Vorlesung
**Statistische Methoden der Datenanalyse**
Prof. Dr. Dr. Wolfgang Rhode

# Punktschätzung, Fitten und Regularisierung



## Überblick

- Punktschätzung
- Minimierung
  - Methode der kleinsten Quadrate
  - Gewichtete Methode der kleinsten Quadrate
  - Maximum-Likelihood-Methode
- Parameterschätzung mit Bayes Theorem
- Regularisierung





## Punktschätzung

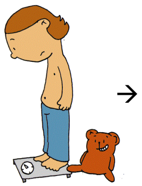




## Motivation

- Warum schätzen?
  - Genaue Messung möglicherweise zeitaufwendig und teuer
  - Messung vielleicht unmöglich
- Was brauchen wir zum Schätzen?
  - Stichprobe (kleine zufällige Auswahl der Messungen)
  - → Schätzwert mit berechenbarer Unsicherheit
- Verschiedene Methoden
  - Punktschätzer (Konkreten unbekannten Parameter)
  - Intervallschätzer (Bereich für unbekannten Parameter)

## Motivation



100 x → Stichprobe ($X_1, X_2, ..., X_{100}$)

→ Waage liefert Messwerte der Normalverteilung N(μ, σ²)

Fragestellungen:

- Wie ist der Erwartungswert μ des Gewichts? **(Punktschätzung)**
- Wie ist die Standardabweichung σ des Gewichts bzw. die Genauigkeit der Waage? **(Punktschätzung)**
- In welchem Bereich liegt das Gewicht mit einer vorgegebenen Sicherheit? **(Intervallschätzung)**





## Punktschätzung

- Definition: Bestimmung eines **einzelnen** Wertes zur Schätzung eines unbekannten Parameters.
  $\theta$ : zu schätzender Parameter
  $\hat{\theta}$ : Schätzwert des zu schätzenden Parameters

- Interessierender unbekannter Parameter ist oft ein Parameter der Wahrscheinlichkeitsverteilung von Beobachtungen.
  Beispiel: Erwartungswert μ der Normalverteilung N(μ, σ²)

- Punktschätzungen ermöglichen keine Ableitungen von Genauigkeiten der Schätzung. Zur Berechnung von Genauigkeiten werden Intervallschätzungen benutzt (→ siehe Kapitel „Intervallschätzung“).





## Gebräuchliche Punktschätzer

- Modus: Häufigster Wert
- Median
- Arithmetisches Mittel
- Quadratisches Mittel
- Geometrisches Mittel
- Harmonisches Mittel

$$\bar{x}_{\text{med}} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & n \text{ ungerade,} \\ \frac{1}{2}\left(x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}\right), & n \text{ gerade.} \end{cases}$$

$$\bar{x}_{\text{arithm}} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

$$\bar{x}_{\text{quadr}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} = \sqrt{\frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n}}$$

$$\bar{x}_{\text{geom}} = \sqrt[n]{\prod_{i=1}^{n} x_i} = \sqrt[n]{x_1 x_2 \cdots x_n}$$

$$\bar{x}_{\text{harm}} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}$$





## Berechnung von Punktschätzern

- Methoden:
  - Momenten-Schätzer
  - Maximum-Likelihood-Schätzer

## Berechnung von Punktschätzern

- Verschiedene Momenten-Schätzer für θ:

$$\frac{1}{n}\sum_{i=1}^{n} X_i,\quad \frac{1}{n}\sum_{i=1}^{n} X_i^2,\quad \frac{1}{n}\sum_{i=1}^{n} X_i^3,\quad \ldots$$

1. Moment    2. Moment    3. Moment



technische universität dortmund
Experimentelle Physik Vb
Astroteilchenphysik

## Berechnung von Punktschätzern

- Beispiel Momenten-Schätzer:
  Schätzen von Mittelwert und Varianz der Normalverteilung

$$X_i \text{ seien u.i.v. Zufallsvariablen mit } X_1 \sim N(\mu, \sigma^2)$$

$$\text{Zu schätzen: } \theta = (\mu, \sigma^2)$$

$$\tau_1(\theta) = \mu = E(X_1) \Rightarrow \hat{\mu} = \bar{X}_n = \frac{1}{n}\sum X_i$$

$$\tau_2(\theta) = \sigma^2 = Var(X_1) = E(X_1^2) - E(X1)^2 \Rightarrow \hat{\sigma}^2 = \frac{1}{n}\sum X_i^2 - \left(\frac{1}{n}\sum X_i\right)^2 = \frac{1}{n}\sum (X_i - \bar{X}_n)^2$$

Prof. Dr. Dr. W. Rhode
Schätzen
Statistische Methoden der Datenanalyse

technische universität dortmund
Experimentelle Physik Vb
Astroteilchenphysik

## Berechnung von Punktschätzern

- Maximum-Likelihood-Schätzer für θ:

$$\text{Likelihood:}$$

$$L(\theta \mid x_1, ..., x_n) = f(x_1, ..., x_n; \theta) = \prod f(x_i; \theta) = \prod L(\theta \mid x_i)$$

$$\text{Log-Likelihood (oft einfacher zu berechnen):}$$

$$l(\theta \mid x_1, ..., x_n) = log(L(\theta \mid x_1, ..., x_n)) = f(x_1, ..., x_n; \theta) = \sum log(f(x_i; \theta)) = \sum l(\theta \mid x_i)$$

→ Aufgrund der Monotonie des Logarithmus‘ sind Maxima gleich

Prof. Dr. Dr. W. Rhode
Schätzen
Statistische Methoden der Datenanalyse

technische universität dortmund
Experimentelle Physik Vb
Astroteilchenphysik

## Berechnung von Punktschätzern

- Beispiel Likelihood-Schätzer:
  Schätzen von Mittelwert und Varianz der Normalverteilung

$$X_i \text{ seien u.i.v. Zufallsvariablen mit } X_1 \sim N(\mu, \sigma^2)$$

$$\text{Zu schätzen: } \theta = (\mu, \sigma^2)$$

$$L(\mu, \sigma^2 | x_1, \ldots, x_n) = \prod \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x_i - \mu)^2}{2\sigma^2}\right)$$

$$L(\mu, \sigma^2 | x_1, \ldots, x_n) = \sum \left[\ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right) - \frac{(x_i - \mu)^2}{2\sigma^2}\right]$$

$$= \sum \left[-\ln\sqrt{2\pi} - 0.5\ln\sigma^2 - \frac{(x_i - \mu)^2}{2\sigma^2}\right]$$

