Machine Learning 2 **Due: 19 May, 2016**

Homework 6

Instructor: Ke Tran Email: m.k.tran@uva.nl

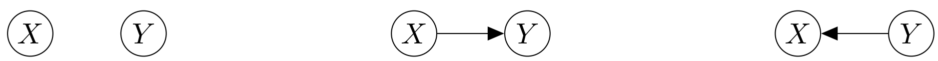
Student: Andrea Jemmett UvA ID: 11162929
Collaborators: N/A Email: andrea.jemmett@student.uva.nl

Problem 1.

Problem 2.

Solution:

(a)



(a) No cause-effect relation (b) X causes Y (c) Y causes X

Figure 1: Two nodes Causal Bayesian Networks

(b) For the Causal Bayesian Networks in Figures 1a, 1b and 1c respectively we have:

$$p(X, Y) = p(X)p(Y) \tag{1}$$
$$p(X, Y) = p(Y|X)p(X) \tag{2}$$
$$p(X, Y) = p(X|Y)p(Y) \tag{3}$$

(c) For the Causal Bayesian Networks in Figures 1a and 1c respectively we have:

$$p(Y|X) = p(X)p(Y) \tag{4}$$
$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} = \frac{p(X|Y)p(Y)}{\sum_Y p(X|Y)p(Y)} \tag{5}$$

while $p(Y|X)$ is already a term of the factorization for the graph in Figure 1b.

(d) For the Causal Bayesian Networks in Figures 1a, 1b and 1c respectively we have:

$$p(Y|do(X)) = p(Y) \tag{6}$$
$$p(Y|do(X)) = \frac{p(X, Y)}{p(X)} = p(Y|X) \tag{7}$$
$$p(Y|do(X)) = p(Y) \tag{8}$$

(e)

∎

Problem 3. Simpson's paradox

Solution:

1a. The recovery rate for *treatment* is 50%, while for *untreated* is 40%.

1b. I would advise to take the drug because the recovery rate is higher for the *treatment* group.

2a.

Recovery rates	Drug	No drug
Male	60%	70%
Female	20%	30%

2b. I would not advice to take the drug nor to male patients nor to female patients because the recovery rate, given the patient's gender supports it.

3. With hindsight I would not advice a patient with unknown gender to take the drug because for both genders the recovery rate does not support it. This is in contradiction with the conclusion given in (1b).

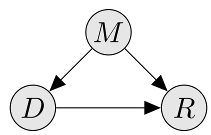


Figure 2: Causal model where M denotes the gender.

4a. By applying the back-door criterion on the causal model in Figure 2 we have:

$$p(R|do(D)) = \sum_M p(R|D,M)p(M) \tag{9}$$

4b. Using normal probability rules we have:

$$p(R|D) = \sum_M p(R,M|D) = \sum_M p(R|D,M)p(M) \tag{10}$$

and so $p(R|do(D)) = p(R|D)$ in this case.

∎