Machine Learning 2 - Homework
Week 6

Minh Ngo
MSc Artificial Intelligence
University of Amsterdam
nlminhtl@gmail.com

May 21, 2016

Collaborators: Riaan Zoetmulder, Arthur Bražinskas

Problem 2

a Possible structures are presented in the figure 1.

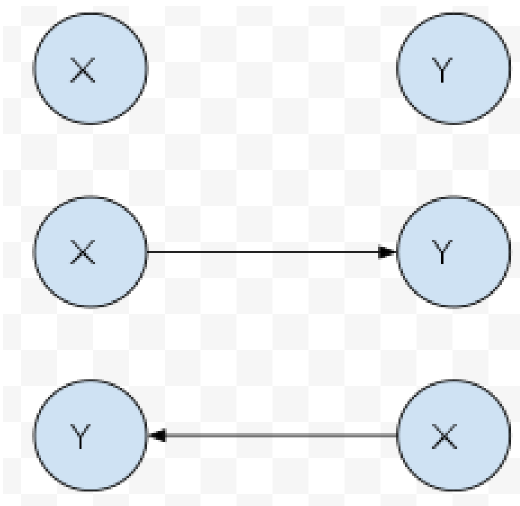

Figure 1: (i) X and Y are independent, (ii) X causes Y, (iii) Y causes X

b (i) $p(X, Y) = p(X)p(Y)$

(ii) $p(X,Y) = p(Y|X)p(X)$

(iii) $p(X,Y) = p(X|Y)p(Y)$

c (i) $p(Y|X) = \dfrac{p(X,Y)}{p(X)} = \dfrac{p(X)p(Y)}{p(X)} = p(Y)$

(ii) $p(Y|X) = \dfrac{p(X,Y)}{p(X)} = \dfrac{p(Y|X)p(X)}{p(X)} = p(Y|X)$

(iii) $p(Y|X) = \dfrac{p(X,Y)}{p(X)} = \dfrac{p(X|Y)p(Y)}{\sum_Y p(X|Y)p(Y)}$

d (i) $p(Y|do(X)) = p(Y)$

(ii) $p(Y|do(X)) = p(Y|X)$

(iii) $p(Y|do(X)) = p(Y)$

e $p(Y|X)$ - probability that somebody gets lung cancer, given the observation that the person smokes.

$p(Y|do(X))$ - probability that somebody gets lung cancer if we force the person to smoke [Slides, page 20].

Problem 3

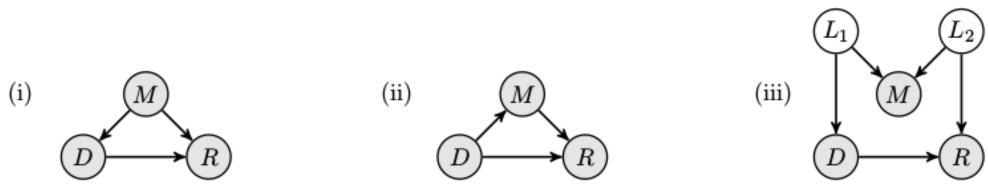

Figure 2: Different hypothetical causal models.

1a $p(R=1|D=1) = \dfrac{20}{40} = 0.5$

$p(R=1|D=0) = \dfrac{16}{40} = 0.4$

1b $p(R=1|D=1) > p(R=1|D=0)$ therefore a new patient should take the drug.

2a $p(R=1|D=1,M=1)=\frac{18}{30}=0.6$

$p(R=1|D=0,M=1)=\frac{7}{10}=0.7$

$p(R=1|D=1,M=0)=\frac{2}{10}=0.2$

$p(R=1|D=0,M=0)=\frac{9}{30}=0.3$

2b Chance to recover is lower if a patient takes the drug, therefore a new patient of both gender shouldn't be recommended to take the drug.

3 $p(R=1|D=1)=p(R=1|D=1,M=1)p(M=1|D=1)+p(R=1|D=1,M=0)p(M=0|D=1)=0.6*0.75+0.2*0.25=0.5$

$p(R=1|D=0)=p(R=1|D=0,M=1)p(M=1|D=0)+p(R=1|D=0,M=0)p(M=0|D=0)=0.7*0.25+0.3*0.75=0.4$

Since $p(R=1|D=1)>p(R=1|D=0)$ we should suggest drugs that contradicts [2b] but doesn't contradict [1b].

