Tobias Lang
Mathias Schickel

Andreas Schilling
Sommersemester 2016

Übungsblatt 4

Ausgabe: 12.05.2016; Abgabe: bis 26.05.2016, 23:59 Uhr.

Aufgabe 1 (Fragen zur Vorlesung) (9 Punkte)

Folgende Fragen sind zu beantworten:

a) In Anwendungen ist die Likelihoodfunktion im Allgemeinen nicht bekannt. Sie muss daher aus statistisch erhobenen Daten $D = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ geschätzt werden. Dabei wird oftmals vorausgesetzt, dass die likelihood durch eine parametrisierte Wahrscheinlichkeitsdichte gegeben ist, deren prinzipielle Form bekannt ist, und dass die erhobenen Daten stochastisch unabhängig sind. Fasst man die unbekannten Parameter in einem Vektor $\boldsymbol{\vartheta}$ zusammen, so ergibt sich für die Wahrscheinlichkeitsdichte p

$$p(D|\boldsymbol{\vartheta}) = \prod_{n=1}^{n} p(\boldsymbol{x}_k|\boldsymbol{\vartheta}).$$

Diesen Term möchte man nun über $\boldsymbol{\vartheta}$ maximieren. In der Regel arbeitet man mit dem Logarithmus dieses Ausdrucks.

Erklärt, weswegen dies sinnvoll und dem Vorhaben nicht abträglich ist, und formt den Ausdruck dazu kurz in die übliche Form um. Geht dabei besonders auf die Eigenschaften des Logarithmus ein. (4 Punkte)

b) Wie berechnen sich die Parameter μ und σ für eine univariate Normalverteilung bei der Maximum-Likelihood-Schätzung?

c) Was ist die Hauptannahme bei der Bayesschen Parameterschätzung?

d) Warum gilt laut Foliensatz 4, Folie 19, $p(x|D) = \mathcal{N}(\mu_n, \sigma^2 + \sigma_n^2)$ und nicht $p(x|D) = \mathcal{N}(\mu_n, \sigma^2)$?

e) Welche Fehler können bei der Parameterschätzung auftreten?

f) Inwiefern unterscheidet sich ein Schätzfehler von einem Modellfehler?

Aufgabe 2 (Maximum-Likelihood-Schätzung) (10 Punkte)

Gegeben seien Ergebnisse $x_1, \dots, x_n$ von unabhängigen und identisch verteilten Zufallsvariablen $X_1, \dots, X_n$, wobei $X_i \sim \mathcal{N}(\mu, \sigma^2)$ gelte. Die Wahrscheinlichkeitsdichte der Normalverteilung $\mathcal{N}(\mu, \sigma^2)$ ist dabei gegeben durch

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

Nun soll angenommen werden, dass σ bekannt und μ unbekannt sei. Unter diesen Voraussetzung ist für den *Maximum-Likelihood-Schätzer* μ_{MLE} zu zeigen:

$$\mu_{\mathrm{MLE}} = \frac{\sum_{i=1}^{n} x_i}{n}.$$

Nachzuweisen ist also, dass der Maximum-Likelihood-Schätzer für den Parameter μ dem Durchschnitt der Stichprobe $\{x_1, \dots, x_n\}$ entspricht.

Hinweise:

- Es empfiehlt sich, mit der logarithmierten Likelihoodfunktion zu argumentieren.
- Der Parameter σ kann als Konstante betrachtet werden.
- Für unabhängige Zufallsvariablen $Z_1, \dots, Z_n$ und mögliche Ergebnisse $z_1, \dots, z_n$ gilt allgemein

$$P\big(Z_1 = z_1 \wedge \dots \wedge Z_n = z_n\big) = \prod_{i=1}^{n} P(Z_i = z_i).$$

Aufgabe 3 (Bayessche Parameterschätzung) (16 Punkte)

In dieser Aufgabe soll die *Bayessche Parameterschätzung* betrachtet werden. Dazu seien unabhängige und normalverteilte Zufallsvariablen mit $\mu = 3$ und $\sigma = 1$ gegeben. Der Parameter μ soll nun als unbekannt betrachtet werden und über ein Schätzverfahren approximiert werden.

Dazu ist μ mit einem Bayesschen Verfahren über eine Stichprobe der Größe $2^i \cdot 10$, $i \in \{0, 1, 2, 3, 4, 5\}$, zu schätzen. Für die Bayesschätzung bei Normalverteilung gilt dabei (vgl. Foliensatz 4, Folie 17)

$$\mu_n := \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \cdot \widehat{\mu_n} + \frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \cdot \mu_0.$$

a) Zunächst sollen die folgenden Parameter beschrieben werden:
   - $\widehat{\mu_n}$, insbesondere für $n = 0$, (0.5 Punkte)
   - μ_0, (0.5 Punkte)
   - σ_0^2, (0.5 Punkte)
   - σ^2. (0.5 Punkte)

   Man beschreibe zudem das Verhalten für große n und kleine n. (Jeweils 1 Punkt)

b) Über was gibt σ_n^2 Auskunft? Die Definition von σ_n^2 lautet (vgl. Foliensatz 4, Folie 17)

$$\sigma_n^2 := \frac{\sigma_0^2\sigma^2}{n\sigma_0^2 + \sigma^2}.$$ (2 Punkte)

c) Lest die zur Verfügung gestellte Datei `samples.csv` ein.

d) Wähle nun für jedes i die ersten $2^i \cdot 10$ Einträge aus dem Stichprobenvektor und berechne den Bayes-Schätzer für μ. Für μ_0 sei der Wert -10 angesetzt und für σ_0^2 der Wert 1. (6 Punkte)

e) Plottet für jedes i jeweils in einer gemeinsamen Grafik die zu approximierende Funktion ($\mathcal{N}(3, 1)$) und deren Bayessche Schätzung. (1 Punkt)

f) Berechnet für jedes i den *quadratischen Fehler* gegenüber der ursprünglichen Verteilung ($\mathcal{N}(3, 1)$) und gebt diesen an. (1 Punkt)

g) Beschreibt in eigenen Worten, wie sich die Bayessche Schätzung im Vergleich zur Maximum-Likelihood-Schätzung verhält. (2 Punkte)

Hinweise:

- Die Bayessche Parameterschätzung wird in Duda, S. 115–117, gut dargestellt. Die Lektüre ist zum besseren Verständnis zu empfehlen.
- Die Variation von μ_0 und die Beobachtung der Auswirkungen auf die Plots kann für das Verständnis ebenfalls hilfreich sein.