Exercise 2 – Information Theory

by Millis Sahar, ID 300420379

Q1

a. Mutual Information - Show that: $I\left(X;Y\right) = H\left(X\right) - H\left(X|Y\right) = H\left(Y\right) - H\left(Y|X\right)$

Prove $I\left(X;Y\right) = H\left(X\right) - H\left(X|Y\right)$:

$I(x;y) =$

$\sum_x \sum_y p(x,y) log\left(\frac{p(x,y)}{p(x)p(y)}\right) =$

$= \sum_x \sum_y p(x,y) log\left(\frac{p(y)p(x|y)}{p(x)p(y)}\right) =$

$= \sum_x \sum_y p(x,y) log\left(\frac{p(x|y)}{p(x)}\right) =$

$= \sum_x \sum_y p(x,y)\left(logp(x|y) - logp(x)\right) =$

$= \sum_x \sum_y p(x,y) logp(x|y) - \sum_x \sum_y p(x,y) logp(x) =$

$= -H\left(X|Y\right) - \sum_x \left(\sum_y p(x,y) logp(x)\right) =$

$= -H\left(X|Y\right) - \sum_x p(x,y) logp(x) =$

$= -H\left(X|Y\right) + H\left(x\right) =$

$= H\left(X\right) - H\left(X|Y\right)$

Prove $I(X;Y) = H(Y) - H(Y|X)$: (almost the same way)

$I(x;y) =$

$\sum_x \sum_y p(x,y) log\left(\frac{p(x,y)}{p(x)p(y)}\right) =$

$= \sum_x \sum_y p(x,y) log\left(\frac{p(x)p(y|x)}{p(x)p(y)}\right) =$

$= \sum_x \sum_y p(x,y) log\left(\frac{p(y|x)}{p(y)}\right) =$

$= \sum_x \sum_y p(x,y)\left(logp(y|x) - logp(y)\right) =$

$= \sum_x \sum_y p(x,y) logp(y|x) - \sum_x \sum_y p(x,y) logp(y) =$

$= -H(Y|X) - \sum_y \left(\sum_x p(x,y) logp(y)\right) =$

$= -H(Y|X) - \sum_y p(x,y) logp(y) =$

$= -H(Y|X) + H(Y) =$

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

b. Mutual information - Conditional vs unconditional

i. Give an example for three random variables such that: $I\big(X;Y|Z\big) < I\big(X;Y\big)$

$$p(Z) = \begin{cases} 0 \text{ with probability of } 0.8 \\ 1 \text{ with probability of } 0.2 \end{cases}$$

$$A = X, Y \sim Uniform(0,1), Uniform(0,1)$$
$$B = X, Y \sim Norm(\mu_x, \sigma_x^2), Norm(\mu_y, \sigma_y^2)$$

$$X, Y \sim \begin{cases} \text{A when Z=0} \\ \text{B when Z=1} \end{cases}$$

$$I(X;Y) = 0.5I(A)_{\text{going to zero}} + 0.5I(B) \to 0.5I(B)$$
$$I(X;Y|Z) = 0.8I(A)_{\text{going to zero}} + 0.2I(B) \to 0.2I(B)$$
$\Rightarrow$ Then: $0.5I(B) > 0.2I(B)$

Therefore: $I(X;Y) > I(X;Y|Z)$

ii. Give an example for three random variables such that: $I\left(X; Y|Z\right) > I\left(X; Y\right)$

it's basically the same example, with a different probability...

$$p(Z) = \begin{cases} 0 \text{ with probability of } 0.4 \\ 1 \text{ with probability of } 0.6 \end{cases}$$

$$A = X, Y \sim Uniform(0,1), Uniform(0,1)$$
$$B = X, Y \sim Norm(\mu_x, \sigma_x^2), Norm(\mu_y, \sigma_y^2)$$

$$X, Y \sim \begin{cases} \text{A when Z=0} \\ \text{B when Z=1} \end{cases}$$

$$I(X;Y) = 0.5I(A)_{\text{going to zero}} + 0.5I(B) \to 0.5I(B)$$
$$I(X;Y|Z) = 0.4I(A)_{\text{going to zero}} + 0.6I(B) \to 0.6I(B)$$
$\Rightarrow$ Then: $0.5I(B) < 0.6I(B)$

