My solutions to

Deep Learning: Foundations and Concepts

Dario Miro Konopatzki

11 Structured Distributions

11.1

$$\begin{aligned}
\int p(x_1, ..., x_K) dx &= \int \prod_{k=1}^{K} p(x_k|\mathrm{pa}(k)) dx \\
&= \int p(x_1)p(x_2|\mathrm{pa}(2)) \dots p(x_K|\mathrm{pa}(x_K)) d(x_K, \dots, x_1) \\
&\overset{(\star)}{=} \int p(x_1) \int p(x_2|\mathrm{pa}(2)) \dots \underbrace{\int p(x_K|\mathrm{pa}(x_K)) dx_K}_{=1 \text{ by assumption}} \dots dx_1 \\
&= \int p(x_1) \dots \int p(x_{K-1}|\mathrm{pa}(x_{K-1})) \cdot 1 \, dx_{K-1} \dots dx_1 \\
&= \int p(x_1) \dots \underbrace{\int p(x_{K-1}|\mathrm{pa}(x_{K-1}))\, dx_{K-1}}_{=1 \text{ by assumption}} \dots dx_1 \\
&\overset{\text{iterate}}{=} 1
\end{aligned}$$

Where in $(\star)$ one uses that for any given l, only lower-numbered nodes are in $\mathrm{pa}(l)$.

11.2

We consider the ordered nodes. If there existed a directed cycle $x_{n_1} \to x_{n_2} \dots \to x_{n_L} \to x_{n_1}$, then $n_1 < n_2 \dots < n_L < n_1$ because there are no links from higher- to lower-numbered nodes, but $n_1 < n_1$ is a contradiction.

11.3

$$
\begin{aligned}
p(a=0) &= \sum_{b,c} p(0,b,c) \\
&= p(0,0,0) + p(0,0,1) + p(0,1,0) + p(0,1,1) \\
&= 0.192 + 0.144 + 0.048 + 0.216 \\
&= 0.6 \\
p(b=0) &= \sum_{a,c} p(a,0,c) \\
&= p(0,0,0) + p(0,0,1) + p(1,0,0) + p(1,0,1) \\
&= 0.192 + 0.144 + 0.192 + 0.064 \\
&= 0.592 \\
p(c=0) &= \sum_{a,b} p(a,b,0) \\
&= p(0,0,0) + p(0,1,0) + p(1,0,0) + p(1,1,0) \\
&= 0.192 + 0.048 + 0.192 + 0.048 \\
&= 0.48 \\
p(c=1) &= 1 - p(c=0) \\
&= 0.52
\end{aligned}
$$

$$
\begin{aligned}
p(a=0,b=0) &= \sum_{c} p(0,0,c) \\
&= p(0,0,0) + p(0,0,1) \\
&= 0.192 + 0.144 \\
&= 0.336 \\
p(a=0,c=0) &= \sum_{b} p(0,b,0) \\
&= p(0,0,0) + p(0,1,0) \\
&= 0.192 + 0.048 \\
&= 0.24 \\
p(a=0,c=1) &= \sum_{b} p(0,b,1) \\
&= p(0,0,1) + p(0,1,1) \\
&= 0.144 + 0.216 \\
&= 0.36
\end{aligned}
$$

$$
\begin{aligned}
p(a=1,c=0) &= \sum_b p(1,b,0) \\
&= p(1,0,0)+p(1,1,0) \\
&= 0.192+0.048 \\
&= 0.24 \\
p(b=0,c=0) &= \sum_a p(a,0,0) \\
&= p(0,0,0)+p(1,0,0) \\
&= 0.192+0.192 \\
&= 0.384 \\
p(b=0,c=1) &= \sum_a p(a,0,1) \\
&= p(0,0,1)+p(1,0,1) \\
&= 0.144+0.064 \\
&= 0.208 \\
p(b=1,c=0) &= \sum_a p(a,1,0) \\
&= p(0,1,0)+p(1,1,0) \\
&= 0.048+0.048 \\
&= 0.096
\end{aligned}
$$