Prof. Dr. Dr. W. Rhode
Schätzen
Statistische Methoden der Datenanalyse

## Berechnung von Punktschätzern

- Beispiel Likelihood-Schätzer:
  Schätzen von Mittelwert und Varianz der Normalverteilung

Maxima bestimmen:

$$\frac{\partial l(\mu,\sigma^2)}{\partial \mu} = \sum \frac{2(x_i-\mu)}{2\sigma^2} \stackrel{!}{=} 0$$
$$\frac{\partial l(\mu,\sigma^2)}{\partial \sigma^2} = \sum \left[ -\frac{1}{2\sigma^2} + \frac{(x_i-\mu)^2}{2\sigma^4} \right] \stackrel{!}{=} 0$$





## Berechnung von Punktschätzern

- Beispiel Likelihood-Schätzer:
  Schätzen von Mittelwert und Varianz der Normalverteilung

Erhalte Gleichungssystem:

$$n\hat{\mu} = \sum x_i \Rightarrow \hat{\mu} = \bar{x}$$
$$n\hat{\sigma}^2 = \sum (x_i - \hat{\mu})^2$$
$$\Rightarrow \sigma^2 = \frac{1}{n}\sum (x_i - \bar{x})^2$$

→ Gleiche Schätzer wie bei Momenten-Methode!





## Kriterien zur Beurteilung der Schätzer

- Erwartungstreue:
  (Erwartungswert von Schätzer = Schätzer)

$$E(\hat{\theta}) = \theta$$

- Verzerrung (Bias):

$$B(\hat{\theta}) = E(\hat{\theta}) - \theta$$

- Mean Squared Error (MSE):

$$MSE(\hat{\theta}) = E\left((\hat{\theta}-\theta)^2\right) = B(\hat{\theta})^2 + Var\hat{\theta}$$





## Erwartungstreue

- Beispiel: Normalverteilung N~($\mu,\sigma^2$)

$$E(\hat{\mu}) = E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}n\mu = \mu$$

→Erwartungstreu bzw. unverzerrt!

## Erwartungstreue

- Beispiel: Normalverteilung N~($\mu$,$\sigma^2$)

$$E(\hat{\sigma}^2) = E(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2) = \frac{1}{n}E\left(\sum\left(X_i - \mu + \mu - \bar{X}\right)^2\right)$$
$$= \frac{1}{n}E\left(\sum\left((X_i-\mu)^2 - 2(X_i-\mu)(\bar{X}-\mu) + (\bar{X}-\mu)^2\right)\right)$$
$$= \frac{1}{n}E\left(\sum(X_i-\mu)^2 - 2\sum(X_i-\mu)(\bar{X}-\mu) + \sum(\bar{X}-\mu)^2\right)$$
$$= \frac{1}{n}E\left(\sum(X_i-\mu)^2 - 2n(\bar{X}-\mu)(\bar{X}-\mu) + n(\bar{X}-\mu)^2\right)$$
$$= \frac{1}{n}E\left(\sum(X_i-\mu)^2 - n(\bar{X}-\mu)^2\right)$$
$$= \frac{1}{n}\left(\sum E(X_i-\mu)^2 - nE(\bar{X}-\mu)^2\right)$$
$$= \frac{1}{n}\left(Var(X_i) - nVar(\bar{X})\right)$$
$$= \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n}\sigma^2$$

→ Verzerrt!





## Erwartungstreue

- Beispiel: Normalverteilung N~($\mu$,$\sigma^2$)

$$E(S^2) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = E\left(\frac{n}{n-1}\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)$$
$$= \frac{n}{n-1}E(\hat{\sigma}^2) = \sigma^2$$

→ Unverzerrt!

→ Empirische Stichprobenvarianz $S^2$





## Wo begegnen wir noch Schätzern?

- Minimierung von Abstandsmaßen

Beispiel: Wir suchen Schätzer für $\vec{a}$ in

$$f(x,\vec{a}) = a_1 f_1(x) + a_2 f_2(x) + \ldots$$

→ Methode der kleinsten Quadrate
→ Log-Likelihood-Methode





# Fitten

## Methode der kleinsten Quadrate

- Allgemeiner Fall:
  - Daten beschrieben durch n-dim. Vektor $\vec{y}(\vec{x})$
  - Verschiedene Standardabweichungen $\sigma$
  - Korrelation beschrieben durch Kovarianzmatrix *V*
- Minimierung der Summe der Abstände zwischen Messung und Modell = Minimierung der Summe der Abstände der Residuen $\Delta y$

$$S = \Delta\vec{y}^T V^{-1} \Delta\vec{y} \stackrel{!}{=} min.$$





## Methode der kleinsten Quadrate

- Häufige Anwendung:
  - Modell:

$$f(x, \vec{a})$$

- Messergebnisse sollen über Funktion *f* von Parametern $a_j$ abhängen
  - Minimierung von S
  - Bestimmung der Parameter
  - Suche nach funktionellen Zusammenhängen
  - Test, ob Form der Parametrisierung verträglich mit Messdaten





## Lineare Modelle

- $f(x, \vec{a})$ hängt linear von Parametern $a_j$ ab

$$y(x) = f(x, \vec{a}) = a_1 f_1(x) + a_2 f_2(x) + ... + a_p f_p(x)$$

- Residuen $r_i$:

$$r_i = y_i - f(x_i, \vec{a})$$





## Lineare Modelle

- Erwartungswerte:

$$E[y_i] = f(x_i, \bar{\vec{a}}) = \bar{y}_i$$

$$\bar{\vec{a}}\text{: wahrer Wert von } \vec{a}$$

$$\text{Für } \bar{\vec{a}} = \vec{a} \text{ gilt:}$$

$$E[\vec{r}] = 0$$

$$E[\vec{r}^2] = Var[\vec{r}] = \sigma^2$$

→ Erwartungstreu
→ Endliche Varianz
→ Keine Annahme über Wahrscheinlichkeitsdichte notwendig

## Lineare Modelle

- Minimiere:

$$S = \sum_i r_i^2 = \sum_i [y_i - a_1 f_1(x_i) - a_2 f_2(x_i) - \ldots - a_p f_p(x_i)]^2$$

- Partielle Ableitungen nach $a_j$ müssen dazu verschwinden:

$$\frac{\partial S}{\partial a_1} = 2 \sum_i f_1(x_i)\, [a_1 f_1(x_i) + a_2 f_2(x_i) + \ldots + a_p f_p(x_i) - y_i] = 0$$
$$\frac{\partial S}{\partial a_2} = 2 \sum_i f_2(x_i)\, [a_1 f_1(x_i) + a_2 f_2(x_i) + \ldots + a_p f_p(x_i) - y_i] = 0$$
$$\vdots$$
u.s.w.