Therefore: $I(X;Y) < I(X;Y|Z)$

Q2

Let X, Y, Z three random variables who form a Markov Chain $X \rightarrow Y \rightarrow Z$
Show that X, Y, Z are conditionally independent given Z, i.e $p\big(x, z|y\big) = p\big(x|y\big)p\big(z|y\big)$

I'll show $p(x, z|y) = p(z|y)p(x|y)$, meaning X,Y are conditionally independent given Z.

Using bayes: $p(x, z|y) = \frac{p(x,y,z)}{p(y)}$

and given the Markov Chain: $X \rightarrow Y \rightarrow Z \Rightarrow \quad p\big(x, y, z\big) = p(x)p(y|x)p(z|y)$

$$p(x, z|y) = \frac{p(x,y,z)}{p(y)} = \frac{p(x)p(y|x)p(z|y)}{p(y)} = p(z|y)p(x|y)$$

Q3

Let p(x,y) be given by:

Pr(x,y)	x_1	x_2
y_1	1/3	1/3
y_2	0	1/3

Find:

a. $H(X), H(Y)$
b. $H(X|Y), H(Y|X)$
c. $H(X,Y)$
d. $H(X) - H(Y)$
e. $H(X), H(Y)$
f. $I(X;Y)$
g. Draw a Venn diagram that illustrates the quantities stated in the above bullets ("a" to "f")

Answers:

a.
$H\left(X\right) = H(\frac{1}{3}, \frac{2}{3}) = -\frac{1}{3}log\frac{1}{3} - \frac{2}{3}log\frac{2}{3} = 0.92$
$H\left(Y\right) = H(\frac{1}{3}, \frac{2}{3}) = -\frac{1}{3}log\frac{1}{3} - \frac{2}{3}log\frac{2}{3} = 0.92$

b.
$H\left(X|Y\right) = -\sum_x \sum_y p(x,y)log\left(p(x|y)\right) = -\frac{1}{3}log\frac{1}{2} - \frac{1}{3}log\frac{1}{2} - \frac{1}{3}log(1) = -\frac{2}{3}log\frac{1}{2} = -\frac{2}{3} \cdot (-1) = \frac{2}{3} = 0.67$
$H\left(Y|X\right) = -\sum_x \sum_y p(x,y)log\left(p(y|x)\right) = -\frac{1}{3}log\frac{1}{2} - \frac{1}{3}log\frac{1}{2} - \frac{1}{3}log(1) = -\frac{2}{3}log\frac{1}{2} = -\frac{2}{3} \cdot (-1) = \frac{2}{3} = 0.67$

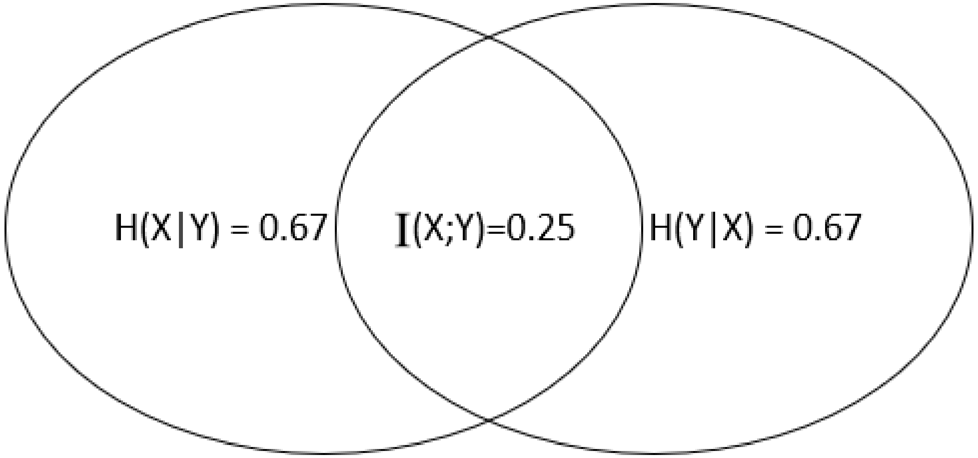
c.
$H\left(X,Y\right) = H\left(X\right) + H\left(Y|X\right) = 0.92 + 0.67 = 1.59$

