

# Maschinelles Lernen Blatt 3

Nikolas Zeitler, Joshua Hartmann, Alexander Diegel

May 18, 2016

## 1 Fragen zur Vorlesung

### a) Darstellung als log-likelihood und warum sinnvoll?

$$\ln(p(D|\vartheta)) = \sum_{k=1}^n \ln(p(D|\vartheta))$$

Sinnvoll:

Kann im Rahmen der Maximum-Likelihood-Methode verwendet werden. Dies ist ein parametrisches Schätzverfahren. dabei wird der parameter als Schätzung gewählt der abhängig von der Verteilung am richtigsten wirkt.

Log ist numerisch stabiler.

Logarithmus wandelt den Produktterm in einen Summenterm um. Eine Summe ist idR. einfacher zu maximieren als ein Produkt (Rechenoperationen gehen schneller aus Computersicht). -Maximum Likelihood such Parameter, die die Trainingsdaten am besten erklären.

### b) Wie berechnet man Erwartungswert und Standardabweichung

Foliensatz 3 S.9

$$\mu = \frac{1}{n} \sum_{k=1}^n x_k$$

$$\sigma^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \mu)^2$$

### c) Was ist die Hauptannahme bei der Bayesschen Parameterschätzung

Da Bayes genutzt wird müssen die Messungen unabhängig sein. Die klassenabhängige Dichte  $p(x|w_i, D)$  sei nur von den Daten der jeweiligen Klasse abhängig, und betrachten daher nur eine Klasse. Das  $w_i$  entfällt also.

### d)

Um die Unsicherheit auszugleichen die wir mit dem geschätzten  $\mu$  gemacht haben wird die bekannte Varianz um  $\sigma_m^2$  vergrößert. Deswegen gilt die Formel wie sie da steht.

### e) Welche Fehler bei Parameterschätzung möglich

Bei einer sehr hohen Standardabweichung braucht man entsprechend viele Messungen sonst ist das Ergebnis "geraten" und damit womöglich Fehlerhaft.

- Bayes Fehler
- Modell- Fehler
- Schätzfehler

### f) Unterschied Schätzfehler und Modellfehler

Wikipedia:

Ein Modellfehler ergibt sich aus einem fehlerhaften Modellansatz. Solch ein Fehler ist kein statistisches Phänomen, sondern ein sog. Wirklichkeitsphänomen, kann also beispielsweise durch eine Korrektur der Modellstruktur behoben werden.

In der Statistik bezeichnet der Schätzfehler die Abweichung einer Schätzfunktion  $\hat{\vartheta}$  vom unbekannten Parameter der Grundgesamtheit  $\vartheta$ . Er ist ein Maß für die Güte der Schätzfunktion (oder Interpolation).

Ein Modellfehler hat also nichts mit einem Statistischen Ergebnis zu tun. (Schätzfehler zu) Ein Modellfehler ist der Fehler den man hat wenn man von Grund auf von einem Fehlerhaften Modell ausgeht.

Der Schätzfehler verbessert sich mit zunehmender Anzahl an Stichproben, der Modellfehler bleibt.

## 2 Maximum-Likelihood-Schätzung

**Nachweis, dass der Maximum-Likelihood-Schätzer für den Parameter  $\mu$  dem Durchschnitt der Stichprobe entspricht. (10 Punkte)**

Es ist:

$$p(x) = 1/(\sqrt{2\pi\sigma^2}) * \exp(-(x_i - \mu)^2/(2\sigma^2))$$

Da  $\sigma$  bekannt ist und als Konstante betrachtet werden kann gilt für MLE:

$$\begin{aligned} MLE : p(X|\mu) &= \prod_{i=1}^n p(x_i) \\ &= \prod_{i=1}^n 1/(\sqrt{2\pi\sigma^2}) * \exp(-(x_i - \mu)^2/(2\sigma^2)) \\ &= 1/((2\pi\sigma)^{n/2}) * \exp(-1/(2\sigma^2) * \sum_{i=1}^n (x_i - \mu)^2) \end{aligned}$$

Mit Log-MLE:

$$\ln(p(X|\mu)) = -n/2 * \ln(2\pi\sigma) - 1/(2\sigma^2) * \sum_{i=1}^n (x_i - \mu)^2$$

Ableiten nach  $\mu$  ergibt:

$$\frac{\partial p(X|\mu)}{\partial \mu} = \frac{1}{\sigma^2} * \sum_{i=1}^n (x_i - \mu)$$

Ableitung = 0 gdw. Summenterm 0 ergibt!

$$\sum_{i=1}^n (x_i - \mu) = 0$$

$$\sum_{i=1}^n (x_i) - n * \mu = 0$$

$$\mu = \sum_{i=1}^n (x_i) / n$$

Damit ist gezeigt, dass der Schätzer von MLE für  $\mu$  dem Durchschnitt der Stichprobe entspricht.

### 3 Bayessche Parameterschätzung

#### a) Parameterbeschreibung

- $\hat{\mu}_n$ :  
Mittelwert der Stichprobe für  $n=0$  wird er unendlich groß da er gemäß  $\sum(x_i)/n$  berechnet wird. Für großes  $n$  nähert sich der Stichprobenmittelwert dem wahren Mittelwert der Verteilung immer mehr an. Für  $n = \infty$  entspricht er dem wahren Mittelwert. Für kleines  $n$  ist der Mittelwert der Stichprobe gegenüber dem wahren Mittelwert der Verteilung mit Unsicherheit behaftet.
- $\mu_0$ :  
A priori Schätzer für den Mittelwert. Ändert sich nicht für unterschiedliches  $n$ .
- $\sigma_0^2$ :  
Gibt Unsicherheit für das geschätzte a priori  $\mu_0$  an. Unabhängig von  $n$ .
- $\sigma^2$ :  
Wahre Varianz der normalverteilten Zufallsvariablen. An der wahren Varianz ändert sich durch die Stichprobengröße ebenfalls nichts.

#### b) Auskunft $\sigma_n^2$

$\sigma_n^2$  gibt die Unsicherheit für den, mit der Bayesschätzung, geschätzten Mittelwert  $\mu_n$  nach  $n$  Samples an. Umso mehr Samples verwendet werden, desto kleiner wird die Varianz und desto mehr entspricht sie der wahren Varianz der Verteilung.

e) Plottet für jedes  $i$  die zu approximierende Funktion  $N(3,1)$  und die Bayessche Schätzung. Bayes ist rote Kurve