## Lineare Modelle

- Umschreiben in Normalgleichungen:

$$a_1 \sum_i f_1^2(x_i) + \ldots + a_p \sum_i f1_(x_i) f_p(x_i) = \sum_i y_i f_1(x_i)$$
$$a_1 \sum_i f_2(x_i) f_1(x_i) + \ldots + a_p \sum_i f2_(x_i) f_p(x_i) = \sum_i y_i f_2(x_i)$$
$$\vdots$$
u.s.w., $p$ Gleichungen





## Lineare Modelle

- Umschreiben in Matrixschreibweise

$$A = \begin{pmatrix} f_1(x_1) & f_2(x_1) & \cdots & f_p(x_1) \\ f_1(x_2) & f_2(x_2) & \cdots & f_p(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ f_1(x_n) & f_2(x_n) & \cdots & f_p(x_n) \end{pmatrix}$$

- *A* heißt auch **Design-Matrix**
- Vektor der Erwartungswerte: $A \cdot \vec{a}$





## Lineare Modelle

- Minimierungsbedingung:

$$(A^T A)\vec{a} = A^T \vec{y}$$

- Es folgt folgender Schätzer:

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

## Lineare Modelle

- Und die Kovarianzmatrix?
- $\hat{\vec{a}}$ ist eine lineare Transformation von $\vec{y}$

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

→ Fehlerfortpflanzung anlog zum vorherigen Kapitel

$$V\left[\hat{\vec{a}}\right] = (A^T A)^{-1} A^T V\left[\vec{y}\right] A (A^T A)^{-1}$$





## Lineare Modelle

- Und die Kovarianzmatrix?
- Ohne Korrelation und gleichen Varianzen:

$$V[\vec{y}] = \sigma^2 \mathbb{1}$$

$$V[\hat{\vec{a}}] = \sigma^2 (A^T A)^{-1}$$





## Lineare Modelle

- Quadratsumme der Residuen:

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

$$\text{einsetzen in:}$$

$$S = \vec{y}^T \vec{y} - 2\vec{a}^T A^T \vec{y} + \vec{a}^T A^T A \vec{a}$$

$$\rightarrow \quad \hat{S} = \vec{y}^T \vec{y} - 2\hat{\vec{a}}^T A^T \vec{y} + \hat{\vec{a}}^T A^T A (A^T A)^{-1} A^T \vec{y}$$

$$= \vec{y}^T \vec{y} - \hat{\vec{a}}^T A^T \vec{y}$$





## Lineare Modelle

- Quadratsumme der Residuen:
  - Summe der Residuen kann direkt berechnet werden
  - Vorsicht: Differenz großer Zahlen
  - Einzelbeiträge interessant: $E[\hat{S}] = \sigma^2 (n-p)$
  - Schätzung bei unbekannter Varianz: $\hat{\sigma}^2 = \dfrac{\hat{S}}{n-p}$

→ Gute Schätzung für große Werte

## Gewichtete Methode der kleinsten Quadraten

- Gegeben:
  - Bekannte Wahrscheinlichkeitsdichte $\hat{S}$
  - $\frac{\hat{S}}{\sigma^2}$ folgt $\chi^2$ -Verteilung mit (n-p) Freiheitsgraden
  - Gaußverteilte Messfehler





## Gewichtete Methode der kleinsten Quadraten

- Diagonale Kovarianzmatrix
  - Datenpunkte mit unterschiedlichen Genauigkeiten
  - Datenpunkte statistisch unabhängig (Kovarianz=0)

$$\Rightarrow V[\vec{y}] = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_n^2 \end{pmatrix}$$





## Gewichtete Methode der kleinsten Quadraten

- Diagonale Kovarianzmatrix
  - Einführung der Gewichtungsmatrix:

$$W[\vec{y}] = V^{-1}[\vec{y}] = \begin{pmatrix} 1/\sigma_1^2 & 0 & \cdots & 0 \\ 0 & 1/\sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1/\sigma_n^2 \end{pmatrix}$$

  - Quadratsumme der Residuen:

$$\begin{aligned} S &= \vec{r}^T W \vec{r} \\ &= (\vec{y} - A\vec{a})^T W[\vec{y}](\vec{y} - A\vec{a}) \end{aligned}$$





## Gewichtete Methode der kleinsten Quadraten

- Allgemeine Kovarianzmatrix
  - Datenpunkte mit unterschiedlichen Genauigkeiten
  - Datenpunkte korreliert untereinander
  - Kovarianzmatrix nicht mehr diagonal, aber symmetrisch

## Gewichtete Methode der kleinsten Quadraten

- Allgemeine Kovarianzmatrix
  - Lineare Algebra: Zu jeder symmetrischen Matrix gibt es orthogonale Matrix U, die die symmetrische Matrix in eine Diagonalmatrix transformiert
  - Ansatz: $\vec{z} = U^T \vec{y}$
  - Fehlerfortpflanzung: $V[\vec{z}] = U^T V[\vec{y}] U$

$$\Rightarrow S = (\vec{z} - U^T A\vec{a})^T \underbrace{W[\vec{z}]}_{\text{diagonal}} (\vec{z} - U^T A\vec{a})$$



technische universität dortmund
Experimentelle Physik Vb
Astroteilchenphysik

## Gewichtete Methode der kleinsten Quadraten

- Allgemeine Kovarianzmatrix
  - Vergleiche mit Gleichung von unkorrelierten Datenpunkten:

$$S = (\vec{y} - A\vec{a})^T W[\vec{y}](\vec{y} - A\vec{a})$$

$$S = (\vec{z} - U^T A\vec{a})^T \underbrace{W[\vec{z}]}_{\text{diagonal}} (\vec{z} - U^T A\vec{a})$$

$$\Rightarrow UW[\vec{z}]U^T = V^{-1}[\vec{z}] = W[\vec{y}]$$

  - Allgemeine Gleichung:

$$S = (\vec{y} - A\vec{a})^T W[\vec{y}](\vec{y} - A\vec{a})$$



technische universität dortmund
Experimentelle Physik Vb
Astroteilchenphysik

## Gewichtete Methode der kleinsten Quadrate

- Allgemeine Lösung:

$$\begin{aligned} \hat{\vec{a}} &= (A^T W A)^{-1} A^T W \vec{y} \\ V[\hat{\vec{a}}] &= (A^T W A)^{-1} \end{aligned}$$