d.
$H\left(X\right) - H\left(Y\right) = 0.92 - 0.92 = 0$

e. Duplication

f.
$I\left(X;Y\right) = H\left(X\right) - H\left(X|Y\right) = 0.92 - 0.67 = 0.25$

g.



Q4

Grouping rule for Entropy

Let $p = (p_1, p_2, \ldots, p_m)$ be a probability distribution on m elements $\left(i.e. 0 \le p_i\right)$ and $\sum_i p_i = 1$.
Define a new distribution on q on $m-1$ elements such that the distribution on the first $m-2$ elements is identical, and the probability of last element in q is the sum of the last two probabilitis in p , i.e.

$$q_1 = p_1, \ \ q_2 = p_2, \ \ \ldots \ , \ q_{m_2} = p_{m-2}, \ \ q_{m-1} = p_{m-1} + p_m$$

Show that: $H\left(p\right) = H\left(q\right) + \left(P_{m-1} + p_m\right) H\left(v\right)$
where $v = \left(\frac{p_{m-1}}{p_{m-1}+p_m}, \frac{p_m}{p_{m-1}+p_m}\right)$ is a binary probability distribution.

Using entropy definition: $H\left(p\right) = -\sum_{i=1}^{m} p_i log\left(p_i\right) = -\sum_{i=1}^{m-1} p_i log\left(p_i\right) - p_m log\left(p_m\right) = -\sum_{i=1}^{m-2} p_i log\left(p_i\right) - p_{m-1} log\left(p_{m-1}\right) - p_m log\left(p_m\right)$
And $q_1 = p_1,\ q_2 = p_2,\ \ldots\ ,\ q_{m_2} = p_{m-2},\ q_{m-1} = p_{m-1} + p_m$

$$Therefore:$$

$$H\left(p\right) =$$

$$= -\sum_{i=1}^{m-2} p_i log\left(p_i\right) - p_{m-1} log\left(p_{m-1}\right) - p_m log\left(p_m\right) =$$

$$= -\sum_{i=1}^{m-2} p_i log\left(p_i\right) - p_{m-1} log\left(p_{m-1}\right) - p_m log\left(p_m\right) + 0 =$$

$$= -\sum_{i=1}^{m-2} p_i log\left(p_i\right) - p_{m-1} log\left(p_{m-1}\right) - p_m log\left(p_m\right) + \left(q_{m-1} log\left(q_{m-1}\right) - q_{m-1} log\left(q_{m-1}\right)\right) =$$

$$= -\sum_{i=1}^{m-2} p_i log\left(p_i\right) + q_{m-1} log\left(q_{m-1}\right) - p_{m-1} log\left(p_{m-1}\right) - p_m log\left(p_m\right) - q_{m-1} log\left(q_{m-1}\right) =$$

$$= \left(-\sum_{i=1}^{m-2} p_i log\left(p_i\right) - q_{m-1} log\left(q_{m-1}\right)\right) - p_{m-1} log\left(p_{m-1}\right) - p_m log\left(p_m\right) + q_{m-1} log\left(q_{m-1}\right) =$$

$$= H\left(q\right) - p_{m-1} log\left(p_{m-1}\right) - p_m log\left(p_m\right) + q_{m-1} log\left(q_{m-1}\right) =$$

