

Solution Manual for Pattern Recognition and Machine Learning by Christopher M. Bishop

Edited by Nicola Fanelli



August 2nd, 2023

1 Introduction

Exercise 1.1 We seek to minimize E by setting its derivative with respect to \mathbf{w} to zero. We can consider each weight w_i separately, so we have

$$\begin{aligned} \frac{\partial E}{\partial w_i} = 0 &\iff \sum_{n=1}^N \{[w_0 + w_1 x_n + w_2 x_n^2 + \cdots + w_M x_n^M - t_n] x_n^i\} = 0 \\ &\iff \sum_{n=1}^N \{w_0 x_n^i + w_1 x_n^{i+1} + w_2 x_n^{i+2} + \cdots + w_M x_n^{i+M}\} = \sum_{n=1}^N (x_n)^i t_n \\ &\iff \sum_{j=0}^M \left\{ \sum_{n=1}^N (x_n)^{i+j} w_j \right\} = \sum_{n=1}^N (x_n)^i t_n \end{aligned}$$

This result corresponds to the set of linear equations shown in the text.

Exercise 1.2 Again, we seek to minimize \tilde{E} , by setting each one of the partial derivatives with respect to w_i to zero. We have

$$\begin{aligned} \frac{\partial \tilde{E}}{\partial w_i} = 0 &\iff \sum_{n=1}^N \{[y(x_n, \mathbf{w}) - t_n] x_n^i\} + \lambda w_i = 0 \\ &\iff \sum_{n=1}^N \{[w_0 + w_1 x_n + w_2 x_n^2 + \cdots + w_M x_n^M - t_n] x_n^i\} + \lambda w_i = 0 \\ &\iff \sum_{n=1}^N \{w_0 x_n^i + w_1 x_n^{i+1} + w_2 x_n^{i+2} + \cdots + w_M x_n^{i+M}\} + \lambda w_i = \sum_{n=1}^N (x_n)^i t_n \end{aligned}$$

The coefficients $\mathbf{w} = \{w_i\}$ that minimize $\tilde{E}(\mathbf{w})$ are given by the solution to the following set of linear equations:

$$\sum_{j=0}^M A_{ij} w_j + \lambda w_i = T_i$$

where $A_{ij} = \sum_{n=1}^N (x_n)^{i+j}$ and $T_i = \sum_{n=1}^N (x_n)^i t_n$.

Exercise 1.3 We want to determine the probability $p(\text{apple})$ of randomly extracting an apple from a randomly chosen box.

By the law of total probability, we have

$$p(\text{apple}) = p(\text{apple}, r) + p(\text{apple}, b) + p(\text{apple}, g)$$

where r , b and g denote the red, blue and green boxes, respectively.

We also apply the product rule to each term of the sum, obtaining:

$$\begin{aligned}
 p(apple) &= p(apple|r)p(r) + p(apple|b)p(b) + p(apple|g)p(g) \\
 &= \frac{3}{10} \frac{1}{5} + \frac{1}{2} \frac{1}{5} + \frac{3}{10} \frac{3}{5} \\
 &= \frac{3}{50} + \frac{1}{10} + \frac{9}{50} \\
 &= \frac{17}{50}
 \end{aligned}$$

Furthermore, the exercise asks use to determine the probability of having selected the green box, given that we have extracted an orange, corresponding to $p(g|orange)$.

We can use Bayes' theorem to obtain

$$p(g|orange) = \frac{p(orange|g)p(g)}{p(orange)}$$

We already know $p(g)$, and we can compute $p(orange)$ in the same way we computed $p(apple)$, obtaining

$$\begin{aligned}
 p(orange) &= p(orange, r) + p(orange, b) + p(orange, g) \\
 &= p(orange|r)p(r) + p(orange|b)p(b) + p(orange|g)p(g) \\
 &= \frac{4}{10} \frac{1}{5} + \frac{1}{2} \frac{1}{5} + \frac{3}{10} \frac{3}{5} \\
 &= \frac{2}{25} + \frac{1}{10} + \frac{9}{50} \\
 &= \frac{18}{50} = \frac{9}{25}
 \end{aligned}$$

Finally, we obtain:

$$\begin{aligned}
 p(g|orange) &= \frac{p(orange|g)p(g)}{p(orange)} \\
 &= \frac{\frac{3}{10} \frac{3}{5}}{\frac{9}{25}} \\
 &= \frac{3}{10} \frac{3}{5} \frac{25}{9} \\
 &= \frac{225}{450} = \frac{1}{2}
 \end{aligned}$$