- Quadratsumme der Residuen:

$$\begin{aligned} \hat{S} &= \vec{y}^T W \vec{y} - \hat{\vec{a}}^T A^T W \vec{y} \\ E[\hat{S}] &= n - p \qquad \text{"freie Parameter"} \end{aligned}$$



technische universität dortmund
Experimentelle Physik Vb
Astroteilchenphysik

## Beispiel: Geraden-Anpassung

$$y = f(x, a_1, a_2) = a_1 + a_2 x$$

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

- Unkorrelierte Messwerte

→ Kovarianzmatrix und Gewichtungsmatrix diagonal

## Beispiel: Geraden-Anpassung

$$A^T W A = \begin{pmatrix} \sum_i W_i & \sum_i W_i x_i \\ \sum_i W_i x_i & \sum_i W_i x_i^2 \end{pmatrix} = \begin{pmatrix} S_1 & S_x \\ S_x & S_{xx} \end{pmatrix}$$

$$A^T W \vec{y} = \begin{pmatrix} \sum_i W_i y_i \\ \sum_i W_i x_i y_i \end{pmatrix} = \begin{pmatrix} S_y \\ S_{xy} \end{pmatrix}$$





## Beispiel: Geraden-Anpassung

$$(A^T W A)^{-1} = \frac{1}{D} \begin{pmatrix} S_{xx} & -S_x \\ -S_x & S_1 \end{pmatrix}$$

$$D = S_1 S_{xx} - S_x^2$$

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





## Beispiel: Geraden-Anpassung

$$\Rightarrow \quad \begin{aligned} \hat{a}_1 &= (S_{xx} S_y - S_x S_{xy}) D \\ \hat{a}_2 &= (-S_x S_y + S_1 S_{xy}) D \end{aligned}$$





## Beispiel: Geraden-Anpassung

- Kovarianzmatrix und Quadratsumme der Residuen

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

$$\hat{S} = S_{yy} - \hat{a}_1 S_y - \hat{a}_2 S_{xy}$$

## Gewichtete Methode der kleinsten Quadrate

- Sonderfall: Fehler in beiden Variablen

$$y = a_1 + a_2 x$$

$$x_i \text{ mit } \sigma_{x_i}$$

$$y_i \text{ mit } \sigma_{y_i}$$





## Gewichtete Methode der kleinsten Quadrate

- Sonderfall: Fehler in beiden Variablen
  - Minimiere:

$$S(a_1, a_2) = \sum_i \frac{(y_i - a_1 - a_2 x)^2}{\sigma_{y_i}^2 + a_2^2 \sigma_{x_i}^2}$$

  - Gefordert:

$$\frac{\partial S}{\partial a_1} = 0 \qquad \text{und} \qquad \frac{\partial S}{\partial a_2} = 0$$





## Gewichtete Methode der kleinsten Quadrate

- Sonderfall: Fehler in beiden Variablen

→ Anwendung von Optimierungsmethoden

→ Numerische Methoden wie z.B. Variation von $a_2$ und Berechnung von

$$\hat{a}_1 = \frac{\sum_i \frac{y_i}{\sigma_{y_i}^2 + a_2^2 \sigma_{x_i}^2} - \sum_i \frac{x_i}{\sigma_{y_i}^2 + a_2^2 \sigma_{x_i}^2}}{\sum_i 1/(\sigma_{y_i}^3 + a_2^2 \sigma_{x_i}^2)}$$





## Gewichtete Methode der kleinsten Quadrate

- Sonderfall: Lineare Regression

→ Minimierung der Quadratsumme der **senkrechten** Abstände zwischen Datenpunkten und Gerade

- Im Falle von gleichen Standardabweichungen:

$$S = \sum_i \frac{(y_i - a_1 - a_2 x_i)^2}{(1 + a_2^2)\sigma^2}$$

## Gewichtete Methode der kleinsten Quadrate

- Sonderfall: Lineare Regression

$$\begin{aligned}
\Rightarrow \hat{a}_1 &= \bar{y} - \hat{a}_2\bar{x} \\
\hat{a}_2 &= q \pm \sqrt{q^2+q} \\
\bar{y} &= \sum_i \frac{y_i}{n} \quad , \quad \bar{x} = \sum_i \frac{x_i}{n} \\
q &= \frac{\sum_i (y_i-\bar{y})^2 - \sum_i (x_i-\bar{x})^2}{2\sum_i (y_i-\bar{y})(x_i-\bar{x})}
\end{aligned}$$

- Vorzeichen testen!





## Methode der kleinsten Quadrate

- Nichtlineare Modelle:
  - $f(x,\vec{a})$ hängt nicht linear ab von $a_i$
  - Bsp.: $f(ax,\vec{a}) = a_1 \cdot \exp(a_2 x)$
  - $\frac{\partial f}{\partial a_1}$ und $\frac{\partial f}{\partial a_2}$ hängen von den Parametern $a_i$ ab





## Nichtlineare Modelle

- Linearisierung: Taylor-Entwicklung

$$f(x,\vec{a}) = f(x,\vec{a}^*) + \sum_{j=1}^{p} \underbrace{\frac{\partial f}{\partial a_j}(a_j - a_j^*)}_{\text{Ableitungen an der Stelle } a^*}$$





## Nichtlineare Modelle

- Korrekturterm:

$$\Delta\vec{a} = \vec{a} - \vec{a}^*$$

- Residuen in linearer Näherung:

$$\vec{r} = \vec{y} - A\Delta\vec{a} - \vec{f}$$

- Jacobi-Matrix mit dem Näherungsvektor $\vec{f}$ an der Stelle $\vec{a}^*$:

$$A = \begin{pmatrix} \partial f(x_1)/\partial a_1 & \partial f(x_1)/\partial a_2 & \cdots & \partial f(x_1)/\partial a_p \\ \partial f(x_2)/\partial a_1 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \partial f(x_n)/\partial a_1 & \cdots & \cdots & \partial f(x_n)/\partial a_p \end{pmatrix}$$

## Nichtlineare Modelle

- Quadratsumme Residuen:

$$\begin{aligned} S &= \vec{r}^T W \vec{r} \\ &= (\vec{y} - A\Delta\vec{a} - \vec{f})^T W (\vec{y} - A\Delta\vec{a} - \vec{f}) \end{aligned}$$

- Normalgleichung:

$$\begin{aligned} (A^T W A)\Delta\vec{a} &= A^T W(\vec{y} - \vec{f}) \\ \Delta\vec{a} &= (A^T W A)^{-1} A^T W (\vec{y} - \vec{f}) \end{aligned}$$





## Nichtlineare Modelle

- Näherungsweise weiter gültig:

$$\begin{aligned} \hat{a} &= (A^T W A)^{-1} A^T W y \\ V[\hat{a}] &= (A^T W A)^{-1} \\ E[\hat{S}] &= n - p \end{aligned}$$





## Nichtlineare Modelle

- <u>Konvergenz</u>
- Hoffnung:

$$S(\vec{a}^* + \Delta\vec{a}) < S(\vec{a}^*)$$

- Sicher:

$$S(\vec{a}^* + \lambda\Delta\vec{a}) < S(\vec{a}^*)$$

→ Suche mit Salami-Taktik nach geeignetem λ





## Nichtlineare Modelle

- <u>Konvergenzkriterium</u>
- Betrachte

$$\Delta S = \Delta\vec{a}^T W (\vec{y} - \vec{f})$$
$$\text{für } \Delta S < 1 \text{ Abweichung innerhalb } 1\sigma$$

- Beende, wenn z.B.

$$\Delta S < 0,1$$

→ Statistische Fehler überwiegen dann

## Maximum-Likelihood-Methode

- **Likelihood**-Funktion:

$$L(\vec{a}) = f(\vec{x}_1|\vec{a}) \cdot f(\vec{x}_2|\vec{a}) \cdots f(\vec{x}_n|\vec{a}) = \prod_{i=1}^{n} f(\vec{x}_i|\vec{a})$$

→ Maß für Wahrscheinlichkeit bei festem $\vec{a}$ Datensatz $\vec{x}_1, ..., \vec{x}_n$ zu messen

→ <u>Keine</u> Wahrscheinlichkeitsdichte!



technische universität dortmund
Experimentelle Physik Vb
Astroteilchenphysik

## Maximum-Likelihood-Methode

- Maximum-Likelihood-Prinzip:
- Beste Schätzung von $\vec{a}$ maximiert $L(\vec{a})$
- Normierung notwendig: $\int f(\vec{x}|\vec{a})\, d\vec{x} = 1 \qquad \forall \vec{a}$

→ Numerischer Aufwand

- Maximum durch Differenzieren:

$$\frac{dL(\vec{a})}{d\vec{a}} = 0 \qquad \frac{\partial L(\vec{a})}{\partial a_i} = 0 \text{ für } i = 1 \ldots k$$

Prof. Dr. Dr. W. Rhode
Schätzen
Statistische Methoden der Datenanalyse

technische universität dortmund
Experimentelle Physik Vb
Astroteilchenphysik

## Maximum-Likelihood-Methode

- **Log-Likelihood**-Funktion:

$$l(\vec{a}) = \ln(L(\vec{a})) = \sum_{i=1}^{n} \ln(f(\vec{x}_i)|\vec{a})$$

- Logarithmus ist monoton
  → Maximum von $l(\vec{a})$ und $L(\vec{a})$ an der gleichen Stelle
- Vereinfachung der Gleichung durch Logarithmen-Regeln
- Oft: Minimierung der negativen Likelihood: $F(\vec{a}) = -l(\vec{a})$
  → Viele Algorithmen zu Minimierungsproblemen existieren...

Prof. Dr. Dr. W. Rhode
Schätzen
Statistische Methoden der Datenanalyse

technische universität dortmund
Experimentelle Physik Vb
Astroteilchenphysik

## Maximum-Likelihood-Methode

- Eigenschaften:

\+ konsistent
\+ nicht immer erwartungstreu
\+ beides für $n \to \infty$
\+ effizient

\- Großer Rechenaufwand
\- a priori Kenntnis der Wahrscheinlichkeitsdichte notwendig
\- Überprüfung der Verträglichkeit zwischen Daten und Parameter
(bei mehreren Dimensionen in allen Teilintervallen!)

Prof. Dr. Dr. W. Rhode
Schätzen
Statistische Methoden der Datenanalyse

## Beispiel: Zerfallsintervallverteilung eines Teilchens

- Modell:

$$f(x|a) = \underbrace{\frac{1}{2}(1+ax)}_{\text{zufällig normiert! s.u.}} \qquad \text{mit } x = \cos\vartheta$$





## Beispiel: Zerfallsintervallverteilung eines Teilchens

- Normierung:

$$\begin{aligned}\int_{-1}^{1} \frac{1}{2}(1+ax)\,dx &= \left.\frac{1}{2}x + \frac{1}{4}ax^2\right|_{-1}^{1} \\ &= \frac{1}{2}1 + \frac{1}{2}1 + \frac{1}{4}a - \frac{1}{4}a = 1\end{aligned}$$





## Beispiel: Zerfallsintervallverteilung eines Teilchens

- $n$ Werte für $x_i$ gemessen
- $\hat{a}$ gesucht
- $-1 \leq x \leq 1$

$$\Rightarrow F(a) = -\sum_{i=1}^{n} \ln\left(\frac{1}{2}(1+ax_i)\right)$$





## Beispiel: Gauß-Verteilung

$$\begin{aligned}f(x_i|a) &= \frac{1}{\sqrt{a\pi\sigma_i}} e^{-\frac{(x_i-a)^2}{a\sigma_i^2}} \\ \Rightarrow F(a) &= \text{konst} + \frac{1}{2}\sum_{i=1}^{n} \frac{(x_i-a)^2}{\sigma_i^2}\end{aligned}$$

## Beispiel: Gauß-Verteilung

$$\frac{dF(a)}{da} \overset{!}{=} 0 \qquad \Rightarrow \qquad -\sum_{i=1}^{n} \frac{x_i - a}{\sigma_i^2} = 0$$

$$\hat{a} = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i} \qquad (\text{mit } w_i = 1/\sigma_i^2)$$

