

87/100

Machine Learning Exercise Sheet 2

Yannick Zelle and Nina Fischer

October 26, 2021

Exercise 1

(a)

Given is the DAG G_1 with the following associated probabilities:

- $p(A) = 0.3$
- $p(B \mid A) = 0.2$
- $p(B \mid \neg A) = 0.4$
- $p(C \mid B) = 0.7$
- $p(C \mid \neg B) = 0.5$

From there we can deduct the following probabilities:

- $p(\neg A) = 0.7$
- $p(\neg B \mid A) = 0.8$
- $p(\neg B \mid \neg A) = 0.6$
- $p(\neg C \mid B) = 0.3$
- $p(\neg C \mid \neg B) = 0.5$

✓

From G_1 we can deduct the joint probability distribution:

$$p(A, B, C) = p(A) \cdot p(B \mid A) \cdot p(C \mid B) \quad \checkmark$$

Using this Probabilitydistribution we can calculate $P(B)$:

$$\begin{aligned}
p(B) &= \sum_{A,C} p(A, B=1, C) \\
&= p(A) \cdot p(B | A) \cdot p(C | B) \\
&\quad + p(\neg A) \cdot p(B | \neg A) \cdot p(C | B) \\
&\quad + p(\neg A) \cdot p(B | \neg A) \cdot p(\neg C | B) \\
&\quad + p(A) \cdot p(B | A) \cdot p(\neg C | B) \quad \checkmark \\
&= 0.3 \cdot 0.2 \cdot 0.7 \\
&\quad + 0.7 \cdot 0.4 \cdot 0.7 \\
&\quad + 0.7 \cdot 0.4 \cdot 0.3 \\
&\quad + 0.3 \cdot 0.2 \cdot 0.3 \\
&= 0.34 \quad \checkmark \quad \text{5 P}
\end{aligned}$$

and P(C):

$$\begin{aligned}
p(C) &= \sum_{A,B} p(A, B, C=1) \\
&= p(A) \cdot p(B | A) \cdot p(C | B) \\
&\quad + p(\neg A) \cdot p(B | \neg A) \cdot p(C | B) \\
&\quad + p(\neg A) \cdot p(\neg B | \neg A) \cdot p(C | \neg B) \quad \checkmark \\
&\quad + p(A) \cdot p(\neg B | A) \cdot p(C | \neg B) \\
&= 0.3 \cdot 0.2 \cdot 0.7 \\
&\quad + 0.7 \cdot 0.4 \cdot 0.7 \\
&\quad + 0.7 \cdot 0.6 \cdot 0.5 \\
&\quad + 0.3 \cdot 0.8 \cdot 0.5 \\
&= 0.568 \quad \checkmark \quad \text{5 P}
\end{aligned}$$

(b)

Given is a DAG G_2 and the associated probabilities:

- $p(A) = 0.3$
- $p(B | A) = 0.2$
- $p(B | \neg A) = 0.4$
- $p(c | A) = 0.7$
- $p(C | \neg A) = 0.6$
- $p(D | B, C) = 0.9$

- $p(D \mid B, \neg C) = 0.5$
- $p(D \mid \neg B, C) = 0.3$
- $p(D \mid \neg B, \neg C) = 0.3$

From there we can deduct the following probabilities:

- $p(\neg A) = 0.7$
- $p(\neg B \mid A) = 0.8$
- $p(\neg B \mid \neg A) = 0.6$
- $p(\neg C \mid A) = 0.3$
- $p(\neg C \mid \neg A) = 0.4$
- $p(\neg D \mid B, C) = 0.1$
- $p(\neg D \mid B, \neg C) = 0.5$
- $p(\neg D \mid \neg B, C) = 0.7$
- $p(\neg D \mid \neg B, \neg C) = 0.7$

From the G_2 we can deduct the following joint Probability distribution:

$$p(A, B, C, D) = p(A) \cdot p(C \mid A) \cdot p(B \mid A) \cdot p(D \mid B, C) \quad \checkmark$$

