Machine Learning
Winter Semester 2019/2020

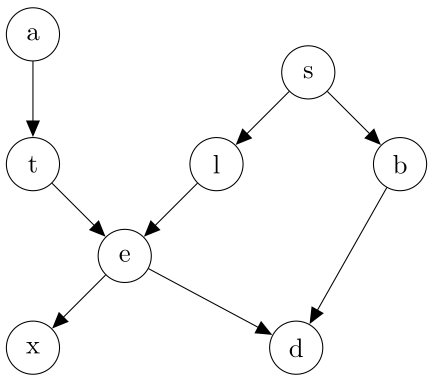
Sheet #9 – Solution

Due:

Fred Hamprecht
Manuel Haußmann

1 A Medical Probabilistic Graphical Model (20 pt +5 pt)

The setup of this exercise is the diagnosis of lung diseases in a clinical environment. Consider the (hypothetical) probabilistic graphical model as specified in Figure 1.



Variable	Meaning
a	visit to Asia?
s	smoking?
t	tuberculosis?
l	lung cancer?
b	bronchitis?
e	either tuberculosis or lung cancer
x	positive X-ray?
d	dyspnea (shortness of breath)

Figure 1: A probabilistic graphical model for the diagnosis of lung diseases.

Assume that we have the following conditional probabilities given by prior knowledge

$$
\begin{aligned}
p(a) &= 0.01 & p(s) &= 0.5 \\
p(t|a) &= 0.05 & p(t|\bar{a}) &= 0.01 \\
p(l|s) &= 0.1 & p(l|\bar{s}) &= 0.01 \\
p(b|s) &= 0.6 & p(b|\bar{s}) &= 0.3 \\
p(e|t,l) &= 1 & p(e|t,\bar{l}) &= 1 \\
p(e|\bar{t},l) &= 1 & p(e|\bar{t},\bar{l}) &= 0 \\
p(x|e) &= 0.98 & p(x|\bar{e}) &= 0.05 \\
p(d|e,b) &= 0.9 & p(d|e,\bar{b}) &= 0.7 \\
p(d|\bar{e},b) &= 0.8 & p(d|\bar{e},\bar{b}) &= 0.1
\end{aligned}
$$

Here we used the shorthand of $p(v)$ to refer to the probability that a variable is true and $p(\bar{v})$ to mean the probability that variable v is wrong (e.g. $p(a)$ gives the probability of a visit to Asia, while $p(\bar{a})$ gives the probability of not having visited Asia.)

i) Having specified a graphical model, we can directly read from the graph certain dependency/independency structures. State whether the following claims are true and false, and explain why not if they aren't.

- tuberculosis $\perp$ smoking | shortness of breath
- lung cancer $\perp$ bronchitis | smoking
- visit to Asia $\perp$ smoking | lung cancer
- visit to Asia $\perp$ smoking | lung cancer, shortness of breath

Solution:

- tuberculosis $\not\perp$ smoking | shortness of breath
  due to common descendant

- lung cancer $\perp$ bronchitis | smoking
  due to observed s in fork structure
- visit to Asia $\perp$ smoking | lung cancer
  due to no path connecting them
- visit to Asia $\not\perp$ smoking | lung cancer, shortness of breath
  due to common descendant

ii) Given the probabilities we have multiple approaches of how to compute the different marginal/conditional distributions we are interested in.

- The first is via variable elimination as discussed in the lecture. Use this approach to compute the following distributions analytically:
  What is the probability of a patient showing up with dyspnea ($p(d)$)? How do these probabilities change if we know that he/she is a smoker ($p(d|s)$) or does not smoke ($p(d|\bar{s})$)?
- Another approach is to use simulations. A sample from this graphical model consists of eight binary variables. Assuming that we have a set of N samples, computing marginals then simply consists of simply counting how many samples have the desired values. Similarly computing conditional probabilities consists of counting the number of samples given some constraints. Implement a function that gives you a sample from this joint distribution. Compute $N = 100000$ samples and compare your numerical approximations to your analytical solutions for the three probabilities above.

Solution: The model is taken from Lauritzen & Spiegelhalter 1988[1].

