Information Theory

Problem Set 06 - Dependent Random Variables

Luís Felipe Ramos Ferreira

lframos_ferreira@outlook.com

1. (a) $H(X,Y)$ is the joint entropy of X and Y. It means how much information, on average, each of the joint outcomes carries. ON onther words, is the expected value of information of the joint outcomes from ensembles X and Y.

$$H(X,Y) = \sum_{xy \in \mathcal{A}_x \mathcal{A}_y} P(x,y) log \frac{1}{P(x,y)}$$

   (b) $H(X|Y)$ is the conditional entropy of X given Y. It represents the average information infomration content of X given each $y \in \mathcal{A}_y$.

$$H(X|Y) = \sum_{xy \in \mathcal{A}_x \mathcal{A}_y} P(x,y) log \frac{1}{P(x|y)}$$

   (c) $I(X,Y)$ is the mutual information between X and Y. It is the average reduction of uncertainty gain of information about X when learning the values of Y, or vice-versa, since the mutual information is symmetric.

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

   (d) $I(X:Y|Z)$ is the conditional mutual information between X and Y given Z. It is the amount of information you gain about X when you learn about Y given that the ensemble Z is known.

$$I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$$

2. (a) The chain rule for entropy states that:

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i | X_1, \ldots, X_{i-1})$$

   (b) The chain rule for mutual information states that:

$$I(X_1, X_2, \ldots, X_n; Y) = \sum_{i=1}^{n} I(X_i; Y | X_1, \ldots, X_{i-1})$$

(c) The data-processing inequality (DPI) staets that, if $X \to Y \to Z$ is a Markov Chain, i. e. $p(x, y, z) = p(x)p(y|x)p(z|y)$ for all values in the ensembles, then $I(X; Z) \leq I(Y; Z)$. In general words, the inequality states that post-processing cannot create information, i.e, the amount of information you have after post-processing data is at most equal to the amount of information before the processing.

3. - $H(X, Y)$

   $$H(X,Y) = H(U,V,V,W) = H(U)+H(V|U)+H(V|U,V)+H(W|U,V,V)$$

   $$H(X,Y) = H(U) + H(V) + 0 + H(W) = H_u + H_v + H_w$$

   $H(V|U) = 0$ since U and V are independent, and $H(V|U, V)$ is obviously zero cause no information is gained.

   - $H(X|Y)$

   $$H(X|Y) = H(U,V|V,W) = H(U|V,W)+H(V|V,W,U) = H(U)+0 = H(U) = H_u$$

   $H(U|V, W) = 0$ since U, V, W are independent and $H(V|V, W, U)$ is obviously zero.

   - $I(X; Y)$

   $$I(X;Y) = H(X)-H(X|Y) = H(U,V)-H(U) = H(U)+H(V)-H(U) = H(V) = H_v$$

4. 4

5. $D_H(X, Y) \equiv H(X, Y) - I(X; Y)$

   - $D_H(X, Y) \geq 0$

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

   $$H(X,Y) - I(X;Y) = H(Y|X) + H(X|Y) \geq 0$$

   We can afirm that given that conditional entropy is always greater or equal to 0.

   - $D_H(X, X) = 0$

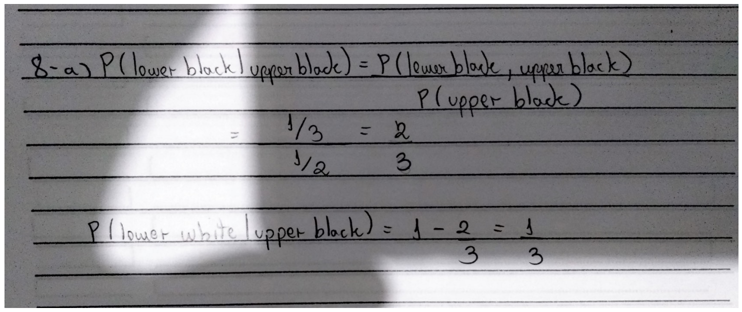
   $$H(X,X)-I(X;X) = H(X)+H(X|X)-(H(X)-H(X|X)) = H(X)+0-H(X)+0 = 0$$

   - $D_H(X, Y) = D_H(Y, X)$ It is easy to proove since conditional entropy and mutual information are symmetric.

