i}{\sigma_i^2} & S_y &= \sum_{i=1}^n \frac{y_i}{\sigma_i^2} \\ S_{xx} &= \sum_{i=1}^n \frac{x_i^2}{\sigma_i^2} & S_{xy} &= \sum_{i=1}^n \frac{x_i y_i}{\sigma_i^2} \end{aligned}$$

Die eindeutig bestimmte Lösung ist hier

$$\begin{aligned} \Delta &= S \cdot S_{xx} - S_x^2 \\ a &= \frac{1}{\Delta} (S_{xx} S_y - S_x S_{xy}) \\ m &= \frac{1}{\Delta} (S \cdot S_{xy} - S_x S_y) \end{aligned}$$

Neben der obigen, anschaulichen Erklärung lässt sich auch hier die **Maximum-Likelihood-Methode** einsetzen, um die Resultate rechtfertigen.

Die obigen Formeln wurden [[Pres92](#), §15.3, p. 661] entnommen.

Octave

```
x = [0.6 ; 0.7 ; 0.8 ; 0.9 ; 1. ; 1.1 ; 1.2 ; 1.3];  
y = [2.61445; 2.78728; 2.69749; 2.89113; 3.0522; 3.22576; 3.4746; 3.68019];  
lnx = log(x); lny = log(y);  
F = [ones(size(x)) lnx];  
p = LinearRegression(F, lny)
```

Das numerische Ergebnis $p = (p_1, p_2)^T = (1.14483, 0.43151)^T$ führt zum Resultat

$$\begin{aligned}\ln(y) &= 1.14483 + 0.43151 \cdot \ln(x) \\ y &= e^{1.14483} x^{0.43151} = 3.1419 x^{0.43151}\end{aligned}$$

(c) Die obigen Parameterwerte für a und b minimieren **nicht** die Funktion

$$\sum_{i=1}^n (a \cdot x_i^b - y_i)^2$$

Um herauszufinden welcher Ausdruck minimiert wird, müssen wir uns die Funktion χ^2 etwas genauer ansehen und die Approximation $\ln(1+z) \approx z$ für $|z| \ll 1$ verwenden.

$$\begin{aligned}\chi^2 &= \sum_{i=1}^n (\ln a + b \ln x_i - \ln y_i)^2 = \sum_{i=1}^n \left(\ln \frac{a \cdot x_i^b}{y_i} \right)^2 \\ &= \sum_{i=1}^n \left(\ln \left(1 + \frac{a \cdot x_i^b - y_i}{y_i} \right) \right)^2 \approx \sum_{i=1}^n \left(\frac{a \cdot x_i^b - y_i}{y_i} \right)^2\end{aligned}$$

Somit entspricht ein Minimum von χ^2 in etwa einem Minimum der Summe der Quadrate der **relativen Fehler** $(ax_i^b - y_i)/y_i$.

Such man das Minimum der Funktion

$$\xi^2 = \sum_{i=1}^n (a \cdot x_i^b - y_i)^2$$

so führt dies auf ein nichtlineares Interpolationsproblem. Die Resultate können durchaus verschieden sein. □

5 Beispiel : Muss eine Kurve der Form

$$f(x) = a \cdot e^{b \cdot x}$$

durch Messpunkte gefittet werden so verwendet man

$$\ln y = \ln a + b \cdot x$$

und somit kann auch das Minimum der Funktion

$$\sum_{i=1}^n (\ln a - b \cdot x_i - \ln y_i)^2$$

gesucht werden. Ähnlich wie im obigen Beispiel verwendet man

Octave

```
F = [ones(size(x)) x];  
p = LinearRegression(F, lny)
```

Das numerische Ergebnis $p = (0.64732, 0.48645)^T$ führt zum Resultat

$$\begin{aligned}\ln(y) &= 0.64732 + 0.48645 \cdot x \\ y &= e^{0.64732} \cdot e^{0.43151 x} = 1.9104 e^{0.43151 x}\end{aligned}$$

◇

4.6 Fehler in x und y -Werten

Bei allen obigen Rechnungen sind wir davon ausgegangen, dass die x -Werte exakt bekannt sind. Ist dies nicht der Fall, so müssen die Rechnungen erheblich modifiziert werden [[Pres92](#), §15.3, p. 666] oder [[Bevi69](#)].

4.7 Octave/MATLAB-Befehl `LinearRegression()`

Der Autor dieser Notizen (A. Stahel) hat einen *Octave* oder *MATLAB*-Befehl `LinearRegression()` geschrieben mit den folgenden Fähigkeiten:

- Lineare Regression kann mit beliebigen Basisfunktionen durchgeführt werden. Die optimalen Parameterwerte werden bestimmt.
- Die Varianz der abhängigen Grösse und die Länge des Residualvektors werden bestimmt.
- Die Varianzen der Parameter werden geschätzt.
- Die Online-Dokumentation ist unten angegeben. Mehr Information finden Sie in [[Octave07](#)].

Octave

```
octave:1> help LinearRegression  
general linear regression  
  
[p,y_var,r,p_var] = LinearRegression(F,y)  
[p,y_var,r,p_var] = LinearRegression(F,y,weight)  
  
determine the parameters p_j (j=1,2,...,m) such that the function  
f(x) = sum_(i=1,...,m) p_j*f_j(x) fits as good as possible to the  
given values y_i = f(x_i)  
  
parameters  
F n*m matrix with the values of the basis functions at the support points  
  in column j give the values of f_j at the points x_i (i=1,2,...,n)  
y n column vector of given values  
weight n column vector of given weights  
  
return values  
p m vector with the estimated values of the parameters  
y_var estimated variance of the error  
r weighted norm of residual  
p_var estimated variance of the parameters p_j
```

Detaillierte Dokumentation und Anwendungsbeispiele sind in den Kursunterlagen zu *MATLAB* und *Octave*-Programmierung enthalten.

Literatur

- [Bevi69] P. R. Bevington. *Data Reduction and Error Analysis for the Physical Sciences*. McGraw–Hill, New York, 1969.
- [Pres92] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C, The Art of Scientific Computing*. Cambridge University Press, second edition, 1992.
- [Octave07] A. Stahel. Octave at the HTI Biel. lecture notes, 2007.