

Supervised Learning: Exercise for Information Theory Part 2

Yawei Li

`yawei.li@stat.uni-muenchen.de`

Date

Exercise 1: Entropy

A fair **die** is rolled at the same time as a fair **coin** is tossed. Let A be the number on the upper surface of the dice and let B describe the outcome of the coin toss, where

$$B = \begin{cases} 1, & \text{head,} \\ 0, & \text{tail.} \end{cases}$$

Two random variables X and Y are given by $X = A + B$ and $Y = A - B$, respectively.

(a) Calculate the entropies $H(X)$ and $H(Y)$, the conditional entropies $H(Y|X)$ and $H(X|Y)$, the joint entropy $H(X, Y)$ and the mutual information $I(X; Y)$.

Solution to Exercise 1 (a)

1. Let a, b, x, y be the realizations of A, B, X, Y , respectively.
2. **If we have observed x and y , then we can calculate the observed a and b .**
Since: $x = a + b$ and $y = a - b$ yields $a = \frac{x+y}{2}$ and $b = \frac{x-y}{2}$.
3. In other words, a pair (x, y) is uniquely associated with a pair (a, b) .
4. For each pair $(a, b) \in \{0, 1, \dots, 6\} \times \{0, 1\}$, it holds that $p_{AB}(a, b) = \frac{1}{6} \cdot \frac{1}{2} = \frac{1}{12}$.
5. Therefore, $p_{XY}(x, y) = p_{AB}(a, b) = \frac{1}{12}$ for all (x, y) .
6. So, the joint entropy

$$\begin{aligned} H(X, Y) &= - \sum_{x,y} p_{XY}(x, y) \log_2 p_{XY}(x, y) = -12 \cdot \frac{1}{12} \log_2 \frac{1}{12} \\ &= 2 + \log_2 3. \end{aligned}$$

Solution to Exercise 1 (a): Continued

Next, we compute $H(X)$ and $H(Y)$. We enumerate all the possible (a, b) events.

x	events (a, b)	$p_X(x)$
1	(1, 0)	1/12
2	(2, 0), (1, 1)	1/6
3	(3, 0), (2, 1)	1/6
4	(4, 0), (3, 1)	1/6
5	(5, 0), (4, 1)	1/6
6	(6, 0), (5, 1)	1/6
7	(6, 1)	1/12

$$\begin{aligned}H(X) &= \sum_x p_X(x) \log_2 p_X(x) \\&= -2 \cdot \frac{1}{12} \log_2 \frac{1}{12} - 5 \cdot \frac{1}{6} \log_2 \frac{1}{6} \\&= \frac{7}{6} + \log_2 3.\end{aligned}$$

y	events (a, b)	$p_Y(y)$
0	(1, 1)	1/12
1	(1, 0), (2, 1)	1/6
2	(2, 0), (3, 1)	1/6
3	(3, 0), (4, 1)	1/6
4	(4, 0), (5, 1)	1/6
5	(5, 0), (6, 1)	1/6
6	(6, 0)	1/12

$$\begin{aligned}H(Y) &= \sum_y p_Y(y) \log_2 p_Y(y) \\&= -2 \cdot \frac{1}{12} \log_2 \frac{1}{12} - 5 \cdot \frac{1}{6} \log_2 \frac{1}{6} \\&= \frac{7}{6} + \log_2 3.\end{aligned}$$

Solution to Exercise 1 (a): Continued

The conditional entropies are

$$H(X|Y) = H(X, Y) - H(Y) = 2 + \log_2 3 - \frac{7}{6} - \log_2 3 = \frac{5}{6}$$

$$H(Y|X) = H(X, Y) - H(X) = 2 + \log_2 3 - \frac{7}{6} - \log_2 3 = \frac{5}{6}$$

The mutual information $I(X; Y)$ can be determined according to

$$I(X; Y) = H(X) - H(X, Y) = \frac{7}{6} + \log_2 3 - \frac{5}{6} = \frac{1}{3} + \log_2 3.$$

Exercise 1: Question (b)

(b) Show that, for independent discrete random variables X and Y ,

$$I(X; X + Y) - I(Y; X + Y) = H(X) - H(Y)$$

Solution to Exercise 1 (b)

$$\begin{aligned}I(X; X + Y) - I(Y; X + Y) &= H(X) - H(X|X + Y) - H(Y) + H(Y|X + Y) \\&= H(X) - H(Y) + (H(Y|X + Y) - H(X|X + Y)) \\&= H(X) - H(Y) + \\&\quad (H(Y, X + Y) - H(X + Y) - H(X, X + Y) + H(X + Y)) \\&= H(X) - H(Y) + \underbrace{H(Y, X + Y) - H(X, X + Y)}_{=0} \\&= H(X) - H(Y)\end{aligned}$$

Note that if we observe $x + y$, and assume we also observe x , we can infer y .

In other words, **each pair $(x + y, x)$ has the same probability as $(x + y, y)$. Therefore, $H(Y, X + Y) = H(X, X + Y)$.**

Exercise 2: Mutual Information of Three Variables

Let X, Y, Z be three discrete random variables. The mutual information of X, Y , and Z is defined as:

$$I(X; Y; Z) = \sum_x \sum_y \sum_z p(x, y, z) \log \left(\frac{p(x, y)p(x, z)p(y, z)}{p(x)p(y)p(z)p(x, y, z)} \right).$$

(a) Prove the lemma:

$$I(X; Y; Z) = I(X; Y) - I(X; Y|Z).$$

Note that the conditional information is defined as:

$$I(X; Y|Z) = \sum_z \sum_x \sum_y p(z)p(x, y|z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)}.$$

Solution to Question 2 (a)

According to the definition of mutual information,

$$\begin{aligned} I(X; Y) - I(X; Y|Z) &= \sum_x \sum_y p(x, y) \log \frac{p(x, y)}{p(x)p(y)} - \sum_z \sum_x \sum_y p(z)p(x, y|z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)} \\ &= \sum_x \sum_y \sum_z p(x, y, z) \log \frac{p(x, y)}{p(x)p(y)} - \\ &\quad \sum_z \sum_x \sum_y p(z)p(x, y|z) \log \frac{p(x, y|z)p(z)^2}{p(x|z)p(y|z)p(z)^2} \\ &= \sum_x \sum_y \sum_z p(x, y, z) \log \frac{p(x, y)}{p(x)p(y)} - \sum_z \sum_x \sum_y p(x, y, z) \log \frac{p(x, y, z)p(z)}{p(x, z)p(y, z)} \\ &= \sum_x \sum_y \sum_z p(x, y, z) \log \left(\frac{p(x, y)p(x, z)p(y, z)}{p(x)p(y)p(z)p(x, y, z)} \right) \\ &= I(X; Y; Z). \end{aligned}$$

Exercise 2: Question (b)

(b) Prove the following relation with the above lemma:

$$I(X; Y) = I(X; Y|Z) + I(Y; Z) - I(Y; Z|X)$$

Recall the lemma: $I(X; Y) - I(X; Y|Z) = I(X; Y; Z)$

Solution to Question 2 (b)

Using the lemma we just proved, we obtain:

$$\begin{aligned} & I(X; Y|Z) + I(Y; Z) - I(Y; Z|X) \\ &= I(X; Y) - I(X; Y; Z) + I(Y; Z) - I(Y; Z) + I(X; Y; Z) \\ &= I(X; Y) \end{aligned}$$