$$= H\left(q\right) - p_{m-1} log\left(p_{m-1}\right) - p_m log\left(p_m\right) + \left(p_{m-1} + p_m\right) log\left(p_{m-1} + p_m\right) =$$

$$= H\left(q\right) - p_{m-1} log\left(p_{m-1}\right) - p_m log\left(p_m\right) + p_{m-1} log\left(p_{m-1} + p_m\right) + p_m log\left(p_{m-1} + p_m\right) =$$

$$= H\left(q\right) + \left(p_{m-1} log\left(p_{m-1} + p_m\right) - p_{m-1} log\left(p_{m-1}\right)\right) + \left(p_m log\left(p_{m-1} + p_m\right) - p_m log\left(p_m\right)\right) =$$

$$= H\left(q\right) + \left(p_{m-1} log\left(p_{m-1} + p_m\right) - p_{m-1} log\left(p_{m-1}\right)\right) + \left(p_m log\left(p_{m-1} + p_m\right) - p_m log\left(p_m\right)\right) =$$

$$= H\left(q\right) + p_{m-1} log\left(\frac{p_{m-1}+p_m}{p_{m-1}}\right) + p_m log\left(\frac{p_{m-1}+p_m}{p_m}\right) =$$

$$= H\left(q\right) - p_{m-1} log\left(\frac{p_{m-1}}{p_{m-1}+p_m}\right) - p_m log\left(\frac{p_m}{p_{m-1}+p_m}\right) =$$

$$= H\left(q\right) - \frac{p_{m-1}}{1} log\left(\frac{p_{m-1}}{p_{m-1}+p_m}\right) - \frac{p_m log}{1}\left(\frac{p_m}{p_{m-1}+p_m}\right) =$$

$$= H\left(q\right) - (p_{m-1} + p_m)\frac{p_{m-1}}{p_{m-1}+p_m} log\left(\frac{p_{m-1}}{p_{m-1}+p_m}\right) - (p_{m-1} + p_m)\frac{p_m}{p_{m-1}+p_m} log\left(\frac{p_m}{p_{m-1}+p_m}\right) =$$

$$= H\left(q\right) - \left(p_{m-1} + p_m\right)\left(\frac{p_{m-1}}{p_{m-1}+p_m} log\left(\frac{p_{m-1}}{p_{m-1}+p_m}\right) - \frac{p_m}{p_{m-1}+p_m} log\left(\frac{p_m}{p_{m-1}+p_m}\right)\right) =$$

$$= H\left(q\right) - \left(p_{m-1} + p_m\right)\left(-H\left(V\right)\right) =$$

$$= H\left(q\right) + \left(p_{m-1} + p_m\right) \cdot H\left(V\right)$$

Q5

In general, Relative Entropy is not symmetric, namely $D\left(p||q\right) \neq D\left(q||p\right)$ **.**
Give an example for two not identical distributions, $p \neq q$ **, such that** $D\left(p||q\right) = D\left(q||p\right)$ **.**

I'll define 2 different(!) distributions P,Q:

	x=0	x=1
p(x)	0.2	0.8
q(x)	0.8	0.2

Therefore:

$D\left(p||q\right) = \sum_x p(x)log(\frac{p(x)}{q(x)}) = p(x = 0)log(\frac{p(x=0)}{q(x=0)}) + p(x = 1)log(\frac{p(x=1)}{q(x=1)}) = 0.2log(\frac{0.2}{0.8}) + 0.8log(\frac{0.8}{0.2}) = 0.2 \cdot (-2) + 0.8 \cdot (2) = -0.4$
$+ 1.6 = 1.2$

$D\left(q||p\right) = \sum_x q(x)log(\frac{q(x)}{p(x)}) = q(x = 0)log(\frac{q(x=0)}{p(x=0)}) + q(x = 1)log(\frac{q(x=1)}{p(x=1)}) = 0.8log(\frac{0.8}{0.2}) + 0.2log(\frac{0.2}{0.8}) = 0.8 \cdot (2) + 0.2 \cdot (-2) = 1.6 - 0.4$
$= 1.2$

$\Rightarrow D\left(p||q\right) = D\left(q||p\right)$

Q6

Relative Entropy $D\left(p||q\right)$ and chi-square $\left(\chi^2\right)$

Show that the χ^2 statistics $\chi^2 = \sum_x \frac{\left(p\left(x\right)-q\left(x\right)\right)^2}{q\left(x\right)}$ is (twice) the first term is the Taylor series expansion of $D\left(p||q\right)$ around q .

