Methode der kleinsten Quadrate

Siehe Kapitel 7 Blobel— Methode der kleinsten Quadrate $\equiv$ Least Square Fit (LSQ)
Zur Idee

- Direkte Meßwerte mit Eigenschaften einer Zufallsvariable (folgen einer Wahrscheinlichkeitsdichte) werden mit y_i bezeichnet
- Abweichung von y_i um σ vom wahren Wert

$\Rightarrow y_i$ Stichprobe mit zugehöriger Wahrscheinlichkeitsdichte.
Weiter Annahme:
Es existiert ein Modell mit Parametern a_i, welche die Werte fü r y_i vorhersagen.

$$f(a_i, y_i) = 0 \tag{1}$$

Gleichung dieser Form heißen Bedingungen.
Prinzip der kleinsten Quadrate besagt, dass

$$S = \sum_{i=1}^{n} \Delta y_i^2 = \text{Minimum} \tag{2}$$

wobei die Residuen definiert sind als

$$\Delta y_i = f(a_i, x_i) - y_i \tag{3}$$

Gilt nur für unkorrelierte Daten.
Einfachster Fall n Messungen y_i des Modells y

$$\sum_{i=1}^{n} (y - y_i)^2 = \text{Minimum} \tag{4}$$

$$\Rightarrow \hat{y} = \sum_{i=1}^{n} \frac{y_i}{n} \tag{5}$$

Allgemeiner Fall

Daten werden beschrieben durch einen n-Vektor $\vec{y}$. Die Residuen sind dann auch ein n-Vektor $\vec{\Delta y}$.

$$S = \vec{\Delta y} V^{-1} \vec{\Delta y} = \text{Minimum} \tag{6}$$

V^{-1} heißt Kovarianzmatrix. Beziehung zu neg. log-likelihood Funktion F

$$S(\vec{a}) = 2F(\vec{a}) \tag{7}$$

Methode ist auf eine große Klasse anwendbar
Messungen werden von einem Modell vorhergesagt. LSQ führt zu einem Kriterium, ob Meßdaten statistisch verträglich mit einer Modellparametrisierung ist.

Bestimmung der Parameter

Annahme Modell hängt linear von Parameter ab

$$f(x, \vec{a}) = a_1 f_1(x) + a_2 f_2(x) + \cdots + a_p f_p(x) \tag{8}$$

Wir haben n Messungen y_i an den Stellen x_i. Erwartungswert jeder Messungen ist

$$\langle y_i \rangle = f(x_i, \bar{a}) \tag{9}$$

wobei $\bar{a}$ die wahren Parameter von $\vec{a}$ sind.
Wir def. den Residuenvektor

$$\vec{r} = \vec{y} - f(\vec{x}, \vec{a}) \tag{10}$$

Residuen haben die folgenden Eigenschaften

$$\langle r_i \rangle = 0 \qquad \langle r_i^2 \rangle = \sigma_i^2 \tag{11}$$

Es wird keine Annahme über die Wahrscheinlichkeitsdichte oder Verteilung der Residuen gemacht! Insbesondere ob Gaußverteilt oder nicht.
Einzige Annahme:

- Unverzerrtheit
- Endliche Varianz

Im Fall gleicher Varianz und unkorrelierter Daten

$$S = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y_i - a_1 f_1(x_i) - a_2 f_2(x_i) - \cdots - a_p f_p(x_i))^2 \tag{12}$$

Bedingungen für Minimum alle partiellen Ableitung müssen verschwinden

$$\frac{\partial S}{\partial a_1} = 2\sum f_1(x_i)(a_1 f_1(x_i) + a_2 f_2(x_i) + \cdots + a_p f_p(x_i) - y_i) = 0 \tag{13}$$

$$\frac{\partial S}{\partial a_2} = 2\sum f_2(x_i)(a_1 f_1(x_i) + a_2 f_2(x_i) + \cdots + a_p f_p(x_i) - y_i) = 0 \tag{14}$$

$$\vdots \qquad \vdots \tag{15}$$

$$\frac{\partial S}{\partial a_p} = 2\sum f_p(x_i)(a_1 f_1(x_i) + a_2 f_2(x_i) + \cdots + a_p f_p(x_i) - y_i) = 0 \tag{16}$$