Exercise 1.4

2 Probability Distributions

Exercise 2.1 Considering the definition $Bern(x|\mu) = \mu^x(1-\mu)^{1-x}$, we have:

$$\begin{aligned}\sum_{x=0}^1 p(x|\mu) &= \mu^0(1-\mu)^{1-0} + \mu^1(1-\mu)^{1-1} = 1 - \mu + \mu = 1 \\ \mathbb{E}[x] &= 0\mu^0(1-\mu)^{1-0} + 1\mu^1(1-\mu)^{1-1} = \mu \\ var[x] &= \sum_{x=0}^1 p(x)(x-\mu)^2 = (1-\mu)(-\mu)^2 + \mu(1-\mu)^2 = \mu(1-\mu) \\ H[x] &= -\sum_{x=0}^1 p(x) \ln p(x) = -\mu \ln \mu - (1-\mu) \ln(1-\mu)\end{aligned}$$

Exercise 2.2 The distribution is normalized if and only if $\sum_{x=0}^N p(x) = 1$, where the sum can be computed as follows:

$$p(-1|\mu) + p(1|\mu) = \frac{1-\mu}{2} + \frac{1+\mu}{2} = 1$$

The expectation value is given by:

$$\mathbb{E}[x] = -1 \frac{1-\mu}{2} + 1 \frac{1+\mu}{2} = \mu$$

The variance is given by:

$$var[x] = (-1)^2 \frac{1-\mu}{2} + (1)^2 \frac{1+\mu}{2} - \mu^2 = 1 - \mu^2$$

The entropy is given by:

$$H[x] = -\sum_{x=-1}^1 p(x) \ln p(x) = -\frac{1-\mu}{2} \ln \frac{1-\mu}{2} - \frac{1+\mu}{2} \ln \frac{1+\mu}{2}$$

Exercise 2.3 We show that:

$$\binom{N}{m} + \binom{N}{m-1} = \frac{N!(N-m+1) + N!m}{(N-m+1)!m!} = \frac{(N+1)!}{(N+1-m)!m!} = \binom{N+1}{m}$$

Then we prove by induction (2.263), where for $N = 1$ we have:

$$\sum_{m=0}^1 \binom{1}{m} x^m = \binom{1}{0} x^0 + \binom{1}{1} x^1 = 1 + x = (1+x)^1$$

Assuming (2.263) holds for N , we have:

$$\begin{aligned} (1+x)^N + 1 &= (1+x)(1+x)^N = (1+x) \sum_{m=0}^N \binom{N}{m} x^m \\ &= \sum_{m=0}^N \binom{N}{m} x^m + \sum_{m=0}^N \binom{N}{m} x^{m+1} \\ &= \sum_{m=0}^N \binom{N}{m} x^m + \sum_{m=1}^{N+1} \binom{N}{m-1} x^m \\ &= \binom{N}{0} x^0 + \sum_{m=1}^N \binom{N}{m} x^m + \sum_{m=1}^N \binom{N}{m-1} x^m + \binom{N}{N} x^{N+1} \\ &= \binom{N+1}{0} x^0 + \sum_{m=1}^N \binom{N+1}{m} x^m + \binom{N+1}{N+1} x^{N+1} \\ &= \sum_{m=0}^{N+1} \binom{N+1}{m} x^m \end{aligned}$$

proving the binomial theorem.

Then we use it to show that the binomial distribution is normalized:

$$\begin{aligned} \sum_{m=0}^N \binom{N}{m} \mu^m (1-\mu)^{N-m} &= \sum_{m=0}^N \binom{N}{m} \mu^m \frac{(1-\mu)^N}{(1-\mu)^m} \\ &= (1-\mu)^N \sum_{m=0}^N \binom{N}{m} \mu^m \frac{1}{(1-\mu)^m} \\ &= (1-\mu)^N \left(1 + \frac{\mu}{1-\mu}\right)^N \\ &= 1 \end{aligned}$$

Exercise 2.4 We differentiate (2.264) with respect to μ to obtain:

$$\begin{aligned}
& \frac{\partial}{\partial \mu} \sum_{m=0}^N \binom{N}{m} \mu^m (1-\mu)^{N-m} = 0 \iff \\
& \sum_{m=0}^N \binom{N}{m} m \mu^{m-1} (1-\mu)^{N-m} - \sum_{m=0}^N \binom{N}{m} \mu^m (N-m) (1-\mu)^{N-m-1} = 0 \iff \\
& \mathbb{E}[m] - \frac{\mu N}{1-\mu} \sum_{m=0}^N \binom{N}{m} \mu^m (1-\mu)^{N-m} + \frac{\mu}{1-\mu} \sum_{m=0}^N \binom{N}{m} m \mu^m (1-\mu)^{N-m} = 0 \iff \\
& \mathbb{E}[m] - \frac{\mu N}{1-\mu} + \mathbb{E}[m] \frac{\mu}{1-\mu} = 0 \iff \\
& \mathbb{E}[m] = \frac{\mu N}{1-\mu} (1-\mu) = \mu N
\end{aligned}$$

proving the result (2.11).