$$
\begin{aligned}
p(a=0|c=0) &= \frac{p(a=0,c=0)}{p(c=0)} \\
&= \frac{0.24}{0.48} \\
&= 0.5 \\
p(a=1|c=0) &= 1-p(a=0|c=0) \\
&= 1-0.5 \\
&= 0.5 \\
p(a=0|c=1) &= \frac{p(a=0,c=1)}{p(c=1)} \\
&= \frac{0.36}{0.52} \\
&= \frac{9}{13} \\
p(a=1|c=1) &= 1-p(a=0|c=1)
\end{aligned}
$$

$$
\begin{aligned}
&= 1 - \frac{9}{13} \\
&= \frac{4}{13} \\
p(b=0|c=0) &= \frac{p(b=0, c=0)}{p(c=0)} \\
&= \frac{0.384}{0.48} \\
&= 0.8 \\
p(b=1|c=0) &= 1 - p(b=0|c=0) \\
&= 1 - 0.8 \\
&= 0.2 \\
p(b=0|c=1) &= \frac{p(b=0, c=1)}{p(c=1)} \\
&= \frac{0.208}{0.52} \\
&= 0.4 \\
p(b=1|c=1) &= 1 - p(b=0|c=1) \\
&= 1 - 0.4 \\
&= 0.6
\end{aligned}
$$

$$
\begin{aligned}
p(a=0, b=0|c=0) &= \frac{p(0,0,0)}{p(c=0)} \\
&= \frac{0.192}{0.48} \\
&= 0.4 \\
&= 0.5 \cdot 0.8 \\
&= p(a=0|c=0)p(b=0|c=0) \\
p(a=0, b=1|c=0) &= \frac{p(0,1,0)}{p(c=0)} \\
&= \frac{0.048}{0.48} \\
&= 0.1 \\
&= 0.5 \cdot 0.2 \\
&= p(a=0|c=0)p(b=1|c=0) \\
p(a=1, b=0|c=0) &= \frac{p(1,0,0)}{p(c=0)}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{0.192}{0.48} \\
&= 0.4 \\
&= 0.5 \cdot 0.8 \\
&= p(a=1|c=0)p(b=0|c=0) \\
p(a=1,b=1|c=0) &= \frac{p(1,1,0)}{p(c=0)} \\
&= \frac{0.048}{0.48} \\
&= 0.1 \\
&= 0.5 \cdot 0.2 \\
&= p(a=1|c=0)p(b=1|c=0)
\end{aligned}
$$

$$
\begin{aligned}
p(a=0,b=0|c=1) &= \frac{p(0,0,1)}{p(c=1)} \\
&= \frac{0.144}{0.52} \\
&= \frac{18}{65} \\
&= \frac{9}{13} \cdot 0.4 \\
&= p(a=0|c=1)p(b=0|c=1) \\
p(a=0,b=1|c=1) &= \frac{p(0,1,1)}{p(c=1)} \\
&= \frac{0.216}{0.52} \\
&= \frac{27}{65} \\
&= \frac{9}{13} \cdot 0.6 \\
&= p(a=0|c=1)p(b=1|c=1) \\
p(a=1,b=0|c=1) &= \frac{p(1,0,1)}{p(c=1)} \\
&= \frac{0.064}{0.52} \\
&= \frac{8}{65} \\
&= \frac{4}{13} \cdot 0.4
\end{aligned}
$$

$$
\begin{aligned}
&= p(a=1|c=1)p(b=0|c=1) \\
p(a=1,b=1|c=1) &= \frac{p(1,1,1)}{p(c=1)} \\
&= \frac{0.096}{0.52} \\
&= \frac{12}{65} \\
&= \frac{4}{13} \cdot 0.6 \\
&= p(a=1|c=1)p(b=1|c=1)
\end{aligned}
$$