Thus, $D\left(p||q\right) = \frac{1}{2}\chi^2 + \ldots$ namely chi-square is a first order approximation of the relative Entropy.

<u>Hint:</u> Write $\frac{p}{q} = 1 + \frac{p-q}{q}$ and expand the $log(\cdot)$.

Answer:

$D(p||q) =$

$= \sum_x p \cdot log(1 + \frac{p-q}{q}) =$

$= \sum_x p \cdot log\big(1 + \frac{p-q}{q}\big) \cong$

$\cong \sum_x p \cdot \left(log\big(1 + \frac{p-q}{q}\big) + \frac{\frac{p}{q^2}}{1+\frac{p-q}{q}} (p-q)^1 + 0.5 \frac{1}{q^2} (p-q)^2 \right) =$

$= \sum_x p \cdot \left(0 + \frac{\frac{p}{q^2}}{1+\frac{p-q}{q}} (p-q) + 0.5 \frac{1}{q^2} (p-q)^2 \right) \Big|_{p=q} =$

$= \sum_x \left(q \cdot \frac{\frac{q}{q^2}}{1+\frac{q-q}{q}} (p-q) + q \cdot 0.5 \frac{1}{q^2} (p-q)^2 \right) =$

$= \sum_x \left(\frac{\frac{q^2}{q^2}}{1+0} (p-q) + 0.5 \frac{q}{q^2} (p-q)^2 \right) =$

$= \sum_x \left((p-q) + 0.5 \frac{(p-q)^2}{q} \right) =$

$= \sum_x (p-q) + 0.5 \sum_x \frac{(p-q)^2}{q} =$

$= \sum_x (p-q) + 0.5 \chi^2$

Q7

Min Relative Entropy under constraints
Let $p(x), q(x), x \in \chi$ two probability mass functions, and let $\mathrm{f_1}, \mathrm{f_2}, \ldots, \mathrm{f_n}$ where $\mathrm{f_j} : \chi \to \mathscr{R}$ be feature function.

Given expectation constrains on the features $\forall j : \sum_x p\big(x\big)\mathrm{f_i} = \alpha_j$, what is the p^* that minimize the relative Entropy $D(p||q)$?

Solve the following $p^* = argmax D\big(p||q\big)$ where $\mathscr{P} = \left\{\forall j : \quad p : E_p\left[\mathrm{f_i}\right] = \alpha_j\right\}$

Use Lagrange multipliers to derive an explicit from to p^* .

$$p^* = argmax \;\; D\big(p||q\big) \text{ where } \mathscr{P} = \left\{\forall j : \quad p : E_p\left[\mathrm{f_i}\right] = \alpha_j\right\}$$

$$\Rightarrow argmin \;\; D\big(p||q\big) \text{ where } \sum_x p(x) f_j(x) - \alpha_j = 0$$

$$= argmin_p \sum_x p(x) log(\tfrac{p(x)}{q(x)}) - \sum_j \lambda_j \big(p(x) f_j(x) - \alpha_j\big) =$$

$$= argmin_p \sum_x p(x) log(\tfrac{p(x)}{q(x)}) - \sum_x \sum_j \lambda_j \big(p(x) f_j(x) - \alpha_j\big) =$$

$$= argmin_p \sum_x [p(x) log(\tfrac{p(x)}{q(x)}) - \sum_j \lambda_j \big(p(x) f_j(x) - \alpha_j\big)]$$

Thanks for reading :)