My solutions to

Deep Learning: Foundations and Concepts

Dario Miro Konopatzki

2 Probabilities

2.1

$$\begin{aligned}
p(C=1|T=1) &= \frac{p(T=1|C=1)p(C=1)}{p(T=1)} && \text{Bayes' theorem} \\
&= \frac{p(T=1|C=1)p(C=1)}{p(T=1,C=0)+p(T=1,C=1)} && \text{sum rule} \\
&= \frac{p(T=1|C=1)p(C=1)}{p(T=1|C=0)p(C=0)+p(T=1|C=1)p(C=1)} && \text{product rule} \\
&= \frac{0.9\cdot 0.001}{0.03\cdot(1-0.001)+0.9\cdot 0.001} \\
&\approx 0.029
\end{aligned}$$

2.2

Let Y denote the yellow die, B the blue die, G the green die and R the red die. We consider throws of pairs of independent dice, i.e. $p(D_1, D_2) = p(D_1)p(D_2)$. Each die takes on a unique value in a given throw, such that e.g. $(G = 5) := (G = 5, (B = 0 \text{ or } B = 4))$ and $(G = 1, B = 0)$ are mutually exclusive events, hence $p(G = 5 \text{ or } (G = 1, B = 0)) = P(G = 5) + P(G = 1, B = 0)$.

$$\begin{aligned}
p(B>Y) &= p(B=4, Y=3) \\
&= p(B=4)p(Y=3) \\
&= \frac{4}{6}\cdot\frac{6}{6} \\
&= \frac{2}{3}
\end{aligned}$$

$$
\begin{aligned}
p(G > B) &= p(G = 5 \text{ or } (G = 1, B = 0)) \\
&= p(G = 5) + p(G = 1)p(B = 0) \\
&= \frac{3}{6} + \frac{3}{6} \cdot \frac{2}{6} \\
&= \frac{2}{3} \\
p(R > G) &= p(R = 6 \text{ or } (R = 2, G = 1)) \\
&= p(R = 6) + p(R = 2)p(G = 1) \\
&= \frac{2}{6} + \frac{4}{6} \cdot \frac{3}{6} \\
&= \frac{2}{3} \\
p(Y > R) &= p(Y = 3, R = 2) \\
&= p(Y = 3)p(R = 2) \\
&= \frac{6}{6} \cdot \frac{4}{6} \\
&= \frac{2}{3}
\end{aligned}
$$

2.4

$$
\begin{aligned}
\int_c^d p(x)dx &= \int_c^d \frac{1}{d-c} dx \\
&= \frac{1}{d-c} \int_c^d dx \\
&= \frac{1}{d-c} [x]_c^d \\
&= \frac{1}{d-c}(d-c) \\
&= 1 \\
\mathbb{E}_p[X] &= \int_c^d xp(x)dx \\
&= \int_c^d x \frac{1}{d-c} dx \\
&= \frac{1}{d-c} \int_c^d xdx \\
&= \frac{1}{d-c} \left[\frac{1}{2}x^2\right]_c^d
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{2(d-c)}(d^2-c^2) \\
&= \frac{1}{2(d-c)}(d-c)(d+c) \\
&= \frac{d+c}{2} \\
\mathbb{E}_p[X^2] &= \int_c^d x^2 p(x) dx \\
&= \int_c^d x^2 \frac{1}{d-c} dx \\
&= \frac{1}{d-c} \int_c^d x^2 dx \\
&= \frac{1}{d-c} \left[\frac{1}{3}x^3\right]_c^d \\
&= \frac{1}{3(d-c)}(d^3-c^3) \\
&= \frac{1}{3(d-c)}(d-c)(d^2+c^2+cd) \\
&= \frac{1}{3}(d^2+c^2+cd) \\
\mathbb{E}_p[X]^2 &= \left(\frac{d+c}{2}\right)^2 \\
&= \frac{d^2+2cd+c^2}{4} \\
\mathbb{V}_p[X] &= \mathbb{E}_p[X^2] - \mathbb{E}_p[X]^2 \\
&= \frac{1}{3}(d^2+c^2+cd) - \frac{(d+c)^2}{2^2} \\
&= \frac{1}{3}(d^2+c^2+cd) - \frac{d^2+2cd+c^2}{4} \\
&= \frac{1}{12}(4d^2+4c^2+4cd-3d^2-6cd-3c^2) \\
&= \frac{1}{12}(d^2-2cd+c^2) \\
&= \frac{1}{12}(d-c)^2
\end{aligned}
$$