We have that the the joint is given as

$$p(a, s, t, l, b, e, x, d) = p(a)p(s)p(t|a)p(l|s)p(s|b)p(e|t, l)p(x|e)p(d|e, b). \tag{1}$$

We first are interested in the marginal probability that a patient has dyspnea, i.e. need to compute

$$\begin{aligned} p(d) &= \sum_{a,s,t,l,b,e,x} p(a)p(s)p(t|a)p(l|s)p(b|s)p(e|t, l)p(x|e)p(d|e, b) && (2)\\ &= \sum_t \sum_l \sum_b \sum_e \sum_s p(s)p(l|s)p(b|s)p(e|t, l)p(d|e, b) \underbrace{\sum_a p(t|a)p(a)}_{=p(t)} \underbrace{\sum_x p(x|e)}_{=1} && (3)\\ &= \sum_b \sum_e p(d|e, b) \underbrace{\sum_l \underbrace{\sum_t p(t)p(e|t, l)}_{=:\phi(e,l)} \underbrace{\sum_s p(s)p(l|s)p(b|s)}_{=:\phi(l,b)}}_{=:\phi(e,b)}, && (4) \end{aligned}$$

where we have overloaded the notation, i.e. $\phi(e, l)$ is a different function from $\phi(l, b)$, etc. (similar to how we overload the notation of $p(\cdot)$). Computing the new terms in turn given the probabilities, we get

$$p(t) = \sum_a p(t|a)p(a) = 0.05 \cdot 0.01 + 0.01 \cdot 0.99 = 0.0104. \tag{5}$$

While for our $\phi(\cdot, \cdot)$ factors we get

$$\phi(l, b) = \sum_s p(s)p(l|s)p(b|s) = p(l|s)p(b|s)p(s) + p(l|\bar{s})p(b|\bar{s})p(\bar{s}) \Rightarrow [0.5265, 0.02345, 0.4185, 0.0315], \tag{6}$$

[1] "Local Computations with Probabilities on Graphical Structures and their Application to Expert Systems" https://www.jstor.org/stable/2345762?seq=1#metadata_info_tab_contents

where the vector gives $[\bar{l}\bar{b}, l\bar{b}, \bar{l}b, lb]$. Analogously we have

$$\phi(e,l) = \sum_t p(t)p(e|t,l) = p(e|t,l)p(t) + p(e|\bar{t},l)p(\bar{t}) \Rightarrow [0.9896, 0.0104, 0, 1]. \tag{7}$$

and

$$\phi(e,b) = \sum_l \phi(e,l)\phi(l,b) = \phi(e,l)\phi(l,b) + \phi(e,\bar{l})\phi(\bar{l},b) \Rightarrow [0.5210, 0.0289, 0.414, 0.0359]. \tag{8}$$

Lastly we get

$$\begin{aligned}
p(d) &= \sum_b \sum_e p(d|e,b)\phi(e,b) && (9)\\
&= p(d|\bar{e},\bar{b})\phi(\bar{e},\bar{b}) + p(d|\bar{e},b)\phi(\bar{e},b) + p(d|e,\bar{b})\phi(e,\bar{b}) + p(d|e,b)\phi(e,b) && (10)\\
&= 0.4358. && (11)
\end{aligned}$$

Now let's assume that we know the patient to be a smoker (the non-smoking case can be derived analogously). Here we compute

$$p(d|s) = \frac{p(d,s)}{p(s)} = \frac{p(s)\sum_{a,t,l,b,e,x} p(a)p(t|a)p(l|s)p(b|s)p(e|t,l)p(x|e)p(d|e,b)}{p(s)} \tag{12}$$

$$\begin{aligned}
p(d|s) &= \sum_{a,t,l,b,e,x} p(a)p(t|a)p(l|s)p(b|s)p(e|t,l)p(x|e)p(d|e,b) && (13)\\
&= \sum_t \sum_l \sum_b \sum_e p(l|s)p(b|s)p(e|t,l)p(d|e,b) \underbrace{\sum_a p(t|a)p(a)}_{=p(t)} \underbrace{\sum_x p(x|e)}_{=1} && (14)\\
&= \sum_b p(b|s) \underbrace{\sum_e p(d|e,b) \underbrace{\sum_l \underbrace{\sum_t p(t)p(e|t,l)}_{=:\phi(e,l)} p(l|s)}_{=:\phi_s(e)}}_{=:\phi(d,b)}. && (15)
\end{aligned}$$

