

Maschinelles Lernen und Data Mining

6. Übung

Prof. Tobias Scheffer
Steffen Bickel

WS06/07

Ausgabe am: 24.11.06
Besprechung am: 01.12.06

Aufgabe 1 (1/3 Punkt):

Nehmen wir an, wir haben eine Menge von Beispielen x_1, \dots, x_n , unabhängig aus einer $N(\mu, 1)$ normalverteilten Grundgesamtheit gezogen. Die Dichte einer univariaten Normalverteilung ist:

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

Wir möchten eine Parameterschätzung für den Parameter μ durchführen. Leiten sie den Maximum-Likelihood-Schätzer für den Parameter μ her. Die Likelihood der Daten ist $P(x_1, \dots, x_n|\mu, \sigma^2) = \prod_{i=1}^n p(x_i|\mu, \sigma^2)$. Hinweis: Da $\log()$ eine streng monotone Funktion ist, befindet sich an einem Maximum von $\log P(x_1, \dots, x_n|\mu, \sigma^2)$ auch ein Maximum von $P(x_1, \dots, x_n|\mu, \sigma^2)$. Finden sie das Maximum von $\log P(x_1, \dots, x_n|\mu, \sigma^2)$, es lässt sich leichter maximieren/ableiten.

Aufgabe 2 (1/3 Punkt):

Werfen Sie einen Blick auf folgende Datenbank mit Supermarkttransaktionen.

TID	Brot	Tiefkühlpizza	Bier	Chips	Waschpulver
0	0	1	1	1	0
1	1	0	1	1	0
2	0	1	1	0	0
3	0	1	1	1	0
4	0	1	1	1	1

Simulieren Sie den Apriori-Algorithmus. Finden Sie alle Assoziationsregeln mit minimalem Support von 0,6 und Konfidenz von 0,8.

Aufgabe 3 (2/6 Punkt):

Nehmen sie an, die längste Assoziationsregel (gemessen als Anzahl der beteiligten Items), die der Apriori-Algorithmus auf einem bestimmten Datensatz gefunden hat ist $ABCDE \Rightarrow FGHIJ$ und für diese Regel gilt $\text{Support} \geq s_{min}$ und $\text{Confidence} \geq c_{min}$.

- Geben Sie eine möglichst große untere Schranke für die Anzahl der gefundenen häufigen Itemsets an.
- Geben Sie eine möglichst große untere Schranke für die Anzahl der Regeln an, in denen die gleichen Items A, \dots, J alle zusammen vorkommen und für die $\text{Support} \geq s_{min}$ und $\text{Confidence} \geq c_{min}$ gilt.