2.5

Exponential distribution

$$
\begin{aligned}
\int p(x|\lambda)dx &= \int_0^\infty \lambda e^{-\lambda x} dx \\
&= \lambda \int_0^\infty e^{-\lambda x} dx \\
&= \lambda \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_0^\infty \\
&= \lambda \left[0 - \left(-\frac{1}{\lambda} e^{-\lambda \cdot 0} \right) \right] \\
&= \lambda \cdot \frac{1}{\lambda} \\
&= 1
\end{aligned}
$$

Laplace distribution

$$
\begin{aligned}
\int p(x|\mu, \gamma) &= \int_{-\infty}^\infty \frac{1}{2\gamma} e^{-\frac{|x-\mu|}{\gamma}} dx \\
&= \frac{1}{2\gamma} \int_{-\infty}^\infty e^{-\frac{|x-\mu|}{\gamma}} dx \\
&= \frac{1}{2\gamma} \left(\int_{-\infty}^\mu e^{-\frac{|x-\mu|}{\gamma}} dx + \int_\mu^\infty \frac{1}{\gamma} e^{-\frac{|x-\mu|}{\gamma}} dx \right) \\
&= \frac{1}{2\gamma} \left(\int_{-\infty}^\mu e^{-\frac{\mu-x}{\gamma}} dx + \int_\mu^\infty e^{-\frac{x-\mu}{\gamma}} dx \right) \\
&= \frac{1}{2\gamma} \left(\int_{-\infty}^\mu e^{\frac{x-\mu}{\gamma}} dx + \int_\mu^\infty e^{\frac{\mu-x}{\gamma}} dx \right) \\
&= \frac{1}{2\gamma} \left(\left[\gamma e^{\frac{x-\mu}{\gamma}} \right]_{-\infty}^\mu + \left[-\gamma e^{\frac{\mu-x}{\gamma}} \right]_\mu^\infty \right) \\
&= \frac{1}{2\gamma} \left(\gamma \left[e^0 - 0 \right] - \gamma \left[0 - (e^0) \right] \right) \\
&= \frac{\gamma}{2\gamma} (1 - (-1)) \\
&= \frac{1}{2} \cdot 2 \\
&= 1
\end{aligned}
$$

2.6

$$\begin{aligned}
\int p(x|\mathcal{D}) &= \int_{-\infty}^{\infty} \frac{1}{N} \sum_{n=1}^{N} \delta(x - x_n) dx \\
&= \frac{1}{N} \sum_{n=1}^{N} \int_{-\infty}^{\infty} \delta(x - x_n) dx \qquad \text{finite sum} \\
&= \frac{1}{N} \sum_{n=1}^{N} 1 \qquad \text{by def. of } \delta \\
&= \frac{1}{N} \cdot N \\
&= 1
\end{aligned}$$

2.8

$$\begin{aligned}
\mathbb{V}[f] &= \mathbb{E}\left[(f(x) - \mathbb{E}[f(x)])^2\right] \\
&= \mathbb{E}\left[f(x)^2 - 2f(x)\mathbb{E}[f(x)] + \mathbb{E}[f(x)]^2\right] \\
&= \mathbb{E}\left[f(x)^2\right] - 2\mathbb{E}[f(x)]\mathbb{E}[f(x)] + \mathbb{E}[f(x)]^2 \qquad \text{linearity of expectation} \\
&= \mathbb{E}\left[f(x)^2\right] - 2\mathbb{E}[f(x)]^2 + \mathbb{E}[f(x)]^2 \\
&= \mathbb{E}\left[f(x)^2\right] - \mathbb{E}[f(x)]^2
\end{aligned}$$

2.9

For independent random variables it holds that

$$\begin{aligned}
\mathbb{E}[xy] &:= \int xyp(x,y)d(x,y) \\
&= \int \int yxp(x)p(y)dxdy \qquad \text{ind.} \\
&= \int y \left(\int xp(x)dx \right) p(y)dy \\
&= \int xp(x)dx \int yp(y)dy \\
&= \mathbb{E}[x]\mathbb{E}[y]
\end{aligned}$$

and consequently

$$\mathrm{cov}[x,y] := \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y]$$

$$
\begin{aligned}
&= \mathbb{E}[x]\mathbb{E}[y] - \mathbb{E}[x]\mathbb{E}[y] \\
&= 0
\end{aligned}
$$

