Pattern Recognition
Solutions to Selected Exercises
Bishop, Chapter 1

1.3 Use the sum and product rules of probability.

Probability of drawing an apple:

$$\begin{aligned} p(a) &= \sum_{\text{box}} p(a, \text{box}) \\ &= \sum_{\text{box}} p(a|\text{box})p(\text{box}) \\ &= p(a|r)p(r) + p(a|b)p(b) + p(a|g)p(g) \\ &= 0.3 \times 0.2 + 0.5 \times 0.2 + 0.3 \times 0.6 = 0.34 \end{aligned}$$

Probability of green box given orange

$$\begin{aligned} p(g|o) &= \frac{p(g,o)}{p(o)} = \frac{p(o|g)p(g)}{\sum_{\text{box}} p(o|\text{box})p(\text{box})} \\ &= \frac{0.18}{0.36} = 0.5 \end{aligned}$$

1.5

$$\begin{aligned} \text{var}[f(x)] &= \mathbb{E}[(f(x) - \mathbb{E}[f(x)])^2] && \text{(1.38;expand)} \\ &= \mathbb{E}[f(x)^2 - 2f(x)\mathbb{E}[f(x)] + \mathbb{E}[f(x)]^2] \\ & && \text{(push expectation inward)} \\ &= \mathbb{E}[f(x)^2] - 2\mathbb{E}[f(x)]\mathbb{E}[f(x)] + \mathbb{E}[f(x)]^2 \\ &= \mathbb{E}[f(x)^2] - \mathbb{E}[f(x)]^2 && \text{(1.39)} \end{aligned}$$

The important thing to notice is that $\mathbb{E}[f(x)]$ is a constant, and so

$$\mathbb{E}[2f(x)\mathbb{E}[f(x)]] = 2\mathbb{E}[f(x)]\mathbb{E}[f(x)] \qquad \text{and} \qquad \mathbb{E}[\mathbb{E}[f(x)]^2] = \mathbb{E}[f(x)]^2$$

1.6

$$\mathrm{cov}[x, y] = \mathbb{E}[xy] - \mathbb{E}[x]\mathbb{E}[y] \tag{1.41}$$

We'll show that $\mathbb{E}[xy] = \mathbb{E}[x]\mathbb{E}[y]$ if x and y are independent.

$$\begin{aligned}\mathbb{E}[xy] &\equiv \sum_x \sum_y xy\ p(x, y) \\ &= \sum_x \sum_y xy\ p(x)p(y) \\ &= \sum_x x\ p(x) \sum_y y\ p(y) = \mathbb{E}[x]\mathbb{E}[y]\end{aligned}$$

1.10 We don't need independence for the first one:

$$\begin{aligned}\mathbb{E}[x+z] &\equiv \sum_x \sum_z (x+z)\ p(x, z) \\ &= \sum_x \sum_z x\ p(x, z) + \sum_z \sum_x z\ p(x, z) \\ &= \sum_x x \sum_z p(x, z) + \sum_z z \sum_x p(x, z) \\ &= \sum_x x\ p(x) + \sum_z z\ p(z) = \mathbb{E}[x] + \mathbb{E}[z]\end{aligned}$$

In exercise 1.5 we have shown that $\mathrm{var}[x+z] = \mathbb{E}[(x+z)^2] - \mathbb{E}[x+z]^2$. Since $\mathbb{E}[x+z] = \mathbb{E}[x] + \mathbb{E}[z]$ as we have shown above, we get

$$\mathrm{var}[x+z] = \mathbb{E}[(x+z)^2] - \mathbb{E}[x]^2 - 2\mathbb{E}[x]\mathbb{E}[z] - \mathbb{E}[z]^2$$

Left to determine $\mathbb{E}[(x+z)^2]$:

$$\begin{aligned}\mathbb{E}[(x+z)^2] &\equiv \sum_x \sum_z (x+z)^2\ p(x, z) \qquad \text{(independence/expand)} \\ &= \sum_x \sum_z (x^2 + 2xz + z^2)\ p(x)p(z) \\ &= \sum_x \sum_z x^2\ p(x)p(z) + \sum_x \sum_z 2xz\ p(x)p(z) + \sum_z \sum_x z^2\ p(x)p(z) \\ &= \sum_x x^2\ p(x) \sum_z p(z) + 2 \sum_x x\ p(x) \sum_z z\ p(z) + \sum_z z^2\ p(z) \sum_x p(x) \\ &= \mathbb{E}[x^2] + 2\mathbb{E}[x]\mathbb{E}[z] + \mathbb{E}[z^2]\end{aligned}$$

So finally we get

$$\begin{aligned}\mathrm{var}[x+z] &= \mathbb{E}[x^2] + 2\mathbb{E}[x]\mathbb{E}[z] + \mathbb{E}[z^2] - \mathbb{E}[x]^2 - 2\mathbb{E}[x]\mathbb{E}[z] - \mathbb{E}[z]^2 \\ &= (\mathbb{E}[x^2] - \mathbb{E}[x]^2) + (\mathbb{E}[z^2] - \mathbb{E}[z]^2) = \mathrm{var}[x] + \mathrm{var}[z]\end{aligned}$$

1.11 Take partial derivatives and equate to zero

$$\frac{\partial}{\partial \mu} = \frac{1}{\sigma^2}\sum_{n=1}^{N}(x_n - \mu) = \frac{1}{\sigma^2}\left(\sum x_n - N\mu\right) = 0$$

