

Schätzen und Regression

- Die zeitliche Abhängigkeit der Sensordaten spielt nun keine Rolle mehr.
- Nach Aufzeichnung der Sensordaten sind wir z.B. daran interessiert, **funktionale Zusammenhänge** zwischen physikalischen Größen zu erkennen und zu formulieren.
- Hierzu dient z.B. die **Ausgleichsrechnung (Regression)**.
- Unbekannte Parameter eines Modells sollen geschätzt werden.
- Hierzu werden üblicherweise **Optimierungsmethoden** eingesetzt.

Beispiel: Optimale Schätzung für Punktdaten (1)

- Zur Identifikation eines unbekannten Punktes $p = (x, y) \in \mathbb{R}^2$ wurden n -Messungen vorgenommen.
- Die Messpunkte seien $p_i = (x_i, y_i)$.
- Was ist eine sinnvolle Schätzung für p ?
- intuitiv der Schwerpunkt: $(\frac{1}{n} \sum_{i=1}^n x_i, \frac{1}{n} \sum_{i=1}^n y_i)$
- Wie kann man dies **mathematisch begründen**?

Beispiel: Optimale Schätzung für Punktdaten (2)

- Fehlerfunktion:

$$\text{err}(p) = \text{err}(x, y) = \sum_{i=1}^n (x - x_i)^2 + (y - y_i)^2$$

- Minimiere $\text{err}(x, y)$
- Herleitung 
- Methode der kleinsten Quadrate

Lineare Regression

- Wir vermuten zwischen den Größen x und y einen linearen Zusammenhang der Form

$$y = m \cdot x + b$$

- Wir haben Messwerte (x_i, y_i) für $i = 1, \dots, n$ vorliegen.
- Wie sollen wir m und b wählen, so dass die entstehende Funktion möglichst gut zu den Messwerten passt?
- Fehlerfunktion:

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

- Herleitung

- Ergebnisse:

$$\begin{aligned} m &= \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \\ &= \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n(\bar{x})^2} \\ &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ b &= \frac{1}{n} \left(\sum_{i=1}^n y_i - m \sum_{i=1}^n x_i \right) \\ &= \bar{y} - m \cdot \bar{x} \end{aligned}$$

Methode der kleinsten Quadrate

- Die durchgeführten Optimierungen waren Beispiele für die **Methode der kleinsten Quadrate**.
- Prinzipiell anwendbar auf beliebige funktionale Modelle

$$y = f(x; \lambda_1, \dots, \lambda_n)$$

- Minimiere

$$\sum_{i=1}^n (f(x_i; \lambda_1, \dots, \lambda_n) - y_i)^2$$

für die λ_i .

- Durch die Gradientenbildung und die notwendige Bedingung

$$\nabla f = 0$$

für ein lokales Minimum entsteht ein lineares Gleichungssystem mit n Gleichungen und den Variablen $\lambda_1, \dots, \lambda_n$.

Parameterschätzer

- Statt von einer Fehlerfunktion können wir bei verrauschten Daten auch von einem **statistischen Modell** ausgehen, das Parameter enthält.
- Als Datenquelle haben wir Daten von Messungen, also Realisationen des statistischen Modells, Ausprägungen von Zufallsvariablen.
- Das Experiment haben wir mehrmals unter gleichen Bedingungen und unabhängig voneinander durchgeführt. Die Messergebnisse sind also Realisationen von unabhängigen und identisch verteilten Zufallsvariablen.
- Wir wollen einen Parameter p (oder auch mehrere Parameter) des statistischen Modells schätzen.

- Beispiel: Wir gehen im Modell davon aus, dass die zu messende Größe exponentialverteilt ist, mit Dichte:

$$f(x) = \lambda e^{-\lambda x} \cdot I_{[0, \infty]}(x)$$

- Wie sollen wir λ wählen, so dass unsere Auswahl möglichst gut zu unseren Messergebnissen passt?
- mögliche Ansätze: Momentenmethode, Maximum Likelihood Schätzer

Momente einer Verteilung

- Erwartungswert und Varianz sind Spezialisierungen allgemeiner Größen einer Verteilung, den **Momenten**.
- Die Größe $E(X^k)$ heißt **k-tes Moment** der Zufallsvariablen X .
- Für diskrete Verteilungen X mit Wahrscheinlichkeitsfunktion f :

$$E(X^k) = \sum_i x_i^k f(x_i)$$

- Für stetige Verteilungen X mit Dichtefunktion f :

$$\mathbb{E}(X^k) = \int_{-\infty}^{\infty} x^k f(x) dx$$

- Der Erwartungswert μ entspricht dem 1. Moment.
- Die Größe $\mathbb{E}((X - \mu)^k)$ heißt *k-tes zentrales Moment*.
- Die Varianz σ^2 entspricht dem 2. zentralen Moment.

Momentenmethode

- Man drückt den zu schätzenden Parameter als Funktion der Momente einer Verteilung aus.
- Beispiel: Bei der Exponentialverteilung gilt $E(X) = \frac{1}{\lambda}$, also

$$\lambda = \frac{1}{E(X)}$$

- In der Funktion drücken wir nun die Momente durch die **Stichprobenmomente** aus.

- Das k-te Stichprobenmoment m_k zur Stichprobe $x_1, \dots, x_n$ ist

$$m_k = \frac{1}{n} \sum_{i=1}^n x_i^k$$

- Für die Exponentialverteilung ergibt sich die Schätzung

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

Maximum Likelihood Schätzer

- Unsere Messergebnisse sind Realisationen von unabhängigen und identisch verteilten Zufallsvariablen.
- Wir bilden die *gemeinsame Dichte/Wahrscheinlichkeitsfunktion* dieser Zufallsvariablen und wählen dann den (oder die) Parameter so, dass wir die *größtmögliche Wahrscheinlichkeit für unsere Messungen* erhalten.
- Nach Einsetzen der Messwerte in die gemeinsame Dichte/Wahrscheinlichkeitsfunktion hängt die Funktion nur noch von den zu schätzenden Parameter ab.
- Diese so entstandene Funktion nennen wir *Likelihood-Funktion*.

Beispiel (1): Maximum Likelihood Schätzer

- Für eine Münze wollen wir die Wahrscheinlichkeit schätzen, dass wir bei einem Münzwurf Zahl erhalten. Wir vermuten, dass die Münze nicht ideal ist.
- Wir führen ein Experiment durch und erhalten bei 20 Würfeln sechsmal Zahl.
- Die Wahrscheinlichkeit hierfür ist

$$P(X = 6) = p^6(1 - p)^{14} =: f(p),$$

$f(p)$ ist die Likelihood-Funktion.

- Wir wählen p so, dass $f(p)$ möglichst groß wird. Also

$$f'(p) = (6p^5(1-p)^{14} - 14p^6(1-p)^{13})$$

Aus $f'(p) = 0$ folgt

$$(6p^5(1-p)^{14} - 14p^6(1-p)^{13}) = 0$$

und daraus

$$6(1-p) - 14p = 0$$

also $p = \frac{3}{10}$.

- $p = \frac{3}{10}$ ist der **Maximum Likelihood-Schätzer** für die Schätzung von p bei den vorliegenden Messwerten.

Bemerkungen

- Es ist nicht notwendig, die originäre Likelihood-Funktion zu maximieren.
- Es sei $f(p)$ die Likelihood-Funktion. Wenn $g(p)$ eine streng monoton steigende Funktion ist, dann nimmt $g(f(p))$ an der gleichen Stelle ein Maximum an wie $f(p)$.
- So kann man z.B. auch $\log f(p)$ maximieren.

Beispiel (2): Maximum Likelihood Schätzer

- Eine Poisson-verteilte Zufallsvariable (mit unbekanntem Parameter μ) hat die Wahrscheinlichkeitsfunktion

$$f(x) = \frac{\mu^x}{x!} e^{-\mu}$$

- Für eine Stichprobe $x_1, \dots, x_n$ ergibt sich damit die Likelihood-Funktion

$$\begin{aligned} f(\mu) &= \prod_{i=1}^n \frac{\mu^{x_i}}{x_i!} e^{-\mu} \\ &= \frac{1}{x_1! \cdots x_n!} \cdot \mu^{x_1 + \cdots + x_n} \cdot e^{-n\mu} \end{aligned}$$

$$= \frac{1}{x_1! \cdots x_n!} \cdot \mu^{n\bar{x}} \cdot e^{-n\mu}$$

Der natürliche Logarithmus dieser Funktion ist

$$\log f(\mu) = -\log(x_1! \cdots x_n!) + n\bar{x} \log \mu - n\mu$$

Daraus ergibt sich

$$\frac{\partial}{\partial \mu} \log f(\mu) = \frac{n\bar{x}}{\mu} - n = 0$$

Damit lautet der Maximum-Likelihood-Schätzer

$$\frac{1}{n}(x_1 + \cdots + x_n) = \bar{x}$$

Beispiel (3): Maximum Likelihood Schätzer

Eine normalverteilte Zufallsvariable hat die Dichte

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

Also lautet die Likelihood-Funktion für die Normalverteilung

$$f(\mu, \sigma) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Durch Zusammenfassen ergibt sich

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

mit

$$h = \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

Nehmen wir den Logarithmus, so folgt

$$\log f(\mu, \sigma) = -n \log \sqrt{2\pi} - n \log \sigma - h$$

Hier sollen zwei Parameter geschätzt werden. Deshalb müssen wir für beide Parameter die partiellen Ableitungen bilden. Wir erhalten

$$\frac{\partial}{\partial \mu} \log f(\mu, \sigma) = -\frac{\partial}{\partial \mu} h = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0$$

und

$$\frac{\partial}{\partial \sigma} \log f(\mu, \sigma) = -\frac{n}{\sigma} - \frac{\partial}{\partial \sigma} h = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^n (x_i - \mu)^2 = 0$$

Aus der ersten Gleichung folgt

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

Eingesetzt in die zweite und aufgelöst ergibt sich

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

Damit sind $\bar{x}$ und $\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ die Maximum Likelihood Schätzer für μ und σ^2 bei der Normalverteilung.

Zusammenfassung

- Empfang von Sensordaten via `SensorManager` und `SensorEventListener`
- Eventuell Integration oder Ableitung der Sensordaten notwendig
- Daten glätten mit Filtern, auf effiziente Berechnung achten
- Daten auswerten mit Methode der kleinsten Quadrate oder Parameterschätzungen
- Methode zur Konstruktion von Parameterschätzern: Maximum Likelihood