$\phi(e,l)$ is given as above, and we get

$$\begin{aligned}
\phi_s(e) &= \sum_l \phi(e,l)p(l|s) = \phi(e,l)p(l|s) + \phi(e,\bar{l})p(\bar{l}|s) = 0.1094 \quad \Big(= p(e|s)\Big), && (16)\\
\phi(d,b) &= \sum_e p(d|e,b)\phi_s(e) = p(d|e,b)\phi_s(e) + p(d|\bar{e},b)\phi_s(\bar{e}) \Rightarrow [0.8344, 0.1656, 0.1891, 0.8109] && (17)
\end{aligned}$$

which gives us the result that

$$p(d|s) = \sum_b p(b|s)\phi(d,b) = \big(p(b|s)\phi(d,b) + p(\bar{b}|s)\phi(d,\bar{b})\big) = 0.5528. \tag{18}$$

If we know that we have a non-smoking case we get that $p(d|\bar{s}) = 0.319$.

iii) Using your simulator give numerical estimates for the probabilities in the following scenario:

According to our model, what is the marginal probability of a patient having lung cancer ($p(l)$)? A patient arrives complaining about shortness of breath. How does that change your estimate for that patient $p(l|d)$ having lung cancer? You decide to take some x-rays, which come back positive. What is the new $p(l|x,d)$? While waiting for the results you discovered that your patient just had a nice vacation in Asia and is a chain smoker. How do each of these new pieces of information to change your lung cancer estimate (i.e. what are your results for $p(l|x,d,a), p(l|x,d,s), p(l|x,d,s,a)$)? Also, discuss how many samples you get to estimate each of the probabilities. As you increase the number of conditions, do you observe any pattern and if so is it problematic or irrelevant?

Solution: See the jupyter notebook.

iv) *(technical +5pt)* Compute the analytical distributions using variable elimination for the distributions discussed in *iii)*. How do they compare?

Solution: The marginal probability of lung cancer is given as

$$p(l) = \sum_a p(a)p(t|a) \sum_x p(x|e) \sum_s \sum_t \sum_b \sum_e \sum_d p(s)p(l|s)p(s|b)p(e|t,l)p(d|e,b) \tag{19}$$

$$= \sum_b \sum_e \underbrace{\sum_s p(s)p(l|s)p(s|b)}_{=:\phi(l,b)} \underbrace{\sum_t p(t)p(e|t,l)}_{=:\phi(e,l)} \underbrace{\sum_d p(d|e,b)}_{=:1} = \sum_b \sum_e \phi(l,b)\phi(e,l) = 0.055. \tag{20}$$

If we assume that the patient comes with a breathing problems we have

$$p(l|d) = \frac{p(l,d)}{p(d)} = \frac{\sum_{a,s,t,b,e,x} p(a,s,t,l,b,e,x,d)}{p(d)} \tag{21}$$

$$= \frac{\sum_{a,s,t,b,e,x} p(a)p(s)p(t|a)p(l|s)p(s|b)p(e|t,l)p(x|e)p(d|e,b)}{p(d)} \tag{22}$$

$$= \frac{\sum_b \sum_e \phi(l,b)\phi(e,l)p(d|e,b)}{p(d)} = 0.1028. \tag{23}$$

Next, we assume that we have taken a set of x-rays which came back positive

$$p(l|d,x) = \frac{p(l,d,x)}{p(d,x)} = \frac{\sum_{a,s,t,b,e} p(a,s,t,l,b,e,x,d)}{p(d,x)} = 0.621, \tag{24}$$

which greatly increases our probability. The final three probabilities are given as

$$p(l|s,d,x) = 0.724 \tag{25}$$

$$p(l|a,d,x) = 0.444 \tag{26}$$

$$p(l|a,s,d,x) = 0.579. \tag{27}$$

As we increase the number of constraints, our sampling based approach gets worse and worse and requires more and more samples to get similar results to the analytical solutions.

