Maschinelles Lernen II

4. Übung

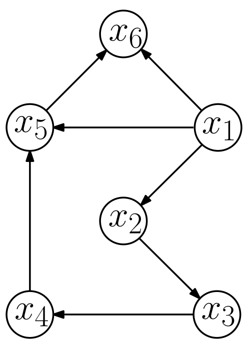
Prof. Tobias Scheffer
Dr. Niels Landwehr
Matthias Bussas

Sommer 2014

Ausgabe am: 06.05.14
Besprechung am: 13.05.14

Aufgabe 1 *Message-Passing Algorithmus*

Betrachten Sie das folgende Bayessche Netz über den sechs binären Zufallsvariablen x_1, x_2, x_3, x_4, x_5, x_6.



$p(x_1 = 1)$
0.5

$p(x_2 = 1 \mid x_1)$	x_1
0.6	0
0.3	1

$p(x_3 = 1 \mid x_2)$	x_2
0.7	0
0.4	1

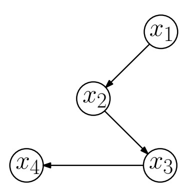
$p(x_4 = 1 \mid x_3)$	x_3
0.2	0
0.5	1

$p(x_5 = 1 \mid x_1, x_4)$	x_1	x_4
0.3	0	0
0.6	1	0
0.5	0	1
0.1	1	1

$p(x_6 = 1 \mid x_1, x_5)$	x_1	x_5
0.5	0	0
0.1	1	0
0.3	0	1
0.4	1	1

Wir möchten die bedingte Verteilung $p(x_3 \mid x_1 = 0, x_4 = 1)$ unter Verwendung des Message-Passing Algorithmus berechnen.

1. Zeigen Sie zunächst, dass die Inferenzanfrage $p(x_3 \mid x_1 = 0, x_4 = 1)$ dasselbe Ergebnis liefert wie die Inferenzanfrage $p(x_3 \mid x_1 = 0, x_4 = 1)$ in dem folgenden graphischen Modell:

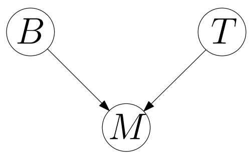


   Dabei sollen dieselben (bedingten) Verteilungen auf den Variablen x_1, x_2, x_3, x_4 angenommen werden.

2. Berechnen Sie $p(x_3 \mid x_1 = 0, x_4 = 1)$ durch Message Passing auf der linearen Kette.

Aufgabe 2 *Parameterlernen*

Wir betrachten das folgende graphische Modell über den drei binären Variablen T (Tank gefüllt), B (Batteriespannung ok) und M (Motor startet):



$p(B=1)$
θ_B

$p(T=1)$
θ_T

$p(M=1\|B,T)$	B	T
$\theta_{M\|00}$	0	0
$\theta_{M\|01}$	0	1
$\theta_{M\|10}$	1	0
$\theta_{M\|11}$	1	1

Die gemeinsame Verteilung $p(B,T,M) = p(B)p(T)p(M|B,T)$ ist entsprechend den gegebenen Tabellen mit 6 Parametern parametrisiert.
Leider kennen wir die echten Parameterwerte nicht. Wir haben aber die folgenden 10 Beobachtungen des Systems gemacht:

B	1	1	1	1	1	1	0	0	0	0
T	1	1	0	0	0	0	1	1	1	0
M	1	1	0	0	0	1	0	0	1	0

1. Wir möchten die echten Parameterwerte aus den Beobachtungen schätzen. Leiten Sie dazu die Likelihood der Beobachtungen als Funktion der 6 Parameter her. Wir nehmen wie üblich an, dass einzelne Beobachtungen unabhängig sind gegeben das Modell. Berechnen Sie die Parameterwerte $\hat{\theta}_B$, $\hat{\theta}_T$,$\hat{\theta}_{M|00}, ..., \hat{\theta}_{M|11}$ die die Likelihood maximieren (Hinweis: Logarithmus).

2. Maximum-Likelihood Parameterschätzungen führen oft zu (unrealistischen) Schätzungen von Null oder Eins für Wahrscheinlichkeiten. Argumentieren Sie, welche Klasse von Prior-Verteilungen geeignet wären, um diese Fälle zu verhindern. Wie sähe die entsprechende a-posteriori Verteilung über Parameter aus, und wie könnte man die entsprechenden maximum-a-posteriori Parameter berechnen?

Aufgabe 3 *Unabhängigkeit, bedingte Unabhängigkeit*

Wir betrachten die folgende Domäne: eine faire Münze wird zwei Mal geworfen; die Ergebnisse der Würfe sind repräsentiert durch die Zufallsvariablen $X, Y \in \{0, 1\}$. Wir definieren eine dritte Zufallsvariable Z durch $Z = xor(X, Y)$, dh. Z hat den Wert Eins falls genau eine der beiden Variablen X, Y den Wert Eins hat.

1. Geben Sie die gemeinsame Verteilung $p(X, Y, Z)$ über die Variablen X, Y, Z an.

2. (a) Sind X, Y, Z paarweise unabhängig, dh. gilt $p(X, Y) = p(X)p(Y)$, $p(X, Z) = p(X)p(Z)$, und $p(Y, Z) = p(Y)p(Z)$?
   (b) Sind X, Y, Z unabhängig, dh. gilt $p(X, Y, Z) = p(X)p(Y)p(Z)$?
   (c) Geben Sie alle Unabhängigkeiten an, die in der Verteilung $p(X, Y, Z)$ gelten, als Menge
   $$I(p) := \{(A \perp B|C) \mid p(A|B, C) = p(A|C)\}$$
   wobei A, B, C beliebige Teilmengen von $\{X, Y, Z\}$ sind.

3. Geben Sie ein Bayessches Netz für die Verteilung $p(X, Y, Z)$ an. Gilt $I(G) = I(p)$?