Die Gleichungen können in Form der Normalgleichungen gebracht werden

$$a_1 \sum_{i=1}^{n} f_1^2(x_i) + \cdots + a_p \sum_{i=1}^{n} f_1(x_i) f_p(x_i) = \sum_{i=1}^{n} y_i f_1(x_i) \tag{17}$$

$$a_1 \sum_{i=1}^{n} f_2(x_i) f_1(x_i) + \cdots + a_p \sum_{i=1}^{n} f_2(x_i) f_p(x_i) = \sum_{i=1}^{n} y_i f_2(x_i) \tag{18}$$

$$\vdots \tag{19}$$

$$a_1 \sum_{i=1}^{n} f_p(x_i) f_1(x_i) + \cdots + a_p \sum_{i=1}^{n} f_p^2(x_i) = \sum_{i=1}^{n} y_i f_p(x_i) \tag{20}$$

LSQ in Matrixform

$$\vec{r} = \vec{y} - A\vec{a} \tag{21}$$

$$A = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \dots & f_p(x_1) \\ \vdots & \vdots & \vdots & \vdots \\ f_1(x_n) & f_2(x_n) & \dots & f_p(x_n) \end{pmatrix} \tag{22}$$

$$\vec{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \vec{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \tag{23}$$

$$S = \vec{r}^T \vec{r} = (y - Aa)^T (y - Aa) \tag{24}$$

$$= y^t y - 2a^T A^T y = a^T A^T Aa \tag{25}$$

$$\Rightarrow \min S \Rightarrow -2A^T y + 2A^T A\hat{a} = 0 \tag{26}$$

$$(A^T A)\hat{a} = A^T y \text{ Normalgleichung} \tag{27}$$

$(A^T A)$ ist symmetrische $p \times p$ Matrix

$$A^T A = \begin{pmatrix} \sum f_1^2 & \sum f_1 f_2 & \dots & \sum f_1 f_p \\ \vdots & \vdots & \vdots & \vdots \\ \sum f_p f_1 & \sum f_p f_2 & \dots & \sum f_p^2 \end{pmatrix} \tag{28}$$

$$A^T y = \begin{pmatrix} \sum y_i f_1 \\ \vdots \\ \sum y_i f_p \end{pmatrix} \tag{29}$$

$$\hat{a} = (A^T A)^{-1} A^T y \tag{30}$$

$\hat{a}$ ist beste Schätzung.
Geht aus lineare Transformation hervor.
Bestimme den Fehler von $\hat{a}$ aus Fehler der $\vec{y}$

$$V(\vec{y}) = \begin{pmatrix} \mathrm{var} y_1 & \mathrm{cov}(y_1, y_2) & \dots & \mathrm{cov}(y_1, y_n) \\ \vdots & \vdots & \vdots & \vdots \\ \mathrm{cov}(y_n, y_1) & \mathrm{cov}(y_n, y_2) & \dots & \mathrm{var} y_n \end{pmatrix} \tag{31}$$

Im Fall unkorrelierter Daten mit konstantem σ

$$V(\vec{y}) = \sigma^2 I_n = \begin{pmatrix} \sigma^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma^2 \end{pmatrix} \tag{32}$$

Damit ist der Fehler nach Fehlerfortpflanzung

$$V(\hat{a}) = BV(y)B^T \tag{33}$$

$$\hat{a} = By \tag{34}$$

$$B = (A^T A)^{-1} A^T \tag{35}$$

$$\Rightarrow V(\hat{a}) = (A^T A)^{-1} A^T V(y) A (A^T A)^{-1} \tag{36}$$

$$V(\hat{a}) = \sigma^2 (A^T A)^{-1} \tag{37}$$

$V(\hat{a})$ hängt nur von A (fester Matrix) und σ^2 ab, aber nicht von r. Bei konstantem σ^2 hat die Varianz σ^2 keinen Einfluss auf Parameter.
Summe der Quadrate für beste Schätzung $\hat{a}$