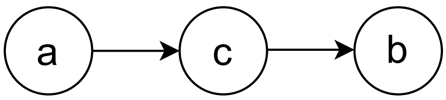
11.4

$$
\begin{aligned}
p(a=1) &= 1 - p(a=0) \\
&= 1 - 0.6 \\
&= 0.4
\end{aligned}
$$

$$
\begin{aligned}
p(c=0|a=0) &= \frac{p(a=0,c=0)}{p(a=0)} \\
&= \frac{0.24}{0.6} \\
&= 0.4 \\
p(c=1|a=0) &= 1 - p(c=0|a=0) \\
&= 1 - 0.4 \\
&= 0.6 \\
p(c=0|a=1) &= \frac{p(a=1,c=0)}{p(a=1)} \\
&= \frac{0.24}{0.4} \\
&= 0.6 \\
p(c=1|a=1) &= 1 - p(c=0|a=1) \\
&= 1 - 0.6 \\
&= 0.4
\end{aligned}
$$

$$
p(0,0,0) = 0.192
$$

$$
\begin{aligned}
&= 0.6 \cdot 0.4 \cdot 0.8 \\
&= p(a=0)p(c=0|a=0)p(b=0|c=0) \\
p(0,0,1) &= 0.144 \\
&= 0.6 \cdot 0.6 \cdot 0.4 \\
&= p(a=0)p(c=1|a=0)p(b=0|c=1) \\
p(0,1,0) &= 0.048 \\
&= 0.6 \cdot 0.4 \cdot 0.2 \\
&= p(a=0)p(c=0|a=0)p(b=1|c=0) \\
p(0,1,1) &= 0.216 \\
&= 0.6 \cdot 0.6 \cdot 0.6 \\
&= p(a=0)p(c=1|a=0)p(b=1|c=1) \\
p(1,0,0) &= 0.192 \\
&= 0.4 \cdot 0.6 \cdot 0.8 \\
&= p(a=1)p(c=0|a=1)p(b=0|c=0) \\
p(1,0,1) &= 0.064 \\
&= 0.4 \cdot 0.4 \cdot 0.4 \\
&= p(a=1)p(c=1|a=1)p(b=0|c=1) \\
p(1,1,0) &= 0.048 \\
&= 0.4 \cdot 0.6 \cdot 0.2 \\
&= p(a=1)p(c=0|a=1)p(b=1|c=0) \\
p(1,1,1) &= 0.096 \\
&= 0.4 \cdot 0.4 \cdot 0.6 \\
&= p(a=1)p(c=1|a=1)p(b=1|c=1)
\end{aligned}
$$



Directed graph corresponding to the factorization from 11.4.

11.5

Consider the case where $x_i = 0$ for all i:

$$\begin{aligned}
p(y=1|\forall i : x_i = 0) &= 1 - (1-\mu_0)\prod_{i=1}^{M}(1-\mu_i)^0 \\
&= 1 - (1-\mu_0)\prod_{i=1}^{M} 1 \\
&= 1 - (1-\mu_0)\cdot 1 \\
&= 1 - 1 + \mu_0 \\
&= \mu_0 \qquad (\star)
\end{aligned}$$

Let $A := \{i : x_i = 1\}$. Then

$$\begin{aligned}
p(y=1|x) &= 1 - (1-\mu_0)\prod_{i=1}^{M}(1-\mu_i)^{x_i} \\
&= 1 - p(y=0|x=0)\prod_{i=1}^{M} p(x_i=0)^{x_i} \quad \text{by } (\star) \text{ \& def. of } \mu \\
&= 1 - p(y=0|x=0)\prod_{i\in A} p(x_i=0)^1 \prod_{i\in A^\complement} p(x_i=0)^0 \\
&= 1 - p(y=0|x=0)\prod_{i\in A} p(x_i=0) \prod_{i\in A^\complement} 1 \\
&= 1 - p(y=0|x=0)\prod_{i\in A} p(x_i=0)
\end{aligned}$$

Hence, observing additional $(x_i = 1)$ introduces corresponding factors $0 \le p(x_i = 0) \le 1$ into the second term, potentially increasing but never decreasing $p(y = 1|x)$.

