COMP2610/COMP6261 Tutorial 3 Solutions[1]

Week 3, Semester 2, 2022

1. (a) Let n_i denote the number of times that we observe outcome $X = i$. The likelihood is

$$\begin{aligned} L(\theta) &= \prod_{i=1}^{N} p(X = x_i|\theta) \\ &= \prod_{i:x_i=1} \left(\frac{\theta}{2}\right) \cdot \prod_{i:x_i=2} \left(\frac{\theta}{2}\right) \cdot \prod_{i:x_i=3} (1-\theta) \\ &= \left(\frac{\theta}{2}\right)^{n_1} \cdot \left(\frac{\theta}{2}\right)^{n_2} \cdot (1-\theta)^{n_3} \\ &= \left(\frac{\theta}{2}\right)^{n_1+n_2} \cdot (1-\theta)^{n_3}. \end{aligned}$$

(b) The log-likelihood is

$$\mathcal{L}(\theta) = (n_1 + n_2) \cdot \log \frac{\theta}{2} + n_3 \cdot \log(1-\theta).$$

The derivative is

$$\mathcal{L}'(\theta) = \frac{n_1 + n_2}{\theta} - \frac{n_3}{1-\theta}.$$

We have that $n_1 = 3, n_2 = 3, n_3 = 4$. So, we need

$$\frac{6}{\theta} = \frac{4}{1-\theta}$$

for which the solution may be checked to be $\theta = 0.6$. Observe then that we estimate

$$p(X = 1) = 0.3 \quad p(X = 2) = 0.3$$

$$p(X = 3) = 0.4,$$

matching the frequencies of observations of each outcome.

[1] Based in part on solutions by Avraham Ruderman or the 2012 version of the course.

2. (a) We can show that X and Y are not statistically independent by showing that $p(x,y)$ 6= $p(x)p(y)$ for at least one value of x and y. For example: $p(X = 1) = 1/8+1/8 = 1/4$ and $p(Y = 2) = 1/8 + 1/16 + 1/16 = 1/4$. From the given table we see that:

$p(X = 1, Y = 2) = 1/8$ which is different from $p(X = 1)p(Y = 2) = 1/16$.

(b) First, we find the marginal probabilities using the sum rule:

$$\mathbf{p}(X) = \big(P(X = 1), P(X = 2), P(X = 3), P(X = 4)\big) = (1/4, 1/4, 1/4, 1/4)$$
$$\mathbf{p}(Y) = \big(P(Y = 1), P(Y = 2), P(Y = 3), P(Y = 4)\big) = (1/4, 1/4, 1/4, 1/4).$$

We see that both $p(X)$ and $p(Y)$ are uniform distributions with 4 possible states. Hence:

$H(X) = H(Y) = \log_2 4 = 2$ bits.

To compute the conditional entropy $H(X|Y)$ we need the conditional distributions $p(X|Y)$ which can be computed by using the definition of conditional probability $p(X = x|Y = y) = p(X = x, Y = y)/p(Y = y)$. In other words, we divide the rows of the given table by the corresponding marginal.

$\mathbf{p}(X|Y = 1) = (0,0,1/2,1/2)$ $\mathbf{p}(X|Y =$
$2) = (1/2,1/4,1/4,0)$ $\mathbf{p}(X|Y = 3) =$
$(1/2,1/2,0,0)$ $\mathbf{p}(X|Y = 4) =$
$(0,1/4,1/4,1/2)$.

Hence the conditional entropy $H(X|Y)$ is given by:

$$\begin{aligned}
H(X|Y) &= \sum_{i=1}^{4} p(Y = i)H(X|Y = i) \\
&= (1/4)H(0, 0, 1/2, 1/2) + (1/4)H(1/2, 1/4, 1/4, 0) \\
&+ (1/4)H(1/2, 1/2, 0, 0) + (1/4)H(0, 1/4, 1/4, 1/2) \\
&= 1/4 \times 1 + 1/4 \times 3/2 + 1/4 \times 1 + 1/4 \times 3/2 \\
&= 5/4 \text{ bits.}
\end{aligned}$$

Here we note that conditioning has indeed decreased entropy. We can compute the joint entropy by using the chain rule:

$$H(X,Y) = H(X|Y) + H(Y) = 5/4 + 2 = 13/4 \text{ bits.}$$

Additionally, we know that by the chain rule $H(X,Y) = H(Y|X) + H(X)$, hence:

$$H(Y|X) = H(X,Y) - H(X) = 13/4 - 2 = 5/4 \text{ bits.}$$

3. (a) $h\left(c=\text{red}, v=\mathrm{K}\right)=\log_2 \frac{1}{P(c=\text{red}, v=\mathrm{K})}=\log_2 \frac{1}{1/26}=4.7004$ bits.

(b) $h(v=\mathrm{K}|f=1)=\log_2 \frac{1}{P(v=\mathrm{K}|f=1)}=\log_2 \frac{1}{1/3}=1.585$ bits.

(c) We have

i. $H\left(S\right)=\sum_s p\left(s\right)\log_2 \frac{1}{p(s)}=4\times\frac{1}{4}\times\log_2\frac{1}{1/4}=2$ bits.

ii. $H(V,S)=\sum_{v,s} p(v,s)\log_2\frac{1}{p(v,s)}=52\times\frac{1}{52}\log_2\frac{1}{1/52}=5.7$ bits.