4a (1) $S=\{M\}$, $D,R\notin S$

(2) M is not a descendant of D

(3) M blocks all back-door path $D\leftarrow ..R$

$\Rightarrow p(R|do(D))=\int p(R|D,m)p(m)dm=\sum_M p(R|D,M)p(M)$

4b

$$\begin{aligned} p(R|D) &= \frac{\int p(R|D,M)p(D|M)p(M)dM}{\int\int p(R|D,M)p(D|M)p(M)dMdD} \\ &= \frac{\sum_M p(R|D,M)p(D|M)p(M)}{\sum_M\sum_D p(R|D,M)p(D|M)p(M)} \end{aligned} \tag{1}$$

Consequently $p(R|do(D))\neq p(R|D)$.

4c

$$\begin{aligned} p(R=1|do(D=1)) &= \sum_M p(R=1|D=1,M)p(M) \\ &= 0.6*0.5+0.2*0.5=0.4 \end{aligned} \tag{2}$$

$$\begin{aligned} p(R=1|do(D=0)) &= \sum_M p(R=1|D=0,M)p(M) \\ &= 0.7*0.5+0.3*0.5=0.5 \end{aligned} \tag{3}$$

$p(R = 1|do(D = 0)) > p(R = 1|do(D = 1))$, therefore we shouldn't recommend drugs.

5a

$$p(R = 1|do(D)) = \sum_M p(R = 1|D, M)p(M|D) = p(R = 1|D) \tag{4}$$

5b Yes, $p(R|do(D)) = p(R|D)$

5c $p(R = 1|do(D = 1)) = 0.6 * 0.75 + 0.2 * 0.25 = 0.5$

$p(R = 1|do(D = 0)) = 0.7 * 0.25 + 0.3 * 0.75 = 0.4$

$p(R = 1|do(D = 1)) > p(R = 1|do(D = 0))$, therefore we should recommend drugs.

6a D - drug / non drug, R - recover / non recover, L2 - good doctor / bad doctor, M - good mood / bad mood, L1 - stress / non stress.

6b Since $D \leftarrow L_1 \rightarrow M \leftarrow L_2 \rightarrow R$ is initially blocked $p(R|do(D)) = p(R|D)$

6c

$$p(R|D) = \frac{\int\int p(L_1)p(L_2)p(R|D, L_2)p(D|L_1)dL_1dL_2}{\int\int\int p(L_1)p(L_2)p(R|D, L_2)p(D|L_1)dL_1dL_2dR} \tag{5}$$

Consequently $p(R|do(D)) \neq p(R|D)$

6d $p(R = 1|do(D = 1)) = p(R = 1|D = 1) = 0.5$

$p(R = 1|do(D = 0)) = p(R = 1|D = 0) = 0.4$

$p(R = 1|do(D = 1)) > p(R = 1|do(D = 0))$, therefore drug should be recommended.

Problem 5

1

$$\begin{aligned} p(Y|do(X), X_{pa(X)}) &= \frac{p(Y, X_{pa(X)}|do(X))}{p(X_{pa(X)}|do(X))} \\ &= \frac{p(Y|X, X_{pa(X)})p(X_{pa(X)})}{p(X_{pa(X)})} \\ &= p(Y|X, X_{pa(X)}) \quad q.e.d. \end{aligned} \tag{6}$$

since do(X) removes all incoming arrows to X and consequently X will be independent of $X_{pa(X)}$ (Slides, page 37) and $p(X_{pa(X)|do(X)} = p(X_{pa(X)})$. In addition, if Y is independent from one of X's parents then conditional probability over this variable will neglect it since dependent on that probability will sum up to 1 and therefore $p(Y|X, X_{pa(X)})$ is the same like $p(Y|Y_{pa(Y)})$.

2

$$p(X_{pa(X)}|do(X)) = p(X_{pa(X)}) \tag{7}$$

Since they are independent. Taking an example: a student doing a homework. His excellent grade (X) is dependent on his material understanding of each topic from the homework ($X_{pa(X)}$). If he will just copy right answers (and let's assume that it's allowed for simplicity) then he will get an excellent grade independent from his understanding.

$$\begin{aligned}
p(Y|do(X)) &= \int p(Y, X_{pa(X)}|do(X))dX_{pa(X)} \\
&= \int p(Y|X_{pa(X)}|, do(X))p(X_{pa(X)}|do(X))dX_{pa(X)} \\
&= \int p(Y|X_{pa(X)}, X)p(X_{pa(X)})dX_{pa(X)} \quad q.e.d.
\end{aligned} \tag{8}$$

It's valuable to mention that we can do marginalization over parents of X even if any of them are independent of Y, since integration over this variable will be equal to 1.