COMP 2610 Tut 3

* Binomial, Bernoulli distribution.

* Likelihood function

$$L(\theta) = p(D/\theta) = \prod_{i=1}^{N} p(x_i/\theta)$$

* Find maximum likelihood estimate by equating

$$\frac{\partial L}{\partial \theta} = 0 \quad \{ \text{ where } L = \log p(D/\theta)$$

$$L_{\log}(\theta) = \log\left(\prod_{i=1}^{N} p(x_i/\theta)\right)$$

$$= \log\left((p(x_1))^{n_1} (p(x_2))^{n_2} \cdots (p(x_n))^{n_n} \right)$$

$$= n_1 \log p(x_1) + n_2 \log (p(x_2)) + \cdots n_n \log (p(x_n))$$

$$= \sum_{\forall i} n_i \log (p(x_i))$$

* Information content of an outcome of RV

$$I(x) = \log_2\left(\frac{1}{p(x)}\right) = -\log_2(p(x)).$$

* Entropy. of a RV

$$H(X) = E_x(h(x))$$

$$= \sum p(x) . h(x)$$

$$= -\sum p(x) \log (p(x)).$$

Conditional Entrophy

$$H(Y/X=x) = -\sum_y p(y/X=x) \log(p(y/X=x))$$

$$H(Y/X) = +\sum_x p(X=x) \cdot H(Y/X=x).$$

$$= -\sum p(x) \sum p(y/X=x) . \log(p(y/X=x))$$

Joint Entroplny

$$H(X, Y) = \sum_y \sum_x p(x, y) \log\left(\frac{1}{p(x, y)}\right)$$

Chain Rule

$$H(X, Y) = H(X) + H(Y/X) = H(Y) + H(X/Y)$$

Relative Entrophy

$$D_{KL}(p \| q) = \sum p(x)\left(\log\left(\frac{1}{q(x)}\right) - \log\left(\frac{1}{p(x)}\right)\right)$$

$$= \sum p(x) \log\left(\frac{p(x)}{q(x)}\right)$$

Mutual Information

$$I(X, Y) = D_{KL}(p(X, Y) \| p(X)p(Y))$$

$$= \sum\sum p(x, y) \log\left(\frac{p(x, y)}{p(x).q(x)}\right)$$

Q1. $p(x=1) = \theta/2$ | out of N observation
$p(x=2) = \theta/2$ | n_1
$p(x=3) = 1-\theta.$ | n_2
| n_3

a. $L(\theta) = \prod_{i=1}^{N} p(x_i/\theta)$

$= \left(\frac{\theta}{2}\right)^{n_1} \cdot \left(\frac{\theta}{2}\right)^{n_2} (1-\theta)^{n_3}$; $n_1+n_2+n_3=N.$

$= \left(\frac{\theta}{2}\right)^{n_1+n_2} (1-\theta)^{n_3}$

b. $\{3,3,1,2,3,2,2,1,3,1\}$

$\therefore n_1 = 3$
$n_2 = 3$
$n_3 = 4$ $\Big\}$ $N=10.$

$L_{log}(\theta) = \sum n_i \log(p(x_i))$

$= 3\log(\theta/2) + 3\log(\theta/2) + 4\log(1-\theta)$
$= 6\log(\theta/2) + 4\log(1-\theta).$

$\frac{\partial L_{log}}{\partial\theta} = \frac{6}{\ln 2} \times \frac{1/2}{\theta/2} + \frac{4}{\ln 2} \times \frac{-1}{1-\theta}$ $\qquad \frac{\partial \ln(x)}{\partial x} = \frac{1}{x}$

for maximum likelyhood estimate
$\frac{\partial L}{\partial\theta} = 0.$

$$\frac{6}{\theta} - \frac{4}{1-\theta} = 0$$

$$3(1-\theta) = 2\theta$$

$$3 - 3\theta = 2\theta$$

$$\theta = 3/5 = 0.6$$

Q2.

		X				
		1	2	3	4	Tot.
Y	1	0	0	1/8	1/8	4/16
	2	1/8	1/16	1/16	0	4/16
	3	1/8	1/8	0	0	4/16
	4	0	1/16	1/16	1/8	4/16.
	Tot	4/16	4/16	4/16	4/16	

Table give $p(X=x_i, Y=y_i)$

a.

Consider $P(X=1, Y=1) = 0$

However $P(X=1) \times P(Y=1) = 4/16 \times 4/16 \neq 0$

$\neq P(X=1, Y=1)$

$\therefore$ X & Y are not independent.

b.(i) $H(X) = -\sum p(x) \log(p(x))$

$$= -\frac{4}{16} \log_2\left(\frac{4}{16}\right) \times 4$$

$$= -\log\left(\frac{1}{4}\right) = -\log(2^{-2}) = 2 \text{ bits.}$$

(ii). $H(Y) = -\sum p(y) \log(p(y))$

$$= -\frac{4}{16} \times \log_2\left(\frac{4}{16}\right) \times 4$$

$$= +2 \text{ bits.}$$

iii $H(X/Y) = \sum_y p(Y=y) \; H(X/Y=y)$

; $H(X/Y=y) = -\sum p(x/Y=y) . \log(p(x/Y=y))$

$p(x/Y=1) = 0, 0, 1/2, 1/2$
$p(x/Y=2) = 1/2, 1/4, 1/4, 0$
$p(x/Y=3) = 1/2, 1/2, 0, 0$
$p(x/Y=4) = 0, 1/4, 1/4, 1/2.$