Using the joint Probability distribution and the probabilities from above, we can calculate $p(B)$:

$$\begin{aligned}
p(B) &= \sum_{A,C,D} p(A, B=1, C, D) \quad \checkmark = \sum_A p(B|A) \cdot p(A) \underbrace{\sum_C p(C|A)}_{=1} \\
&= p(A) \cdot p(C|A) \cdot p(B|A) \cdot p(D|B, C) \\
&+ p(A) \cdot p(C|A) \cdot p(B|A) \cdot p(\neg D|B, C) \\
&+ p(A) \cdot p(\neg C|A) \cdot p(B|A) \cdot p(D|B, \neg C) \\
&+ p(A) \cdot p(\neg C|A) \cdot p(B|A) \cdot p(\neg D|B, \neg C) \\
&+ p(\neg A) \cdot p(C|\neg A) \cdot p(B|\neg A) \cdot p(D|B, C) \\
&+ p(\neg A) \cdot p(C|\neg A) \cdot p(B|\neg A) \cdot p(\neg D|B, C) \\
&+ p(\neg A) \cdot p(\neg C|\neg A) \cdot p(B|\neg A) \cdot p(D|B, \neg C) \\
&+ p(\neg A) \cdot p(\neg C|\neg A) \cdot p(B|\neg A) \cdot p(\neg D|B, \neg C) \\
&= 0.7 \cdot 0.7 \cdot 0.2 \cdot 0.9 \\
&+ 0.7 \cdot 0.3 \cdot \cancel{0.8} \cdot 0.1 \\
&+ 0.7 \cdot 0.3 \cdot 0.2 \cdot 0.5 \\
&+ 0.7 \cdot 0.3 \cdot 0.2 \cdot 0.5 \\
&+ 0.3 \cdot 0.6 \cdot 0.4 \cdot 0.9 \\
&+ 0.3 \cdot 0.6 \cdot 0.4 \cdot 0.1 \quad f \\
&+ 0.3 \cdot 0.4 \cdot 0.4 \cdot 0.5 \\
&+ 0.3 \cdot 0.4 \cdot 0.4 \cdot 0.5 \\
&= 0.27 \quad 0.34 \quad 2/5p
\end{aligned}$$

and $p(D)$:

$$\begin{aligned}
p(D) &= \sum_{A,C,B} p(A, B, C, D = 1) \quad \checkmark \\
&= p(A) \cdot p(C | A) \cdot p(B | A) \cdot p(D | B, C) \\
&+ p(A) \cdot p(C | A) \cdot p(\neg B | A) \cdot p(D | \neg B, C) \\
&+ p(A) \cdot p(\neg C | A) \cdot p(B | A) \cdot p(D | B, \neg C) \\
&+ p(A) \cdot p(\neg C | A) \cdot p(\neg B | A) \cdot p(D | \neg B, \neg C) \quad \checkmark \\
&+ p(\neg A) \cdot p(C | \neg A) \cdot p(B | \neg A) \cdot p(D | B, C) \\
&+ p(\neg A) \cdot p(C | \neg A) \cdot p(\neg B | \neg A) \cdot p(D | \neg B, C) \\
&+ p(\neg A) \cdot p(\neg C | \neg A) \cdot p(B | \neg A) \cdot p(D | B, \neg C) \\
&+ p(\neg A) \cdot p(\neg C | \neg A) \cdot p(\neg B | \neg A) \cdot p(D | \neg B, \neg C) \\
&= 0.7 \cdot 0.7 \cdot 0.2 \cdot 0.9 \\
&+ 0.7 \cdot 0.7 \cdot 0.8 \cdot 0.3 \quad \text{9.7} \\
&+ 0.7 \cdot 0.3 \cdot 0.2 \cdot 0.5 \\
&+ 0.7 \cdot 0.3 \cdot 0.8 \cdot 0.3 \quad \text{f} \\
&+ 0.3 \cdot 0.6 \cdot 0.4 \cdot 0.9 \quad \text{f} \\
&+ 0.3 \cdot 0.6 \cdot 0.6 \cdot 0.3 \\
&+ 0.3 \cdot 0.4 \cdot 0.4 \cdot 0.5 \\
&+ 0.3 \cdot 0.4 \cdot 0.6 \cdot 0.3 \\
&= 0.36 \quad \text{f}
\end{aligned}$$