→ gewichtetes Mittel aller $x_i$

→ Gleiches Ergebnis wie bei Methode der kleinsten Quadrate





## Fehlerbestimmung bei Maximum-Likelihood-Methode

- **Ein** Parameter

- Betrachte Zeichnung *F(a)* gegen *a*
- Schnittpunkte mit folgender Geraden markieren den 1σ-Bereich:

$$F = F_{min} + \frac{1}{2}$$





## Fehlerbestimmung bei Maximum-Likelihood-Methode

- **Mehrere** Parameter

- Entwicklung der negativen Likelihood:

$$\begin{aligned} F(\vec{a}) &= F(\hat{\vec{a}}) + \frac{1}{2}\sum_{i,k}^{n} \frac{\partial^2 F}{\partial a_i \partial a_k}(a_i - \hat{a}_i)(a_k - \hat{a}_k) + \ldots \\ &= F(\hat{\vec{a}}) + \frac{1}{2}\sum_{i,k}^{n} G_{ik}(a_i - \hat{a}_i)(a_k - \hat{a}_k) + \ldots \\ G_{ik} &= \frac{\partial^2 F}{\partial a_i \partial a_k} \qquad V = G^{-1} = \text{Kovarianzmatrix} \end{aligned}$$





## Fehlerbestimmung bei Maximum-Likelihood-Methode

- **Zwei** Parameter

- Konturlinien als Linien gleicher Likelihood

$$F(\vec{a}) = F(\hat{\vec{a}} + \frac{1}{2}r^2 \leftrightarrow F(\hat{\vec{a}}) + \frac{1}{2}$$

- Abweichung vom asymptotischen Verhalten
  → asymmetrische Fehler

## Maximum-Likelihood-Methode

- Eigenschaften: Konsistenz und Erwartungstreue
  - Konsistenz: Es kann gezeigt werden (siehe Blobel), dass

$$\lim_{n\to\infty} \hat{a} = a_0$$

  - Erwartungstreu bei symmetrischer Likelihood im Intervall [-pσ, pσ]
  - Asymptotisch erwartungstreu bei asymmetrischer Likelihood





## Maximum-Likelihood-Methode

- Eigenschaften: Varianz

$$\sigma(\hat{a}) = \left( \left. \frac{d^2 F}{da^2} \right|_{\hat{a}} \right)^{-1/2}$$

→ kleinstmögliche Varianz

→ asymptotisch effizient





# Parameterschätzung mit Bayes Theorem





## Bayes Theorem

$$p(H_i|D,I) = \frac{p(H_i|I)p(D|H_i,I)}{p(D|I)}$$

- $H_i$ Hypothese
- $I$ Prior
- $D$ Daten

## Bayes Theorem

$$p(H_i|D,I) = \frac{p(H_i|I)p(D|H_i,I)}{p(D|I)}$$

- $p(D|H_i,I)$ Wahrscheinlichkeit die Daten D zu messen, wenn $H_i$ und $I$ wahr sind (auch Likelihood-Funktion $\mathcal{L}(H_i|D)$)
- $p(H_i|I)$ Prior p.d.f. der Hypothese $H_i$ (Information vor der Beobachtung)
- $p(H_i|D,I)$ Posterior p.d.f. der Hypothese $H_i$
- $p(D|I) = \sum_i p(H_i|I)p(D|H_i,I)$ Normierung des Posteriors
- Im kontinuierlichen Fall gehen alle Summen in Integrale über



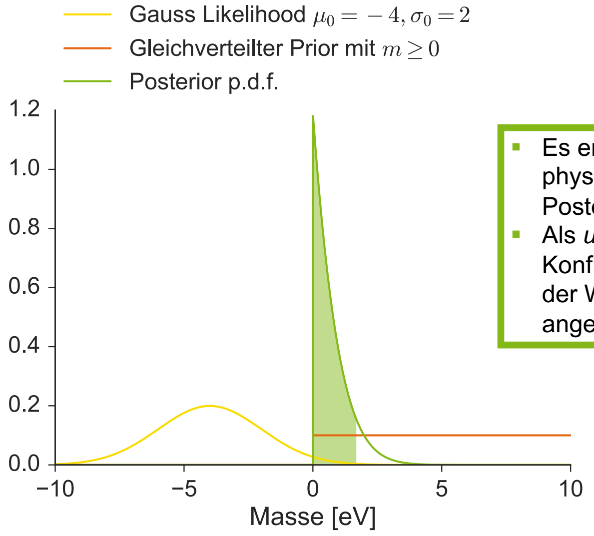


## Ein Beispiel – Messung der Neutrinomasse

- Messung der Neutrinomasse über den Zerfall von Neutronen
- Da die Neutrinomasse sehr klein ist, wäre es nicht allzu verwunderlich wenn ein Experiment einen negativen Wert messen würde
- Gauß Annahme für die Likelihood: $p(x|\mu,\sigma) = N(x|\mu,\sigma)$
- Prior p.d.f. wird durch eine Gleichverteilung für $m \geq 0$ beschrieben, da Massen nur positive Werte annehmen können $p(m) = \begin{cases} \text{const.}, & m \geq 0 \\ 0, & m < 0 \end{cases}$
- Die Messung der Neutrinomasse liefert: $x_0 = -4\,\text{eV}$ mit einer bekannten Standardabweichung von $\sigma_0 = 2\,\text{eV}$








- Es ergibt sich eine physikalisch sinnvolle Posterior p.d.f.
- Als *upper limit* kann mit einer Konfidenz von $\alpha = 0.9$ der Wert $m_{\text{ul}} \simeq 1.68\,\text{eV}$ angegeben werden





## Bayesische Statistik

- Bayesische Wahrscheinlichkeitsdichte:
  - Messung des Zustandes unseres Wissens über den Wert eines Parameters
  - Der Parameter hat jedoch einen FESTEN, aber unbekannten Wert
- Bayesische Statistik kann als ein Lernprozess verstanden: Wenn neue Daten genommen werden, kann der alte Posterior als neuer Prior verwendet werden.