$\therefore H(X/Y) = \underbrace{1/4 \times H(0, 0, 1/2, 1/2)}_{\times 2} + \underbrace{1/4 H(1/2, 1/4, 1/4, 0)}_{\times 2}$

$$= \frac{1}{2}\left[-\frac{1}{2}\log\left(\frac{1}{2}\right) \times 2\right] + \frac{1}{2}\left[-\frac{1}{2}\log\left(\frac{1}{2}\right) - \frac{2}{4}\log\left(\frac{1}{4}\right)\right]$$

$$= \frac{1}{2}\left[+1\right] + \frac{1}{2}\left[\frac{1}{2} + \frac{1}{\cancel{2}} \times +\cancel{2}\right]$$

$$= \frac{1}{2} + \frac{3}{4} = \frac{5}{4}$$

iv $H(Y/X)$??.

$$H(X,Y) = H(X) + H(Y/X) = H(Y) + H(X/Y).$$

$$\cancel{2} + H(Y/X) = \cancel{2} + 5/4.$$

$$\therefore H(Y/X) = 5/4.$$

v. $H(X,Y) = H(X) + H(Y/X)$

$$= 2 + 5/4$$

$$= 13/4$$

Q3. 13 cards (v) (J, Q, K are face cards).
4 suits.
2 colors.

a. $h(c=\text{red}, v=K) = -\log p(c=\text{red}, v=K)$

$$p(c=\text{red}, v=K) = p(v=K, \heartsuit) + p(v=K, \diamondsuit)$$

$$= \frac{1}{52} + \frac{1}{52} = \frac{1}{26}$$

$\therefore$ $h(c=\text{red}, v=K) = -\log\left(\frac{1}{26}\right)$

$= 4.7$ bits.

b. $h(v=K/f=1) = -\log(p(v=K/f=1))$

$$p(v=K/f=1) = \frac{p(v=K, f=1)}{p(f=1)}$$

$$= \frac{p(v=K)}{(3\times4)/52}$$

$$= \frac{4/52}{12/52}$$

$$= 1/3$$

$\therefore h(v=K/f=1) = -\log\left(\frac{1}{3}\right)$

$= 1.585$

c. $H(S) = -\sum p(s) \log(p(s))$

$= -\frac{1}{4} \times \log\left(\frac{1}{4}\right) \times 4$

$= 2$ bits

$H(V,S) = -\sum\sum p(v,s) \log(p(v,s))$

$= -\left[\frac{1}{52} \log\left(\frac{1}{52}\right) \times 52\right]$

$= 5.70$ bits

Q4. We are going to use the chain rule.

$$H(X) + H(Y/X) = H(Y) + H(X/Y).$$

So if we can find whether $H(Y/X)$ & $H(X/Y)$ are zero or >0, then can find the inequality relationship.

Consider $Y = 2^X$.

we know that this is a 1 to 1 mapping expression., i.e., for every unique value of X, we can find a unique Y.

$\therefore$ if $p(X = x_i) = \theta$, then $p(Y = 2^{x_i}) = \theta$.

Considering this

$p(Y = 2^{x_i}/X = x_i) = 1$ and $p(Y = y_i/X = x_i) = 0$ for $y_i \neq 2^{x_i}$.

Generalizing this observation, we can write.

if we can represent Y as a function of X, i.e., $Y = f(X)$, the

$$p(Y = y_i/X = x_i) = \begin{cases} 1 \; ; & y_i = f(x_i) \\ 0 \; ; & \text{elsewhere.} \end{cases} \quad \text{—} \; R_1$$

As for
$$H(Y/X) = -\sum P(X=x_i) \sum p(Y=y_i/X=x_i) \log\left(p(Y=y_i/X=x_i)\right).$$

when $P(Y=y_i/X=x_i) = 1$, $\log(P(Y=y_i/X=x_i)) = 0$
$P(Y=y_i/X=x_r) = 0$, $p.(Y=y_i/X=x_i)\log[P(Y=y_i/X=x_i)] = 0.$

$\therefore$ We have $H(Y/x) = 0.$

Generalizing this observation we have.

if $B = f(A)$

$H(B/A) = 0.$ ⟵ R_2.

Using R_2, we can conclude

$H(Y/\cancel{X}) = 0$ since $y = 2^x = f(x)$

Also $H(X/Y) = 0$, since $x = \log_2(x) = g(Y).$

$\therefore$ $H(X) = H(Y)$ //.

b. $Y = \cos(x)$

we caen conclude

$H(Y/X) = 0$, since $Y = \cos(x) = f(x)$.

But we cannot express X as a function of Y, since $Y = \cos(x)$ is a many to 1 mapping ~~e~~expression.

$\therefore H(X/Y) \neq 0$.

$$\therefore H(X) + \underbrace{H(Y/X)}_{=0} = H(Y) + \underbrace{H(X/Y)}_{\neq 0}.$$

$\therefore H(x) = H(Y) + H(x/Y)$.

$\therefore H(X) > H(Y)$ //