Maschinelles Lernen Blatt 3

Nikolas Zeitler, Joshua Hartmann, Alexander Diegel

May 26, 2016

1 Fragen zur Vorlesung

a) Darstellung als log-likelihood und warum sinnvoll?

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

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 die Daten am besten erklärt.
Der Logarithmus wandelt den Produktterm in einen Summenterm um. Eine Summe ist idR. einfacher zu maximieren als ein Produkt (Rechenoperationen gehen schneller aus Computersicht) da jeder Term für sich abgeleitet werden kann und keine Produktregel angewandt werden muss.

Das Maximum des Ausdrucks bleibt auch weiterhin das Maximum des logarithmierten Ausdrucks, daher ist die Form dem Vorhaben nicht abträglich.

Maximum Likelihood such Parameter, die die Trainingsdaten am besten erklären.

b) Wie berechnet man Erwartungswert und Standardabweichung

Foliensatz 3 S.9

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

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

Eigentlich jeweils \widehat{...}

c)Was ist die Hauptannahme bei der Bayesschen Parameterschätzung

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.

Noch wichtiger: Dichte hat bekannte parametrische Form!

d)

Um die Unsicherheit auszugleichen die wir mit dem geschätzten μ gemacht haben wird die bekannte Varianz um σ_m^2 vergößert. Deswegen gilt die Formel wie sie da steht.
Aber warum gerade um diese Varianz?

e) Welche Fehler bei Parameterschätzung möglich

Bayes Fehler Man entscheidet sich falsch. Aber warum?

Modell- Fehler Falsche parametrische Form für μ und σ^2 gewählt. Man arbeitet also mit dem falschen Modell

Schätzfehler Zu wenige Trainingsdaten.

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 ϑ. 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. Da muss ich Wikipedia wohl gut bewerten! ;-)

2 Maximum-Likelihood-Schätzung

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

Es ist:
$p(x) = \frac{1}{(\sqrt{2\pi\sigma^2})} \cdot \exp(-\frac{(x_i-\mu)^2}{2\sigma^2})$
Da σ bekannt ist und als Konstante betrachtet werden kann gilt für MLE:
$MLE : p(X|\mu) = \prod_{i=1}^{n} p(x_i)$
$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp(-\frac{(x_i-\mu)^2}{2\sigma^2})$
$= \frac{1}{(2\pi\sigma)^{n/2}} \cdot \exp(-\frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n}(x_i-\mu)^2)$

Mit Log-MLE:

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

Ableiten nach μ ergibt:
$\frac{\partial p(X|\mu)}{\partial \mu} = \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n}(x_i - \mu)$

Ableitung = 0 gdw. Summenterm 0 ergibt!
$\sum_{i=1}^{n}(x_i - \mu) = 0$
$\sum_{i=1}^{n}(x_i) - n \cdot \mu = 0$
$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$
Damit ist gezeigt, dass der Schätzer von MLE für μ dem Durchschnitt der Stichprobe entspricht.

3 Bayessche Parameterschätzung

a) Paramterbeschreibung

- $\hat{\mu_n}$:
  Mittelwert der Stichprobe für n=0 ist nicht definiert. Die Formel lautet: $\sum(x_i)/n$ Division durch null ist nicht gestattet und wenn $n = 0$ ist gibt es auch kein x_i da es keine Stichproben gibt. Für großes n nähert sich der Stichprobenmittelwert dem wahren Mittelwert der Verteilung immer mehr an. Für n = ∞ entspricht er dem wahren Mittelwert. Für kleines n ist der Mittelwert der Stichprobe gegenüber dem wahren Mittelwert der Verteilung mit Unsicherheit behaftet.

- μ_0:
  A priori Schätzer für den Mittelwert. Ändert sich nicht für unterschiedliches n.

- σ_0^2:
  Gibt Unsicherheit für das geschätzte a priori μ_0 an. Unabhängig von n.

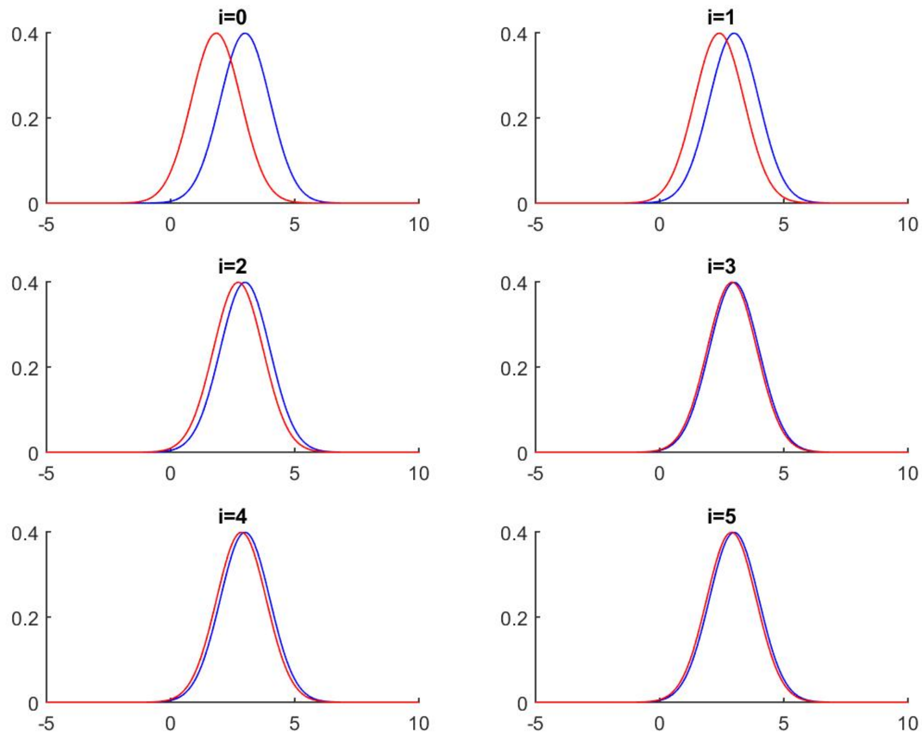
- σ^2:
  Wahre Varianz der normalverteilten Zufallsvariablen. An der wahren Varianz ändert sich durch die Stichprobengröße ebenfalls nichts.

b) Auskunft σ_n^2

σ_n^2 gibt die Unsicherheit für den, mit der Bayesschätzung, geschätzten Mittelwert μ_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.

c)-f) siehe Matlab-File für Berechnungen

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



f)Angabe des quadratischen Fehlers

- i=0: 0.0108
- i=1: 0.0032
- i=2: 7.0277e-04
- i=3: 6.7479e-05
- i=4: 2.0726e-04
- i=5: 1.0817e-04

g) Beschreibt in eigenen Worten, wie sich Bayes im Vergleich zu Maximum-Likelihood verhält

Bayes: Liefert Wahrscheinlichkeit für die geschätzten Parameter.
ML: Sucht Parameter, die Trainingsdaten am besten repräsentieren.