Assignment 3 Machine Learning COS4852

Adriaan Louw (53031377)

September 8, 2018

Contents

1 Question 1 **1**

2 Question 2 **2**
2.1 Question 2a 2
2.2 Question 2b 2

3 Question 3 **2**

1 Question 1

Given Bayes' theorem where

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{1}$$

We need to calculate the probability of Joe being a librarian (L) given that he has glasses. (G) This can be expressed as follows using Bayes' theorem from above

$$P(L|G) = \frac{P(G|L)P(L)}{P(G)} \tag{2}$$

We know that $P(G|L)$ is the probability that a librarian is wearing glasses, $P(L)$ is the probability that a man is a librarian an $P(G)$ is just the probability of a man wearing glasses. These probabilities are provided. Thus putting them into the above equation gives

$$\begin{aligned} P(L|G) &= \frac{P(G|L)P(L)}{P(G)} \\ &= \frac{(1)(\frac{1}{12000})}{0.12} \\ &= 0.00069 \end{aligned} \tag{3}$$

The probability that Joe is a librarian is 0.00069

Following the same argument we need to determine the probability that Joe is a salesman (S) given that he wears glasses. Using Bayes' theorem again we have

$$P(S|G) = \frac{P(G|S)P(S)}{P(G)} \tag{4}$$

$P(G|S)$ is the probability that a salesman wears glasses. $P(S)$ is the probability that a man is a salesman and $P(G)$ is again the probability that a person wears glasses. We are given this information. We now work out $P(S|G)$

$$
\begin{aligned}
P(S|G) &= \frac{P(G|S)P(S)}{P(G)} \\
&= \frac{(\frac{1}{25})(\frac{1}{200})}{0.12} \\
&= 0.00166
\end{aligned} \tag{5}
$$

We find that it is much more likely for Joe to be a salesman than a librarian. This is not what we would expect. We would expect for him to be a librarian given that every librarian wears glasses. Without more information there is no better way to calculate whether Joe is a librarian or a salesman.

2 Question 2

2.1 Question 2a

The network has 5 nodes namely Smoker (named variable S), Lung Cancer (named variable C), Pollution (named variable P), X-Ray (named variable X) and Dyspnoea (named variable D). Refer to Figure 1.

Smoker (S) has 2 states: Yes (y) and No (n)

Lung Cancer (C) has 2 states: True (t), False (f)

Pollution (P) has 2 states: High (h) and low (l)

X-Ray has 2 states: Abnormal (a) and Normal (n)

Dyspnoea has 2 states: Present (p) and Absent (a)

Lung Cancer is dependant on Smoker and Pollution. While X-Ray and Dyspnoea is dependant on Lung Cancer. This relationship can be seen in Figure 1.

2.2 Question 2b

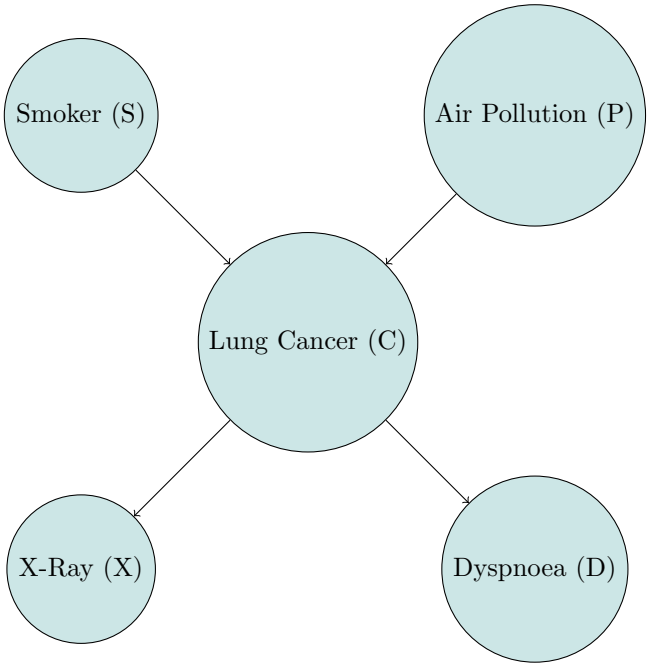
The bayesian network is described by Figure 1

3 Question 3

References

P(S=y)	P(S=n)
0.33	0.66

P(P=h)	P(P=l)
0.91	0.09



S	P	C	P(C|S,P)
n	l	f	0.999
n	l	t	0.001
n	h	f	0.97
n	h	t	0.03
y	l	f	0.98
y	l	t	0.02
y	h	f	0.96
y	h	t	0.04

X	C	P(X|C)
n	f	0.83
n	t	0.05
a	f	0.17
a	t	0.95

D	C	P(D|C)
a	f	0.75
a	t	0.3
p	f	0.25
p	t	0.7

Figure 1: Bayesian Network