

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

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

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

2.22

$$f(p) := - \sum_{i=1}^M p_i \ln p_i$$

$$g(p) := \sum_{i=1}^M p_i - 1$$

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

$$\begin{aligned} \frac{\partial L}{\partial p_i} &= -1 \cdot \ln p_i - p_i \cdot \frac{1}{p_i} + \lambda \cdot (1 - 0) \\ &= -\ln p_i - 1 + \lambda \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial \lambda} &= - \left(\sum_{i=1}^M p_i - 1 \right) \\ &= 1 - \sum_{i=1}^M p_i \end{aligned}$$

$$\begin{aligned} &\frac{\partial L}{\partial p_i} \stackrel{!}{=} 0 \\ \Leftrightarrow & -\ln p_i - 1 + \lambda = 0 \\ &\Leftrightarrow -\ln p_i = 1 - \lambda \\ &\Leftrightarrow \ln p_i = \lambda - 1 \\ &\Leftrightarrow p_i = e^{\lambda-1} \quad (1) \end{aligned}$$

$$\begin{aligned} &\frac{\partial L}{\partial \lambda} \stackrel{!}{=} 0 \\ \Leftrightarrow & 1 - \sum_{i=1}^M p_i = 0 \\ \stackrel{(1)}{\Rightarrow} & 1 - \sum_{i=1}^M e^{\lambda-1} = 0 \\ \Leftrightarrow & 1 - M \cdot e^{\lambda-1} = 0 \end{aligned}$$

$$\Leftrightarrow 1 = M \cdot e^{\lambda-1}$$

$$\Leftrightarrow \frac{1}{M} = e^{\lambda-1}$$

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

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

and consequently

$$\begin{aligned} H[p] &= - \sum_{i=1}^M p_i \ln p_i \\ &= - \sum_{i=1}^M \frac{1}{M} \ln \frac{1}{M} \\ &= -M \cdot \frac{1}{M} (-\ln M) \\ &= \ln M \end{aligned}$$