7/10 P

19 / 25 P

Exercise 2

(a)

$$p(A, B, C, D, E, F) = p(A) \cdot p(B | A) \cdot p(C | B) \cdot p(D | B) \cdot p(E | C, D) \cdot p(F | E) \quad \checkmark \quad 5P$$

(b)

$$p(A, B, C, D, E) = p(A) \cdot p(B | A) \cdot p(C | B) \cdot p(D | B) \cdot p(E | C, D) \quad \checkmark \quad 5P$$

(c)

Task: Find a minimal set S that d-separates A and F and prove that this is the case.

Proof. We propose $S = \{B\}$ to prove that the criteria of S holds we have actually to prove two statements:

1. S d-separates A and F \checkmark
2. there is no set with fewer elements that also d-separates A and F \checkmark

We will start with the first statement. Between A and F exist two paths:

$$\begin{aligned} p_1 &= A \rightarrow B \rightarrow C \rightarrow E \rightarrow F \\ p_2 &= A \rightarrow B \rightarrow D \rightarrow E \rightarrow F \end{aligned}$$

We have p_1 is blocked by S because with $i_k = B$ we have : $i_k \in S$ and $A \rightarrow B \rightarrow C \leftrightarrow i_{k-1} \rightarrow i_k \rightarrow i_{k+1}$ and p_2 is blocked by S because with $i_k = B$ we have : $i_k \in S$ and $A \rightarrow B \rightarrow D \leftrightarrow i_{k-1} \rightarrow i_k \rightarrow i_{k+1}$ It is left to show that there is no set with fewer elements that also d-seperates A and F. The only set that has fewer elements is the empty set but the empty set is neither blocking p_1 nor p_2 according to the definition. \square

10P

(d)

Proof. We will proof that $C \perp\!\!\!\perp_G D \mid B$ i.e B d-seperates C and D holds. This is the case if every path is blocked by $S = \{B\}$. There are two paths from C to D:

$$\begin{aligned} p_1 &= C \leftarrow B \rightarrow D \\ p_2 &= C \rightarrow E \leftarrow D \end{aligned}$$

p_1 is blocked by S because with $i_k = B$ we have $i_k \in S$ and $C \leftarrow B \rightarrow D \leftrightarrow i_{k-1} \leftrightarrow i_k \rightarrow i_{k+1}$. Also p_2 is blocked by S because with $i_k = E$ we have : $i_k \notin S$ and $C \rightarrow E \leftarrow D \leftrightarrow i_{k-1} \rightarrow i_k \leftarrow i_{k+1}$ So every path between C and D is blocked by S and therefore $C \perp\!\!\!\perp_G D \mid B$ holds. \square

5P / 25P

Exercise 3

(a)

We will proof that

$$A \rightarrow B \rightarrow C \implies A \perp\!\!\!\perp C \mid B$$

Proof.

$$\begin{aligned} p(A, B, C) &= p(A) \cdot p(B \mid A) \cdot p(C \mid B) \\ \Leftrightarrow p(A, B, C) &= p(A, B) \cdot p(C \mid B) \\ \Leftrightarrow p(A, B, C) &= p(A \mid B) \cdot p(B) \cdot p(C \mid B) \\ \Leftrightarrow \frac{p(A, B, C)}{p(B)} &= p(A \mid B) \cdot p(C \mid B) \\ \Leftrightarrow p(A, C \mid B) &= p(A \mid B) \cdot p(C \mid B) \\ \implies A \perp\!\!\!\perp C \mid B \end{aligned}$$

\square

(b)

We will proof that

$$A \leftarrow B \leftarrow C \implies A \perp\!\!\!\perp C \mid B$$

and we know from a that $p(C) \cdot p(B \mid C) \Leftrightarrow p(B) \cdot p(C \mid B)$. ✓

Proof.