2.10

$$
\begin{aligned}
\mathbb{E}[x+z] &= \int (x+z)p(x,z)d(x,z) \\
&= \int\int (x+z)p(x)p(z)dxdz \qquad \text{ind.} \\
&= \int\int xp(x)p(z)dxdz + \int\int zp(x)p(z)dxdz \\
&= \int \underbrace{\left(\int xp(x)dx\right)}_{=\mathbb{E}[x]} p(z)dz + \int z \underbrace{\left(\int p(x)dx\right)}_{=1} p(z)dz \\
&= \mathbb{E}[x] \underbrace{\int p(z)dz}_{=1} + \int zp(z)dz \\
&= \mathbb{E}[x] + \mathbb{E}[z]
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{V}[x+z] &:= \mathbb{E}[(x+z)^2] - \mathbb{E}[x+z]^2 \\
&= \mathbb{E}[x^2 + 2xz + z^2] - \mathbb{E}[x+z]^2 \qquad \text{linearity (cf. above)} \\
&= \mathbb{E}[x^2] + 2\mathbb{E}[xz] + \mathbb{E}[z^2] - (\mathbb{E}[x] + \mathbb{E}[z])^2 \\
&= \mathbb{E}[x^2] + 2\mathbb{E}[xz] + \mathbb{E}[z^2] - \mathbb{E}[x]^2 - 2\mathbb{E}[x]\mathbb{E}[z] - \mathbb{E}[z]^2 \qquad \text{ind} \\
&= \mathbb{E}[x^2] + \cancel{2\mathbb{E}[x]\mathbb{E}[z]} + \mathbb{E}[z^2] - \mathbb{E}[x]^2 \cancel{-2\mathbb{E}[x]\mathbb{E}[z]} - \mathbb{E}[z]^2 \\
&= \mathbb{E}[x^2] - \mathbb{E}[x]^2 + \mathbb{E}[z^2] - \mathbb{E}[z]^2 \\
&= \mathbb{V}[x] + \mathbb{V}[z]
\end{aligned}
$$

2.22

$$
\begin{aligned}
f(p) &:= -\sum_{i=1}^{M} p_i \ln p_i \\
g(p) &:= \sum_{i=1}^{M} p_i - 1 \\
L(p,\lambda) &:= f(p) + \lambda g(p)
\end{aligned}
$$

$$= -\sum_{i=1}^{M} p_i \ln p_i + \lambda \left(\sum_{i=1}^{M} p_i - 1 \right)$$

$$\frac{\partial L}{\partial p_i} = -1 \cdot \ln p_i - p_i \cdot \frac{1}{p_i} + \lambda \cdot (1 - 0)$$
$$= -\ln p_i - 1 + \lambda$$
$$\frac{\partial L}{\partial \lambda} = -\left(\sum_{i=1}^{M} p_i - 1 \right)$$
$$= 1 - \sum_{i=1}^{M} p_i$$

$$\frac{\partial L}{\partial p_i} \overset{!}{=} 0$$
$$\Leftrightarrow \quad -\ln p_i - 1 + \lambda = 0$$
$$\Leftrightarrow \quad -\ln p_i = 1 - \lambda$$
$$\Leftrightarrow \quad \ln p_i = \lambda - 1$$
$$\Leftrightarrow \quad p_i = e^{\lambda - 1} \quad (1)$$

$$\frac{\partial L}{\partial \lambda} \overset{!}{=} 0$$
$$\Leftrightarrow \quad 1 - \sum_{i=1}^{M} p_i = 0$$
$$\overset{(1)}{\Rightarrow} \quad 1 - \sum_{i=1}^{M} e^{\lambda - 1} = 0$$
$$\Leftrightarrow \quad 1 - M \cdot e^{\lambda - 1} = 0$$
$$\Leftrightarrow \quad 1 = M \cdot e^{\lambda - 1}$$
$$\Leftrightarrow \quad \frac{1}{M} = e^{\lambda - 1}$$

Hence for all $i \in \{1, ..., M\}$:

$$p_i = \frac{1}{M}$$

and consequently

$$H[p] = -\sum_{i=1}^{M} p_i \ln p_i$$

$$
\begin{aligned}
&= -\sum_{i=1}^{M} \frac{1}{M} \ln \frac{1}{M} \\
&= -M \cdot \frac{1}{M} \left(-\ln M\right) \\
&= \ln M
\end{aligned}
$$