Advanced data analysis II

9. Exercise

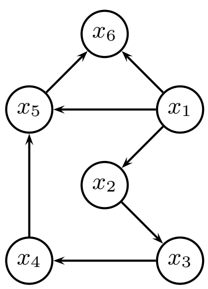
Prof. Tobias Scheffer
Dr. Niels Landwehr
Ahmed Abdelwahab

Winter 2015

From: 15.12.15
To: 05.01.16

Exercise 1 *Message Passing Algorithm*

Consider the following Bayesian network over the six binary random variables 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

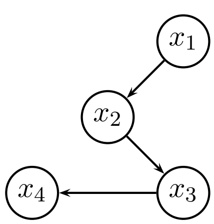
$p(x_4 = 1 \mid x_3)$	x_3
0.2	0
0.5	1

$p(x_5 = 1 \mid x_1, x_4)$	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

We want to calculate the conditional distribution $p(x_3 \mid x_1 = 0, x_4 = 1)$ using the Message Passing algorithm.

1. First show that the inference query $p(x_3 \mid x_1 = 0, x_4 = 1)$ in the original graphical model yields the same result as the inference query $p(x_3 \mid x_1 = 0, x_4 = 1)$ in the following simpler graphical model:



Here, we assume the same (conditional) distributions for the variables x_1, x_2, x_3, x_4.

2. Calculate $p(x_3 \mid x_1 = 0, x_4 = 1)$ by Message Passing on the linear chain.

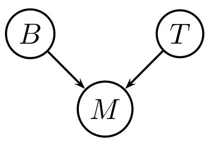
Exercise 2 *Independence, conditional independence*

We study the following domain: a fair coin is tossed twice; the results of tosses are represented by the random variables $X, Y \in \{0, 1\}$. We define a third random variable Z by $Z = xor(X, Y)$, that is, Z is one if exactly one of the two variables X, Y has the value one.

1. What is the joint distribution $p(X, Y, Z)$ over the three random variables X, Y, Z?
2. (a) Are X, Y, Z pairwise independent? That is, does it hold that $p(X, Y) = p(X)p(Y)$, $p(X, Z) = p(X)p(Z)$, and $p(Y, Z) = p(Y)p(Z)$?
   (b) Are X, Y, Z independent? That is, does it hold that $p(X, Y, Z) = p(X)p(Y)p(Z)$?
   (c) Specify all independencies that hold in the distribution $p(X, Y, Z)$, in the form of a set
   $$I(p) := \{(A \perp B|C) \mid p(A|B, C) = p(A|C)\}$$
   where A, B, C are any subsets of $\{X, Y, Z\}$.
3. Construct a Bayesian network for the distribution $p(X, Y, Z)$. Does it hold that $I(G) = I(p)$?

Exercise 3 *Parameter learning*

We study the following simple graphical model over the three binary variables T (tank filled), B (battery voltage ok) and M (motor starts):



$p(B = 1)$
θ_B

$p(T = 1)$
θ_T

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

The joint distribution $p(B, M, T) = p(B)p(T)p(M|B, T)$ is parameterized by the six parameters θ_B, θ_T,$\theta_{M|00}, ..., \theta_{M|11}$ given in the tables. Unfortunately, we do not know the true parameter values. However, we have made the following 10 observations of the system:

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

1. We would like to estimate the true parameter values from the observations. In order to do that, derive the likelihood of the observations as a function of the six parameters. We assume as usual that individual observations are independent given the model. Compute estimates $\hat{\theta}_B$, $\hat{\theta}_T$,$\hat{\theta}_{M|00}, ..., \hat{\theta}_{M|11}$ of the true parameters that maximize the likelihood (hint: logarithm, derivative).

2. Maximum likelihood parameter estimates often lead to unrealistic estimates of zero or one for probabilities.

   Argue which prior distribution would be suitable to prevent these cases. What is the corresponding posterior distribution over parameters, and how can we calculate the corresponding maximum a posteriori parameters?