Machine Learning Exercise Sheet 2

Yannick Zelle and Nina Fischer

October 30, 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:

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

Using this Probabilitydistribution we can calculate P(B):

$$
\begin{aligned}
p(B) &= \sum_{A,C} p(A, B=1, C) \\
&= p(A) \cdot p(B \mid A) \cdot p(C \mid B) \\
&+ p(\neg A) \cdot p(B \mid \neg A) \cdot p(C \mid B) \\
&+ p(\neg A) \cdot p(B \mid \neg A) \cdot p(\neg C \mid B) \\
&+ p(A) \cdot p(B \mid A) \cdot p(\neg C \mid B) \\
&= 0.3 \cdot 0.2 \cdot 0.7 \\
&+ 0.7 \cdot 0.4 \cdot 0.7 \\
&+ 0.7 \cdot 0.4 \cdot 0.3 \\
&+ 0.3 \cdot 0.2 \cdot 0.3 \\
&= 0.34
\end{aligned}
$$

and P(C):

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

(b)

Given is a DAG G_2 and the associated probabilities:

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

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

From there we can deductthe 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 Probabilitydistribution:

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

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

$$
\begin{aligned}
p(B) &= \sum_{A,C,D} p(A, B=1, C, D) \\
&= p(A) \cdot p(C \mid A) \cdot p(B \mid A) \cdot p(D \mid B, C) \\
&+ p(A) \cdot p(C \mid A) \cdot p(B \mid A) \cdot p(\neg D \mid B, C) \\
&+ p(A) \cdot p(\neg C \mid A) \cdot p(B \mid A) \cdot p(D \mid B, \neg C) \\
&+ p(A) \cdot p(\neg C \mid A) \cdot p(B \mid A) \cdot p(\neg D \mid B, \neg C) \\
&+ p(\neg A) \cdot p(C \mid \neg A) \cdot p(B \mid \neg A) \cdot p(D \mid B, C) \\
&+ p(\neg A) \cdot p(C \mid \neg A) \cdot p(B \mid \neg A) \cdot p(\neg D \mid B, C) \\
&+ p(\neg A) \cdot p(\neg C \mid \neg A) \cdot p(B \mid \neg A) \cdot p(D \mid B, \neg C) \\
&+ p(\neg A) \cdot p(\neg C \mid \neg A) \cdot p(B \mid \neg A) \cdot p(\neg D \mid B, \neg C) \\
&= 0.7 \cdot 0.7 \cdot 0.2 \cdot 0.9 \\
&+ 0.7 \cdot 0.3 \cdot 0.8 \cdot 0.1 \\
&+ 0.7 \cdot 0.3 \cdot 0.2 \cdot 0.5 \\
&+ 0.7 \cdot 0.3 \cdot 0.2 \cdot 0.5 \\
&+ 0.3 \cdot 0.6 \cdot 0.4 \cdot 0.9 \\
&+ 0.3 \cdot 0.6 \cdot 0.4 \cdot 0.1 \\
&+ 0.3 \cdot 0.4 \cdot 0.4 \cdot 0.5 \\
&+ 0.3 \cdot 0.4 \cdot 0.4 \cdot 0.5 \\
&= 0.27
\end{aligned}
$$

and p(D):

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

Exercise 2

(a)

$p(A,B,C,D,E,F) = p(A)\cdot p(B \mid A)\cdot p(C \mid B)\cdot p(D \mid B)\cdot p(E \mid C,D)\cdot p(F \mid E)$

(b)

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

(c)

Task: Find a minimal set S that d-seperates 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-seperates A and F
2. there is no set with fewer elements that also d-seperates A and F

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

$$p_1 = A \to B \to C \to E \to F$$
$$p_2 = A \to B \to D \to E \to F$$

We have p_1 is blocked by S because with $i_k = B$ we have : $i_k \in S$ and $A \to B \to C \leftrightarrow i_{k-1} \to i_k \to i_{k+1}$ and p_2 is blocked by S because with $i_k = B$ we have : $i_k \in S$ and $A \to B \to D \leftrightarrow i_{k-1} \to i_k \to i_{k+1}$ 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. □

(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:

$$p_1 = C \leftarrow B \to D$$
$$p_2 = C \to E \leftarrow D$$

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

Exercise 3

(a)

We will proof that

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

Proof.

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

□

(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.

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

□

(c)

We will proof that

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

Proof.

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

□

(d)

We will proof that

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

Proof.

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

□

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.

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

□

(b)

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

Proof.

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

□

Exercise 5

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

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

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

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

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

$\square$

Exercise 6

Our Implementation is in the Jupyter Notebook ”Exercise 6”