The interpretation of y as 'noisy or' makes sense because:

- The probability $p(y = 1)$ of observing the effect is lowest if none of the causes occur (i.e. if $(x = 0)$).
- Observing additional causes does potentially increase, but never decrease, the probability of observing the effect.
- The probability of observing the effect is highest if all of the causes occur (i.e. if $(x = 1)$).

- The probability of observing the effect can be non-zero even if none of the causes occur (and this probability is given by μ_0), and the effect will not in general be observed even if all causes occur. This is what makes y 'noisy'.

11.11

$$\begin{aligned}
p(a, b|d) &= \int p(a, b, c|d)dc \\
&= \int p(a|d)p(b, c|d)dc && \text{by assumption } a \perp\!\!\!\perp b, c \mid d \\
&= p(a|d) \int p(b, c|d)dc \\
&= p(a|d)p(b|d)
\end{aligned}$$

11.12

Let $A = \{x\}$, C Markov blanket of x, and B the remaining variables. Consider any path from x to any node in B. Such a path has to pass through some node $c \in C$.

1. If c parent of x and
    - path via child of c, then tail-to-tail at c.
    - path via parent of c, then head-to-tail at c.
2. If c child of x and
    - path via child of c, then head-to-tail at c.
    - path via parent d of c, then d co-parent of x (i.p. $d \in C$) and one continues by considering 3. below for d.
3. c co-parent of x and
    - path via child of c, then tail-to-tail at c.
    - path via parent of c, then head-to-tail at c

By the above, and since $c \in C$, the path is blocked. Because it was chosen arbitrarily, all paths are blocked. Hence $\{x\}$ is d-separated from B by C.

11.13

$\boldsymbol{a} \perp\!\!\!\perp \boldsymbol{b} \mid \emptyset$

$$
\begin{aligned}
p(a, b) &= \sum_{\{c,d\}} p(a, b, c, d) \\
&= \sum_{\{c,d\}} p(a)p(b)p(c|a, b)p(d|c) \qquad \text{graph structure} \\
&= p(a)p(b) \sum_{c} p(c|a, b) \sum_{d} p(d|c) \\
&= p(a)p(b) \sum_{c} (p(c|a, b) \cdot 1) \\
&= p(a)p(b)
\end{aligned}
$$

$\boldsymbol{a} \not\perp\!\!\!\perp \boldsymbol{b} \mid \boldsymbol{d}$

Consider the distribution and results from exercise 11.14 below. Then

$$
\begin{aligned}
p(B = 0, F = 0|D = 0) &= \frac{p(B = 0, F = 0, D = 0)}{p(D = 0)} \\
&= \frac{p(F = 0|B = 0, D = 0)p(B = 0, D = 0)}{p(D = 0)} \\
&= \frac{\frac{41}{374} \cdot 0.0748}{0.352} \\
&= \frac{41}{1760} \\
&\neq \frac{289}{6400} \\
&= \frac{0.0748}{0.352} \cdot 0.2125 \\
&= \frac{p(B = 0, D = 0)}{p(D = 0)} p(F = 0|D = 0) \\
&= p(B = 0|D = 0)p(F = 0|D = 0)
\end{aligned}
$$

11.14

Auxiliary Calculations

$$\begin{aligned}
p(B=0) &= 1 - p(B=1) \\
&= 1 - 0.9 \\
&= 0.1 \\
p(F=0) &= 1 - p(F=1) \\
&= 1 - 0.9 \\
&= 0.1 \\
p(D=0) &= \sum_G p(D=0|G)p(G) \\
&= p(D=0|G=0)p(G=0) \\
&\quad + p(D=0|G=1)p(G=1) \\
&= p(D=0|G=0)p(G=0) \\
&\quad + (1 - p(D=1|G=1))(1 - p(G=0)) \\
&= 0.9 \cdot 0.315 + (1 - 0.9) \cdot (1 - 0.315) \\
&= 0.352
\end{aligned}$$