$$\hat{S} = y^T Y - 2\hat{a}^T A^T y + \hat{a}^T A^T \underbrace{A(A^T A)^{-1} A^T y}_{y} \tag{38}$$

$$= y^t Y - \hat{a}^T A^T y \tag{39}$$

Erwartungswert für $\hat{S}$

$$E[\hat{S}] = \sigma^2 (n - p) \tag{40}$$

Wenn die Varianz der Messdaten nicht bekannt ist kann diese geschätzt werden

$$\hat{\sigma}^2 = \frac{\hat{S}}{n - p} \tag{41}$$

Korrektur der Datenwerte

Nachdem Parameter bestimmt wurden lassen sich Modellvorhersagen zu beliebigem x treffen.

$$\hat{y}(x) = f(x, \hat{a}) = \sum_{i=1}^{n} \hat{a}_i f_i(x) \tag{42}$$

Speziell für die Messdaten y_i sind die korrigierte Datenpunkt $\hat{y}(x_i)$. Diese sollten auf Auffälligkeiten überprüft werden.

Allgemeiner Fall unterschiedlicher Fehler

$$V(y) = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \sigma_n^2 \end{pmatrix} \tag{43}$$

Dann sucht man ein Minimum für

$$S = \sum_{i=1}^{n} \frac{r_i^2}{\sigma_i^2} \tag{44}$$

Führe Gewichte ein

$$W(y) = \begin{pmatrix} 1/\sigma_1^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & 1/\sigma_n^2 \end{pmatrix} \tag{45}$$

$$S = r^T W r = (y - Aa)^T W (y - Aa) \tag{46}$$

Gilt auch im Fall korrelierter Daten.
Lösung

$$\hat{a} = (A^T W A)^{-1} A^T W y \tag{47}$$
$$V[\hat{a}] = (A^T W A)^{-1} \tag{48}$$
$$\hat{S} = y^T W y - \hat{a}^T A^T W y \tag{49}$$
$$E[\hat{S}] = n - p \tag{50}$$
$$V[\hat{y}] = A(A^T W A) - 1 A^T \tag{51}$$

Beispiel Geradenanpassung

Konstruiere die Matrizen für $f(x) = a_1 + a_2 x$

$$A = \begin{pmatrix} f_1(x_1) & f_2(x_1) \\ f_1(x_2) & f_2(x_2) \\ \vdots & \vdots \\ f_1(x_n) & f_2(x_n) \end{pmatrix} \tag{52}$$

$$= \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix} \tag{53}$$

Die Fehler sind

$$V = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & \sigma_n^2 \end{pmatrix} \tag{54}$$

$$W = \begin{pmatrix} 1/\sigma_1^2 & 0 & \dots & 0 \\ 0 & 1/\sigma_2^2 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 1/\sigma_n^2 \end{pmatrix} \tag{55}$$

Fitparameter und Werte sind

$$a = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \tag{56}$$

$$\hat{a} = (A^T W A)^{-1} A^T W y \tag{57}$$

$$A^T W A = \begin{pmatrix} \sum w_i & \sum w_i x_i \\ \sum w_i x_i & \sum w_i x_i^2 \end{pmatrix} = \begin{pmatrix} S_1 & S_x \\ S_x & S_{xx} \end{pmatrix} \tag{58}$$

$$A^T W y = \begin{pmatrix} \sum w_i y_i \\ \sum w_i x_i y_i \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix} \tag{59}$$

$$\hat{a}_1 = \frac{1}{S_1 S_{xx} - S_x^2} \left(S_{xx} S_y - S_x S_{xy} \right) \tag{60}$$

$$\hat{a}_2 = \frac{1}{S_1 S_{xx} - S_x^2} \left(i - S_x S_y - S_1 S_{xy} \right) \tag{61}$$

$$V[\hat{a}] = \frac{1}{S_1 S_{xx} - S_x^2} \begin{pmatrix} S_{xx} & -S_x \\ -S_x & S_1 \end{pmatrix} \tag{62}$$

$$V[\hat{y}] = \frac{1}{S_1 S_{xx} - S_x^2} \left(S_{xx} + x^2 S_1 - 2x S_x \right) \tag{63}$$