## Parameterschätzung

- Verwendung von Bayes Theorem, um aus Daten etwas über Werte von Parametern zu lernen
- Problem: Unter der Annahme eines Modells mit dem Parameter $\theta$ wird dessen p.d.f. gesucht
- Gegeben: Gemessene Daten *x*, sowie ein Prior $p(\theta)$
- Die Posterior p.d.f. lässt sich dann direkt berechnen:

$$p(\theta|x) = \frac{p(\theta)p(x|\theta)}{p(x)} = \frac{p(\theta)p(x|\theta)}{\int \mathrm{d}\theta p(\theta)p(x|\theta)}$$





## Parameterschätzung

- Oft ist es nützlich die Posterior p.d.f. mit einer ausgewählten Parameterkonfiguration („best-fit") und einer Konfidenzregion zusammenfassend zu beschreiben
- Für den „best-fit" der p.d.f. bieten sich der Modus (wahrscheinlichster Wert) oder der Mittelwert der Verteilung an

$$\langle\theta\rangle = \int \mathrm{d}\theta\, \theta p(\theta|x)$$

- Sind Modus und Mittelwert sehr unterschiedlich ist die p.d.f. sehr asymmetrisch und es sollte auf eine Zusammenfassung verzichtet werden
- → Auf bayesische Konfidenzintervalle wird im Kapitel Testen eingegangen





# Regularisierung





## Grundlage

- **Korrekt gestellte Probleme** (nach J. Hadamard):
  1. **Existenz:** Problem hat eine Lösung
  2. **Eindeutigkeit:** Lösung ist eindeutig bestimmt
  3. **Stabilität:** Lösung hängt stetig von den Eingangsdaten ab
- Ansonsten: **inkorrekt/schlecht gestelltes Problem**
- Aufgrund endlicher Präzision beim Finden der Lösung und fehlerbehafteter Eingangsdaten kann ein korrekt gestelltes Problem unter Umständen instabil werden → **schlecht konditioniert**

## Regularisierung

- Anwendungen:
  - Zur Lösung von schlecht gestellten Problemen
  - Zur Vermeidung von Überanpassung (overfitting)
- Regularisierung = Einbringen zusätzlicher Informationen
- Zusätzliche Informationen über Strafterme (Kostenterme) oder Einschränkungen des Parameterbereiches
- Verwendung oft im Bereich von Inversen Problemen (→ Kapitel „Entfaltung“) oder von maschinellem Lernen





## Tikhonov-Regularisierung

- Approximation einer positiv definiten (und somit invertierbaren) Matrix für die positiv semi-definite Matrix (auch Ridge-Regression genannt)
- Lineare Modelle: mittels L2-Norm
- Nicht lineare Modelle: mittels zweiter Ableitung
- Beispiel: Methode der kleinsten Quadrate mit Tikhonov-Regularisierung mittels der L2-Norm





## Lineare Modelle

Erinnerung!

- Minimierungsbedingung:

$$(A^T A)\vec{a} = A^T \vec{y}$$

- Es folgt folgender Schätzer:

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





## Lineare Modelle

Problem: $(A^\top A)$ kann singulär
oder
$(A^\top A)$ schlecht konditioniert
(annähernd singulär) sein

- Minimierungsbedingung:

$$\boxed{(A^T A)}\vec{a} = A^T \vec{y}$$

- Es folgt folgender Schätzer:

$$\hat{\vec{a}} = \cancel{(A^T A)^{-1}} A^T \vec{y}$$

Inverse Matrix kann nicht berechnet werden

## Regularisierte Methode der kleinsten Quadrate

- Für die Lösung muss S minimiert werden
- S wird von:

$$S = (A\vec{a} - \vec{y})^\top (A\vec{a} - \vec{y})$$

zu:

$$S^{\mathrm{reg}} = (A\vec{a} - \vec{y})^\top (A\vec{a} - \vec{y}) + (\Gamma\vec{a})^\top (\Gamma\vec{a})$$





## Regularisierte Methode der kleinsten Quadrate

- Für die Lösung muss S minimiert werden
- S wird von:

$$S = (A\vec{a} - \vec{y})^\top (A\vec{a} - \vec{y})$$

Regularisierung-matrix

zu:

$$S^{\mathrm{reg}} = (A\vec{a} - \vec{y})^\top (A\vec{a} - \vec{y}) + (\boxed{\Gamma}\vec{a})^\top (\boxed{\Gamma}\vec{a})$$





## Regularisierte Methode der kleinsten Quadrate

- Für die Lösung muss S minimiert werden
- S wird von:

$$S = (A\vec{a} - \vec{y})^\top (A\vec{a} - \vec{y})$$

zu:

$$S^{\mathrm{reg}} = (A\vec{a} - \vec{y})^\top (A\vec{a} - \vec{y}) + (\Gamma\vec{a})^\top (\Gamma\vec{a})$$

- Der regularisierte Schätzer wird zu:

$$\vec{a}^{\mathrm{reg}} = \left(A^\top A + \Gamma^\top \Gamma\right)^{-1} A^\top \vec{y}$$





## Regularisierte Methode der kleinsten Quadrate

- Für die Lösung muss S minimiert werden
- S wird von:

$$S = (A\vec{a} - \vec{y})^\top (A\vec{a} - \vec{y})$$

zu:

$$S^{\mathrm{reg}} = (A\vec{a} - \vec{y})^\top (A\vec{a} - \vec{y}) + (\Gamma\vec{a})^\top (\Gamma\vec{a})$$

- Der regularisierte Schätzer wird zu:

$$\vec{a}^{\mathrm{reg}} = \boxed{\left(A^\top A + \Gamma^\top \Gamma\right)}^{-1} A^\top \vec{y}$$

Auch invertierbar, wenn $(A^\top A)$ singulär oder schlecht konditioniert ist

## Regularisierte Methode der kleinsten Quadrate

- Die Art der Regularisierung hängt von der Wahl von $\Gamma$ ab
- Häufig wird $\Gamma = \sqrt{\lambda}\underline{\underline{1}}$ als ein Vielfaches der Einheitsmatrix gewählt
  - Der zu minimierende Audruck wird damit zu:

$$S^{\text{reg}} = (A\vec{a} - \vec{y})^\top (A\vec{a} - \vec{y}) + \lambda \vec{a}^\top \vec{a}$$

- Der regularisierte Schätzer wird zu:

$$\vec{a}^{reg} = (A^\top A + \lambda \underline{\underline{1}})^{-1} A^\top \vec{y}$$





## Regularisierte Methode der kleinsten Quadrate

- Die Art der Regularisierung hängt von der Wahl von $\Gamma$ ab
- Häufig wird $\Gamma = \sqrt{\lambda}\underline{\underline{1}}$ als ein Vielfaches der Einheitsmatrix gewählt
  - Der zu minimierende Audruck wird damit zu:

$$S^{\text{reg}} = (A\vec{a} - \vec{y})^\top (A\vec{a} - \vec{y}) + \lambda \underbrace{\boxed{\vec{a}^\top \vec{a}}}_{\text{L2-Norm}}$$