2 HMM: Robot localization in a 1D world (10 pt)

The task of this exercise is robot localization following a similar setup to the one presented in the lecture. Our robot lives in a circular world consisting of 10 discrete locations (i.e. if the robot is in the tenth location and moves to the right it comes out at the first). The robot has access to a map of this world which shows that all of the locations consist of squares except for three, which are circular. The map is given as

$$[X, X, X, O, X, O, X, O, X, X]. \tag{28}$$

At each time step the robot tries to move either to the right or to the left with equal probability. This movement succeeds in 90% of the cases, while 5% of the time it does not move at all and the last 5% it moves a step in the opposite direction. After moving, its sensors try to read whether the current location contains a square or a circle. This sensor reading is correct 95% of the time.

Implement the algorithm to update the beliefs of the robot as to where it is given a sequence of actions $\mathbf{a}_T$ and sensor readings $\mathbf{r}_T$ and visualize how the beliefs $p(x_t|a_{1:t}, r_{1:t})$ as to where the robot is change from an initial belief that assigns equal probability ($p(x_0 = \text{loc l}) = 1/10$) as we get new information with each time step.

Solution: Before having observed anything, i.e. at time step $t = 0$, we have

$$p(x_0 = \text{Robot is at location l}) = 1/10. \tag{29}$$

In general after t time steps we have

$$p(x_t|a_{1:t}, r_{1:t}) = \frac{p(x_t, a_{1:t}, r_{1:t})}{p(a_{1:t}, r_{1:t})} = \frac{p(x_t, a_{1:t}, r_{1:t})}{\sum_{x_t} p(x_t, a_{1:t}, r_{1:t})} \tag{30}$$

i.e. we need the joint for which we proceed as discussed in the lecture.

$$\begin{align}
p(x_t, a_{1:t}, r_{1:t}) &= \int p(x_t, x_{t-1}, a_{1:t}, r_{1:t}) dx_{t-1} \tag{31}\\
&= \int p(x_t, r_t, a_t | x_{t-1}, \cancel{a}_{1:t-1}, \cancel{r}_{1:t-1}) p(x_{t-1}, a_{1:t-1}, r_{1:t-1}) dx_{t-1} \tag{32}\\
&= \int p(r_t|x_t, \cancel{a_t}, \cancel{x_{t-1}}) p(x_t|x_{t-1}, a_t) p(a_t| \cancel{x_{t-1}}) p(x_{t-1}, a_{1:t-1}, r_{1:t-1}) dx_{t-1} \tag{33}\\
&= p(r_t|x_t) \int p(x_t|x_{t-1}, a_t) p(x_{t-1}|a_{1:t-1}, r_{1:t-1}) dx_{t-1} p(a_t) p(a_{1:t-1}, r_{1:t-1}). \tag{34}
\end{align}$$

Notice that the $p(a_t)p(a_{1:t-1}, r_{1:t-1})$ factor appears both in the numerator as well as the denominator and cancels. Given that we work with discrete locations, the integral becomes

$$\int p(x_t|x_{t-1}, a_t) p(x_{t-1}|a_{1:t-1}, r_{1:t-1}) dx_{t-1} = \sum_{x_{t-1}} p(x_t|x_{t-1}, a_t) p(x_{t-1}|a_{1:t-1}, r_{1:t-1}), \tag{35}$$

which leaves us with an overall tractable set of operations. Given a state $x_t = l$ the transition probability is nonzero for exactly three states $l-1, l, l+1$, and the sum simplifies further to

$$\begin{align}
\sum_{x_{t-1}} p(x_t = l|x_{t-1}, a_t) p(x_{t-1}|a_{1:t-1}, r_{1:t-1}) &= p(x_t = l|x_{t-1} = l-1, a_t) p(x_{t-1} = l-1|a_{1:t-1}, r_{1:t-1}) \tag{36}\\
&+ p(x_t = l|x_{t-1} = l, a_t) p(x_{t-1} = l|a_{1:t-1}, r_{1:t-1}) \tag{37}\\
&+ p(x_t = l|x_{t-1} = l+1, a_t) p(x_{t-1} = l+1|a_{1:t-1}, r_{1:t-1}) \tag{38}
\end{align}$$

For the implementation see `practical09.ipynb`.