Hence we have $\mu_{\mathrm{ML}} = \bar{x}$, where $\bar{x} = \frac{1}{N}\sum x_n$ is the sample mean.

$$\frac{\partial}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^3}\sum_{n=1}^{N}(x_n - \bar{x})^2 = 0$$

Hence, we have

$$\sigma^3 N = \sigma \sum_{n=1}^{N}(x_n - \bar{x})^2, \text{ so } \sigma^2_{\mathrm{ML}} = \frac{1}{N}\sum_{n=1}^{N}(x_n - \bar{x})^2$$

1.29 It is tempting to use

$$H[x] = -\sum_{i=1}^{M} p(x_i)\log p(x_i)$$

and the fact that $-\log u$ is a convex function, but this gives us an inequality in the wrong direction. We should use the equivalent

$$H[x] = \sum_{i=1}^{M} p(x_i)\log\frac{1}{p(x_i)}$$

instead. Since $\log u$ is concave, we have

$$\sum_{i=1}^{M} p(x_i)\log\frac{1}{p(x_i)} \le \log\left(\sum_{i=1}^{M} p(x_i)\frac{1}{p(x_i)}\right) = \log M.$$

1.37

$$\begin{aligned}
H[\mathbf{x},\mathbf{y}] &= -\sum_{\mathbf{x}}\sum_{\mathbf{y}} p(\mathbf{x},\mathbf{y})\ln p(\mathbf{x},\mathbf{y}) && \text{(apply product rule)}\\
&= -\sum_{\mathbf{x}}\sum_{\mathbf{y}} p(\mathbf{x},\mathbf{y})\ln\{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})\}\\
&= -\sum_{\mathbf{x}}\sum_{\mathbf{y}} p(\mathbf{x},\mathbf{y})\{\ln p(\mathbf{y}|\mathbf{x}) + \ln p(\mathbf{x})\}\\
&= -\sum_{\mathbf{x}}\sum_{\mathbf{y}} p(\mathbf{x},\mathbf{y})\ln p(\mathbf{y}|\mathbf{x}) - \sum_{\mathbf{x}}\sum_{\mathbf{y}} p(\mathbf{x},\mathbf{y})\ln p(\mathbf{x})\\
& && \text{(apply sum rule)}\\
&= -\sum_{\mathbf{x}}\sum_{\mathbf{y}} p(\mathbf{x},\mathbf{y})\ln p(\mathbf{y}|\mathbf{x}) - \sum_{\mathbf{x}} p(\mathbf{x})\ln p(\mathbf{x})\\
&= H[\mathbf{y}|\mathbf{x}] + H[\mathbf{x}]
\end{aligned}$$

1.39 a) $H[\mathbf{x}] = -\sum_{\mathbf{x}} p(\mathbf{x}) \ln p(\mathbf{x}) = -2/3 \ln 2/3 - 1/3 \ln 1/3 \approx 0.64$.

b) $H[\mathbf{y}] \approx 0.64$ (same as $H[\mathbf{x}]$. Why?)

c) $H[\mathbf{y}|\mathbf{x}] = -1/3 \ln 1/2 - 1/3 \ln 1/2 - 0 \ln 0 - 1/3 \ln 1 \approx 0.46$.

d) $H[\mathbf{x}|\mathbf{y}] \approx 0.46$.

e)

$$\begin{aligned} H[\mathbf{x}, \mathbf{y}] &= -\sum_{\mathbf{x},\mathbf{y}} p(\mathbf{x}, \mathbf{y}) \ln p(\mathbf{x}, \mathbf{y}) \\ &= -1/3 \ln 1/3 - 1/3 \ln 1/3 - 0 \ln 0 - 1/3 \ln 1/3 \approx 1.1 \end{aligned}$$

f) $I[\mathbf{x}, \mathbf{y}] = H[\mathbf{x}] - H[\mathbf{x}|\mathbf{y}] \approx 0.64 - 0.46 = 0.18$.

1.41 Prove that

$$I[\mathbf{x}, \mathbf{y}] = H[\mathbf{x}] - H[\mathbf{x}|\mathbf{y}] \tag{1.121}$$

Note that

$$H[\mathbf{x}] = -\sum_{\mathbf{x}} \sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y}) \log p(\mathbf{x})$$

since

$$p(\mathbf{x}) = \sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y}) \qquad \text{(sum rule of probability)}$$

Also we have

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})} \qquad \text{(product rule of probability)}$$

so

$$H[\mathbf{x}|\mathbf{y}] = -\sum_{\mathbf{x}} \sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y}) \log \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})}$$

Then

$$\begin{aligned} H[\mathbf{x}] - H[\mathbf{x}|\mathbf{y}] &= -\sum_{\mathbf{x}} \sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y}) \log p(\mathbf{x}) + \sum_{\mathbf{x}} \sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y}) \log \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{y})} \\ &= \sum_{\mathbf{x}} \sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y}) \log \frac{p(\mathbf{x}, \mathbf{y})}{p(\mathbf{x})p(\mathbf{y})} \\ &= -\sum_{\mathbf{x}} \sum_{\mathbf{y}} p(\mathbf{x}, \mathbf{y}) \log \frac{p(\mathbf{x})p(\mathbf{y})}{p(\mathbf{x}, \mathbf{y})} = I[\mathbf{x}, \mathbf{y}] \end{aligned}$$

Venn diagram showing the relationship between entropy and mutual information: