Teoria da Informação - Homework 06 Dependent Random Variables

Yuri Niitsuma

REVIEW QUESTIONS.

1) The entropy $H(X) = -\sum_x \log p(x)$ can be interpreted as the uncertainty one has about the random variable X. With that in mind, for each of the items below, give its name, its mathematical formula and explain its meaning in terms of uncertainty.
   a) $H(X, Y)$
   b) $H(X|Y)$
   c) $I(X; Y)$
   d) $I(X; Y|Z)$

Respostas

a) Joint Entropy

$$H(X, Y) = \sum_{x,y} p(x, y) \log \frac{1}{p(x, y)}$$

Quantifica a incerteza quando as variáveis aleatórias X e Y são utilizadas juntas.

b) Conditional entropy

$$\begin{aligned} &H(X, Y) = \\ &= \sum_{y} p(y) H(X|Y = y) \\ &= \sum_{x,y} p(x, y) \log \frac{1}{p(x|y)} \end{aligned}$$

incerteza da variável aleatória X dado o valor Y.

c) Mutual Information

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

quanto o conhecimento de Y reduz a incerteza a X e análogo a simetria.

d) Mutual information with Z?

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

quanto o conhecimento de Y reduz a incerteza de X, assumindo que a variável aleatória Z é conhecida. Também é simétrico.

2) State the following "laws"of information theory.
   a) The chain rule for entropy $H(X_1, X_2, \ldots, X_n)$.
   b) The chain rule for mutual information $I(X_1, X_2, \ldots, X_n; Y)$.
   c) The data-processing inequality (DPI), and explain what it intuitively means.

Respostas

a)

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

b)

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

c) Se temos uma cadeia de Markov:

$$X \to Y \to Z$$

Indicando qual variável aleatória causa qual v.a. através da indicação da seta. Temos, desenvolvendo a regra da cadeia:

$$\begin{aligned}
&p(x, y, z) = \\
&= p(x)p(y|x)\underbrace{p(z|x, y)}_{p(z|x)} \\
&= p(x)p(y|x)p(z|x)
\end{aligned}$$

Então a informação ganha por X afeta Z que a variável Y que é causal. Logo,

$$I(X; Z) \leq I(Y; Z)$$

Exercises.

3) (MacKay 8.1) [Medium] Consider thrre independent random variables u, v, w with entropies H_u, H_v, H_w . Let $X \equiv (U, V)$ and $Y \equiv (V, W)$. What is $H(X, Y)$? What is $H(X|Y)$? What is $I(X; Y)$

Resposta

Utilizando propriedades da entropia e informações do enunciado (como independência), temos:

- $H(X, Y)$?

$$\begin{aligned}
&H\left(X, Y\right) = H\left((U, V), (V, W)\right) = H\left(U, V, V, W\right) = \\
&= H(U) + \underbrace{H(V|U)}_{H(V)} + \underbrace{H(V|U, V)}_{0} + \underbrace{H(W|U, V, V)}_{H(W)} \\
&= H(U) + H(V) + H(W) = H_u + H_v + H_w
\end{aligned}$$

  Note que $H(V|U, V) = 0$ pois não gera incerteza já que V é dados.

- $H(X|Y)$?

$$\begin{aligned}
&H(X|Y) = H\left((U, V)|(V, W)\right) = H\left(U, V|V, W\right) = \\
&= \underbrace{H(U|V, W)}_{H(U)} + \underbrace{H(V|V, W, U)}_{0} = H(U) + 0 = H_u
\end{aligned}$$

- $I(X; Y)$?

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

4) (MacKay 8.2) [Medium] Referring to the definitions of conditional entropy (8.3-8.4), confirm (with an example) that it is possible for $H(X|y = b_k)$ to exceed $H(X)$, but that the average, $H(X|Y)$, is less than $H(X)$. So data are helpful - they do not increase uncertainty, on average.

Resposta

Utilizando o teorema de Bayes e a inequação de Gibbs':

$$
\begin{aligned}
H(X|Y) &\equiv \sum_{y\in A_Y} P(y)\left[\sum_{x\in A_X} P(x|y)\log\frac{1}{P(x|y)}\right]\\
&= \sum_{y\in A_Y}\sum_{x\in A_X} P(y)P(x|y)\log\frac{1}{P(x|y)}\\
&= \sum_{xy\in A_X A_Y} P(x,y)\log\frac{1}{P(x|y)}\\
&= \sum_{xy} P(x|y)\log\frac{1}{\left(\frac{P(y|x)P(x)}{P(y)}\right)}\\
&= \sum_{xy} P(x)P(y|x)\left[\log\frac{P(y)}{P(y|x)}+\log\frac{1}{P(x)}\right]\\
&= \sum_{xy} P(x)P(y|x)\log\frac{P(y)}{P(y|x)}+\sum_{xy} P(x)P(y|x)\log\frac{1}{P(x)}\\
&= \sum_{x}\sum_{y} P(x)P(y|x)\log\frac{P(y)}{P(y|x)}+\sum_{x}\sum_{y} P(x)P(y|x)\log\frac{1}{P(x)}\\
&= \sum_{x} P(x)\sum_{y} P(y|x)\log\frac{P(y)}{P(y|x)}+\sum_{y} P(y|x)\sum_{x} P(x)\log\frac{1}{P(x)}\\
&= \sum_{x} P(x)\underbrace{\sum_{y} P(y|x)\log\frac{P(y)}{P(y|x)}}_{0}+\underbrace{\sum_{x} P(x)\log\frac{1}{P(x)}}_{H(X)}
\end{aligned}
$$

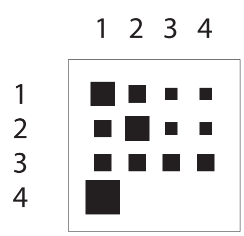
Adquirindo uma soma de entropias entre a distribuição $P(y|x)$ e $P(y)$. Assim, temos:

$$H(X|Y) \leq H(X) + 0$$

Teremos a equalidade somente se X e Y forem independentes.

5) (MacKay 8.6) [Easy] A joint ensemble XY has the following joint distribution.

$P(x,y)$		x 1	2	3	4
	1	$1/8$	$1/16$	$1/32$	$1/32$
y	2	$1/16$	$1/8$	$1/32$	$1/32$
	3	$1/16$	$1/16$	$1/16$	$1/16$
	4	$1/4$	0	0	0



What is the joint entropy $H(X,Y)$? What are the marginal entropies $H(X)$ and $H(Y)$? For each value of y, what is the conditional entropy $H(X|y)$? What is the conditional entropy $H(X|Y)$? What is the conditional entropy of Y given X? What is the mutual information between X and Y?

Resposta

Calculando as marginais

- $P(X = x) = \sum_y P(x,y)$
- $P(X = 1) = \frac{1}{8} + \frac{1}{16} + \frac{1}{16} + \frac{1}{4} = \frac{1}{2}$
- $P(X = 2) = \frac{1}{16} + \frac{1}{8} + \frac{1}{16} + 0 = \frac{1}{4}$
- $P(X = 3) = \frac{1}{32} + \frac{1}{32} + \frac{1}{16} + 0 = \frac{1}{8}$

- $P(X=4) = \frac{1}{32} + \frac{1}{32} + \frac{1}{16} + 0 = \frac{1}{8}$
- $P(Y=y) = \sum_x P(x,y)$
- $P(Y=1) = \frac{1}{8} + \frac{1}{16} + \frac{1}{32} + \frac{1}{32} = \frac{1}{4}$
- $P(Y=2) = \frac{1}{16} + \frac{1}{8} + \frac{1}{32} + \frac{1}{32} = \frac{1}{4}$
- $P(Y=3) = \frac{1}{16} + \frac{1}{16} + \frac{1}{16} + \frac{1}{16} = \frac{1}{4}$
- $P(Y=4) = \frac{1}{4} + 0 + 0 + 0 = \frac{1}{4}$

$$H(X,Y) = \sum_{xy} P(x,y) \log \frac{1}{P(x,y)} = \frac{1}{2} \cdot \log 2 + \ldots + 0 \cdot \log 0 = \frac{27}{8}$$

Calculando:

- $H(X,Y) = \sum_{xy} P(x,y) \log \frac{1}{P(x,y)} = \frac{1}{2} \cdot \log 2 + \ldots + 0 \cdot \log 0 = \frac{27}{8}$
- $H(X) = \frac{1}{2} \log 2 + \frac{1}{4} \log 4 + \frac{1}{8} \log 8 + \frac{1}{8} \log 8 = \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{3}{8} = \frac{7}{4}$
- $H(Y) = \frac{1}{4} \log 4 + \frac{1}{4} \log 4 + \frac{1}{4} \log 4 + \frac{1}{4} \log 4 = \log 4 = 2$

Para calcular $H(X|y)$, temos que calcular todos os $P(x|y)$.

- $P(x=1|y=1) = \frac{1}{2}$
- $P(x=2|y=1) = \frac{1}{4}$
- $P(x=3|y=1) = \frac{1}{8}$
- $P(x=4|y=1) = \frac{1}{8}$
- $P(x=1|y=2) = \frac{1}{4}$
- $P(x=2|y=2) = \frac{1}{2}$
- $P(x=3|y=2) = \frac{1}{8}$
- $P(x=4|y=2) = \frac{1}{8}$
- $P(x=1|y=3) = \frac{1}{4}$
- $P(x=2|y=3) = \frac{1}{4}$
- $P(x=3|y=3) = \frac{1}{4}$
- $P(x=4|y=3) = \frac{1}{4}$
- $P(x=1|y=4) = 1$
- $P(x=2|y=4) = 0$
- $P(x=3|y=4) = 0$
- $P(x=4|y=4) = 0$

Calculando $H(X|y)$:

- $H(X|y=1) = \frac{7}{4}$
- $H(X|y=2) = \frac{7}{4}$
- $H(X|y=3) = 2$
- $H(X|y=4) = 0$ $H(X|Y) = \frac{11}{8}$

Lembrando que:

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

$$H(Y|X) = H(X,Y) - H(X) = \frac{27}{8} - \frac{7}{4} = \frac{13}{8}$$

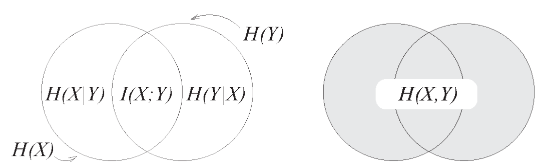
E por último.

$$I(X;Y) = \underbrace{H(X)}_{7/4} - \underbrace{H(X|Y)}_{11/8} = \frac{3}{8}$$

6) (MacKay 8.7) [Medium] Consider the ensemble XYZ in which $A_X = A_Y = A_Z = \{0,1\}$, x and y are independent with $P_X = \{p, 1-p\}$ and $P_Y = \{q, 1-q\}$ and

$$z = (x+y) \bmod 2$$

a) If $q = \dfrac{1}{2}$, what is P_Z? What is $I(Z;X)$?
b) For general p and q, what is P_Z? What is $I(Z;X)$? Notice that this ensemble is related to the binary symmetric channel, with $x =$ input, $y =$ noise, and $z =$ output.



Resposta

a) Se $q = \frac{1}{2}, P_Z = \{1/2, 1/2\}$ e $I(Z;X) = H(Z) - H(Z|X) = 1 - 1 = 0$

b) No geral q e p, $P_Z = \{pq + (1-p)(1-q), p(1-q) + q(1-p)\}$. A informação mútua é $I(Z;X) = H(Z) - H(Z|X) = H_2(pq + (1-p)(1-q)) - H_2(q)$

7) (MacKay 8.9) [Hard] Prove this theorem by considering an ensemble WDR in which w is the state of the world, d is data gathered, and r is the processed data, so that these three variables form a *Markov chain*

$$w \to d \to r,$$

that is, the probability $P(w,d,r)$ can be written as

$$P(w,d,r) = P(w)P(d|w)P(r|d)$$

Show that the average information that R conveys about W, $I(W;R)$, is less than or equal to the average information that D conveys about W, $I(W;D)$.
This theorem is as much a caution about our definition of 'information' as it is a caution about data processing!

Resposta

Para qualquer ensemble conjunto XYZ, a regra da cadeia para informação mútua é:

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

Agora, no caso $w \to d \to r$, w e r são independentes dado d, então $I(W;R|D) = 0$. Utilizando a regra da cadeia duas vezes, temos:

$$I(W;D,R) = I(W;D) + \underbrace{I(W;R|D)}_{0}$$
$$I(W;D,R) = I(W;R) + I(W;D|R)$$

Com isso, temos:

$$I(W;D,R) = I(W;D) + \underbrace{I(W;R|D)}_{0}$$
$$I(W;D,R) = I(W;R) + I(W;D|R)$$
$$I(W;D) = I(W;R) + I(W;D|R) \Rightarrow I(W;D) \geq I(W;R) \Rightarrow I(W;R) - I(W;D) \leq 0$$

8) (MacKay 8.10) [Medium] The three cards.
   a) One card is white on both faces; one is black on both faces; and one is white on one side and black on the other. The three cards are shuffled and their orientations randomized. One card is drawn and placed on the table. The upper face is black. What is the colour of its lower face? (Solve the inference problem.)
   b) Does seeing the top face convey *information* about the colour of the bottom face? Discuss the *information contents* and *entropies* in this situation. Let the value of the upper face's colour be u and the value of the lower face's colour be l. Imagine that we draw a random card and learn both u and l. What is the entropy of u, $H(U)$? What is the entropy of l, $H(L)$? What is the mutual information between U and L, $I(U;L)$?

Resposta

O problema é semelhante das bolas na urna.

a) Seja $\{w,b\}$ representando a face branca e preta da carta respectivamente, e l,u indicando quais eventos *lower* e *upper* da carta. Utilizando Bayes, temos:

$$p(l=w|u=b) = \frac{p(l=w,u=b)}{p(w=b)} = \frac{\left(\frac{1}{6}\right)}{\left(\frac{3}{6}\right)} = \frac{1}{3}$$

Já o complementar é:

$$p(l=b|u=b) = 1 - p(l=w|u=b) = \frac{2}{3}$$

É mais provável que parte de baixo da carta tenha a cor preta.

b)

$$H(U) = H(L) = 1\ bit$$

1 bit pois só responde uym nível da árvore de perguntas.

$$H(L|U=b) = \frac{1}{3}\log 3 + \frac{2}{3}\log\frac{3}{2} = \frac{1}{3}\left(\log 3 + 2\log 3 - 2\log 2\right) = \frac{1}{3}(3\log 3 - 2) = \log 3 - \frac{2}{3}$$

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

Informação mútua:

$$I(U;L) = H(L) - H(L|U) = 1 - \left(\log 3 - \tfrac{2}{3}\right) = 1 + \tfrac{2}{3} - \log 3 = \tfrac{5}{3} - \log 3$$