   $$H(X,Y) - I(X;Y) = H(Y,X) - I(Y;X) = D_H(Y,X)$$

   - $D_H(X, Z) \leq D_H(X, Y) + D_H(Y, Z)$

6. Handmade exercise.



8-a) $P(\text{lower black} \mid \text{upper black}) = \dfrac{P(\text{lower black}, \text{upper black})}{P(\text{upper black})}$

$= \dfrac{1/3}{1/2} = \dfrac{2}{3}$

$P(\text{lower white} \mid \text{upper black}) = 1 - \dfrac{2}{3} = \dfrac{1}{3}$

7. (a) $\mathcal{P}_z = \{1/2, 1/2\}$

$$\mathcal{P}(z=0) = \mathcal{P}(x=0)\mathcal{P}(y=0) + \mathcal{P}(x=1)\mathcal{P}(y=1)$$

$$\mathcal{P}(z=0) = \mathcal{P}(x=0)\frac{1}{2} + \mathcal{P}(x=1)\frac{1}{2}$$

$$\mathcal{P}(z=0) = \frac{p+1-p}{2} = \frac{1}{2}$$

$$I(Z;X) = H(Z) - H(Z|X) = 1 - 1 = 0$$

(b) For genral p and q we have

$$\mathcal{P}(z=0) = \mathcal{P}(x=0)\mathcal{P}(y=0) + \mathcal{P}(x=1)\mathcal{P}(y=1)$$

$$\mathcal{P}(z=0) = pq + (1-p)(1-q)$$

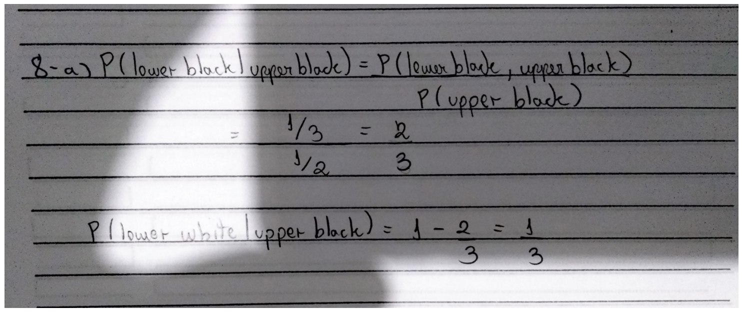
Consequently,

$$\mathcal{P}(z=1) = 1 - pq - (1-p)(1-q) = p(1-q) + q(1-p)$$

Therefora, for general p and q, we have $\mathcal{P}_z = \{pq+(1-p)(1-q), p(1-q)+q(1-p)\}$.
For the mutual information, we have:

$$I(Z;Y) = H(Z) - H(Z|X) = H(pq + (1-p)(1-q)) - H(q)$$

8. • Handmade exercise.



8-a) $P(\text{lower black} \mid \text{upper black}) = \dfrac{P(\text{lower black}, \text{upper black})}{P(\text{upper black})}$

$= \dfrac{1/3}{1/2} = \dfrac{2}{3}$

$P(\text{lower white} \mid \text{upper black}) = 1 - \dfrac{2}{3} = \dfrac{1}{3}$

As we can see, the probability that the lower face of the card is white is 1/3 and the probability that the lower face of the card is black is 2/3.

- Yes, it certainly does. Since the probability distribution over the color of the upper or lower face of a random selected card is uniform, we can say that both $H(U)$ and $H(L)$ are qual to 1 bit. If we know the color of the upper side of the card, we have:

$$H(L|U = black) = H(L|U = white) = H_2(1/3, 2/3) = -\frac{1}{3}log\frac{1}{3} - \frac{2}{3}log\frac{2}{3}$$

$$H(L|U) = P(U = black)H(L|U = black) + P(U = white)H(L|U = white)$$

$$H(L|U) = \frac{1}{2}(log3 - \frac{2}{3})2$$

$$H(L|U) = (log3 - \frac{2}{3})$$

With this value, we can finally compute the mutual information $I(L;U) = H(L) - H(U|L) = 1 - log3 + \frac{2}{3} = \frac{5}{3} - log3$. So we can say that learning the color of the upper side of the card gives a non zero amount of bits of information about the color of the lower side of the card.