$$\begin{aligned} p(A, B, C) &= p(A \mid B) \cdot p(B \mid C) \cdot p(C) \quad \checkmark \\ \Leftrightarrow p(A, B, C) &= p(A \mid B) \cdot p(C \mid B) \cdot p(B) \\ \Leftrightarrow \frac{p(A, B, C)}{p(B)} &= p(A \mid B) \cdot p(C \mid B) \quad \checkmark \\ \Leftrightarrow p(A, C \mid B) &= p(A \mid B) \cdot p(C \mid B) \quad \checkmark \\ \implies A \perp\!\!\!\perp C \mid B &\quad \checkmark \quad \text{SP} \end{aligned}$$

□

(c)

We will proof that

$$A \leftarrow B \rightarrow C \implies A \perp\!\!\!\perp C \mid B$$

Proof.

$$\begin{aligned} p(A, B, C) &= p(A \mid B) \cdot p(B) \cdot p(C \mid B) \quad \checkmark \\ \Leftrightarrow \frac{p(A, B, C)}{p(B)} &= p(A \mid B) \cdot p(C \mid B) \quad \checkmark \\ \Leftrightarrow p(A, C \mid B) &= p(A \mid B) \cdot p(C \mid B) \quad \checkmark \\ \implies A \perp\!\!\!\perp C \mid B &\quad \checkmark \quad \text{SP} \end{aligned}$$

□

(d)

We will proof that

$$A \rightarrow B \leftarrow C \implies A \perp\!\!\!\perp C$$

Proof.

$$p(A, B, C) = p(A) \cdot p(B | A, C) \cdot p(C) \quad \checkmark$$

$$\Leftrightarrow \frac{p(A, B, C)}{p(B | A, C)} = p(A) \cdot p(C) \quad \checkmark$$


$$\Leftrightarrow \frac{p(A, B, C)}{\frac{p(A, B, C)}{p(A, C)}} = p(A) \cdot p(C) \quad \checkmark$$

$$\Leftrightarrow p(A, C) = p(A) \cdot p(C) \quad \checkmark$$

$$\Rightarrow A \perp\!\!\!\perp C$$

\checkmark

SP

□  20 P

Exercise 4

(a)

We will proof that $E(a \cdot X + Y) = a \cdot E(X) + E(Y)$ with the sum rule and the factor rule.

Proof.

$\rightarrow \sum_x \sum_y (ax+y) p(x,y)$ discrete values

$$\begin{aligned} E(a \cdot X + Y) &= \int a \cdot X + Y dx \\ &= \int a \cdot X dx + \int Y dx \\ &= a \cdot \int X dx + \int Y dx \\ &= a \cdot E(X) + E(Y) \end{aligned}$$

7/5

□

(b)

We will proof that $Var(a \cdot X) = a^2 \cdot Var(X)$ with the result of a.

Proof.

$$\begin{aligned} Var(a \cdot X) &= E(a \cdot X - \mu)^2 \\ &= E(a^2 \cdot (X - \mu)^2) \\ &= a^2 \cdot E(X - \mu)^2 \\ &= a^2 \cdot Var(X) \end{aligned}$$

SP

□  6

Exercise 5

In this Exercise we attempt to show that the mean μ is also the mode of the Gaussian normal distribution:

$$N(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \checkmark$$

Proof. The mode of N is the point where N obtains its maximum value. We will therefore search for the point where N is maximal. This is the case only if $\frac{dN}{dx} = 0$

✓ easier with $\log N$ (as in the hint)

$$\frac{dN}{dx} = \frac{-2(x-\mu)}{2\sigma^2\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Since $e^x > 0$ for all x this function becomes only 0 if $x = \mu$. We can also deduce that for $x \in (-\infty, \mu] : \frac{-2(x-\mu)}{2\sigma^2\sqrt{2\pi\sigma^2}} \geq 0$ and for $x \in [\mu, \infty) : \frac{-2(x-\mu)}{2\sigma^2\sqrt{2\pi\sigma^2}} \leq 0$, since $e^x > 0$ this determines the behaviour of the derivative and there for $(\mu, N(\mu \mid \mu, \sigma))$ is the maximum.

✓

□

10 P

Exercise 6

Our Implementation is in the Jupyter Notebook "Exercise 6"

Please submit your outcomes next time as well

histograms are missing

7 P