$$\begin{aligned}
p(G=1|F=0) &= 1 - p(G=0|F=0) \\
&= 1 - 0.81 \\
&= 0.19 \\
p(D=0|G=1) &= 1 - p(D=1|G=1) \\
&= 1 - 0.9 \\
&= 0.1
\end{aligned}$$

$$\begin{aligned}
p(G=0|B=0,F=0) &= 1 - p(G=1|B=0,F=0) \\
&= 1 - 0.1 \\
&= 0.9
\end{aligned}$$

$$\begin{aligned}
p(G=0|B=0,F=1) &= 1 - p(G=1|B=0,F=1) \\
&= 1 - 0.2 \\
&= 0.8
\end{aligned}$$

$$p(B=0,D=0) = \sum_{F,G} p(B=0,F,G,D=0)$$

$$
\begin{aligned}
&= \sum_{F,G} p(B=0)p(F)p(G|B=0,F)p(D=0|G) \\
&= p(B=0) \sum_{F,G} p(F)p(G|B=0,F)p(D=0|G) \\
&= p(B=0) \sum_{F} p(F) \underbrace{\sum_{G} p(G|B=0,F)p(D=0|G)}_{\star} \\
&= p(B=0)(p(F=0) \cdot \star_{[F=0]} + p(F=1) \cdot \star_{[F=1]}) \\
&= 0.1 \cdot (0.1 \cdot 0.82 + 0.9 \cdot 0.74) \\
&= 0.0748
\end{aligned}
$$

$$
\begin{aligned}
\star_{[F=0]} &= p(G=0|B=0,F=0)p(D=0|G=0) \\
&\quad + p(G=1|B=0,F=0)p(D=0|G=1) \\
&= 0.9 \cdot .9 + 0.1 \cdot 0.1 \\
&= 0.82 \\
\star_{[F=1]} &= p(G=0|B=0,F=1)p(D=0|G=0) \\
&\quad + p(G=1|B=0,F=1)p(D=0|G=1) \\
&= 0.8 \cdot 0.9 + 0.2 \cdot 0.1 \\
&= 0.74
\end{aligned}
$$

Part 1

$$
\begin{aligned}
p(F=0|D=0) &= \sum_{G} p(F=0,G|D=0) \\
&= \frac{1}{p(D=0)} \sum_{G} p(F=0,G,D=0) \\
&= \frac{p(F=0)}{p(D=0)} \sum_{G} p(G|F=0)p(D=0|G) \\
&= \frac{p(F=0)}{p(D=0)}(p(G=0|F=0)p(D=0|G=0) \\
&\quad + p(G=1|F=0)p(D=0|G=1)) \\
&= \frac{0.1}{0.352}(0.81 \cdot 0.9 + 0.19 \cdot 0.1) \\
&= 0.2125
\end{aligned}
$$

Part 2

$$\begin{aligned}
p(F=0|B=0,D=0) &= \sum_G p(F=0,G|B=0,D=0) \\
&= \frac{1}{p(B=0,D=0)} \sum_G p(F=0,G,B=0,D=0) \\
&= \frac{1}{p(B=0,D=0)} \sum_G p(B=0)p(F=0) \\
&\quad \cdot p(G|B=0,F=0)p(D=0|G) \\
&= \frac{p(B=0)p(F=0)}{p(B=0,D=0)} \\
&\quad \cdot \sum_G p(G|B=0,F=0)p(D=0|G) \\
&= \frac{p(B=0)p(F=0)}{p(B=0,D=0)} \\
&\quad \cdot (p(G=0|B=0,F=0)p(D=0|G=0) \\
&\quad + p(G=1|B=0,F=0)p(D=0|G=1)) \\
&= \frac{0.1 \cdot 0.1}{0.0748}(0.9 \cdot 0.9 + 0.1 \cdot 0.1) \\
&\approx 0.109 \\
&< 0.2125 \qquad \text{Part 1} \\
&= p(F=0|D=0)
\end{aligned}$$

- Observing $(D=0)$ increases probability that $(F=0)$ compared to prior probability $p(F=0)$.
- Observing $(B=0)$ explains away observation $(D=0)$.