- Der regularisierte Schätzer wird zu:

$$\vec{a}^{reg} = (A^\top A + \lambda \underline{\underline{1}})^{-1} A^\top \vec{y}$$





## Regularisierte Methode der kleinsten Quadrate

- Die Art der Regularisierung hängt von der Wahl von $\Gamma$ ab
- Häufig wird $\Gamma = \sqrt{\lambda}\underline{\underline{1}}$ als ein Vielfaches der Einheitsmatrix gewählt
  - Der zu minimierende Audruck wird damit zu:

$$S^{\text{reg}} = (A\vec{a} - \vec{y})^\top (A\vec{a} - \vec{y}) + \underbrace{\boxed{\lambda}}_{\text{Steuerung der Regularisierungsstärke}} \vec{a}^\top \vec{a}$$

- Der regularisierte Schätzer wird zu:

$$\vec{a}^{reg} = (A^\top A + \lambda \underline{\underline{1}})^{-1} A^\top \vec{y}$$



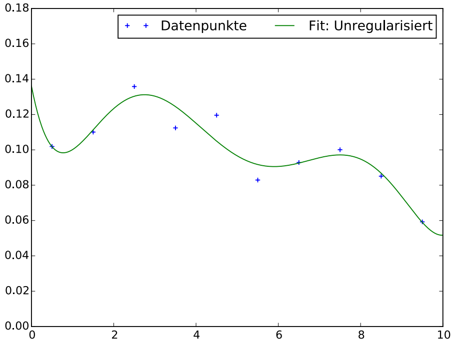


## Regularisierte Methode der kleinsten Quadrate

- Die Art der Regularisierung hängt von der Wahl von $\Gamma$ ab
- Eine andere Möglichkeit ist: $\Gamma = \sqrt{\lambda} C A$ mit $C = \begin{pmatrix} -1 & 1 & 0 & 0 & \dots & & \\ 1 & -2 & 1 & 0 & \dots & & \\ 0 & 1 & -2 & 1 & \dots & & \\ & \dots & & & \dots & & \\ & \dots & & 1 & -2 & 1 & \\ & \dots & & & 1 & -1 & \end{pmatrix}$
  - Mit diesem $\Gamma$ wird die numerischen zweite Ableitung der gefitteten Funktion für die Regularisierung genutzt
- Der regularisierte Schätzer wird zu:

$$\vec{a}^{\text{reg}} = \left(A^\top A + \lambda (CA)^\top (CA)\right)^{-1} A^\top \vec{y}$$

## Beispiel:




$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$$

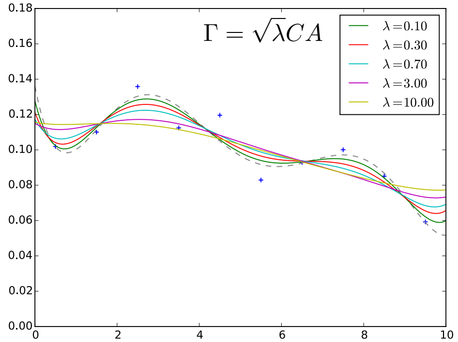
| | $a_0$ | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ |
|---|---|---|---|---|---|---|
| unregularisiert | 0.1361 | -0.1194 | 0.1233 | -0.0476 | 0.0084 | -0.0007 |

- Terme hoher Ordnung führen zu „Oszillationen" der Lösung





## Beispiel:




$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$$

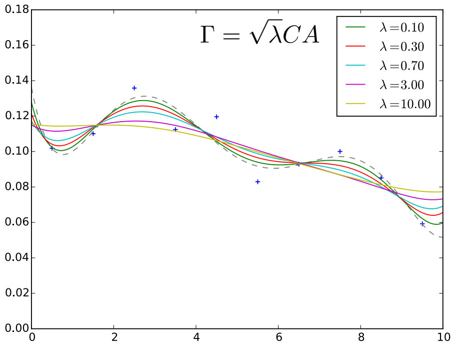
| | $a_0$ | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ |
|---|---|---|---|---|---|---|
| unregularisiert | 0.1361 | -0.1194 | 0.1233 | -0.0476 | 0.0084 | -0.0007 |
| $\lambda = 0.1$ | 0.1277 | -0.0908 | 0.0981 | -0.0386 | 0.0069 | -0.0006 |
| $\lambda = 0.3$ | 0.1212 | -0.0634 | 0.0715 | -0.0285 | 0.0051 | -0.0004 |
| $\lambda = 0.7$ | 0.1172 | -0.0406 | 0.0473 | -0.0189 | 0.0034 | -0.0003 |
| $\lambda = 3.0$ | 0.1150 | -0.0136 | 0.0164 | -0.0065 | 0.0011 | -0.0001 |
| $\lambda = 10.0$ | 0.1156 | -0.0045 | 0.0052 | -0.0022 | 0.0004 | -0.0000 |

- Die Regularisierung über die 2. Ableitung → Glättung des Ergebnisses
- Oszillationen der Lösung verschwinden





## Beispiel:




$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + a_5 x^5$$

| | $a_0$ | $a_1$ | $a_2$ | $a_3$ | $a_4$ | $a_5$ |
|---|---|---|---|---|---|---|
| unregularisiert | 0.1361 | -0.1194 | 0.1233 | -0.0476 | 0.0084 | -0.0007 |
| $\lambda = 0.1$ | 0.1277 | -0.0908 | 0.0981 | -0.0386 | 0.0069 | -0.0006 |
| $\lambda = 0.3$ | 0.1212 | -0.0634 | 0.0715 | -0.0285 | 0.0051 | -0.0004 |
| $\lambda = 0.7$ | 0.1172 | -0.0406 | 0.0473 | -0.0189 | 0.0034 | -0.0003 |
| $\lambda = 3.0$ | 0.1150 | -0.0136 | 0.0164 | -0.0065 | 0.0011 | -0.0001 |
| $\lambda = 10.0$ | 0.1156 | -0.0045 | 0.0052 | -0.0022 | 0.0004 | -0.0000 |

Mit stärkerer Regularisierung werden die Koeffizienten hoher Ordnung kleiner

- Die Regularisierung über die 2. Ableitung → Glättung des Ergebnisses
- Oszillationen der Lösung verschwinden