ECE 515 Fall 2025 Assignment 2 Solutions

1. $$\rho = 1 - \frac{H(Y|X)}{H(X)}$$

a) $$\rho = \frac{H(X) - H(Y|X)}{H(X)}$$

$$= \frac{H(Y) - H(Y|X)}{H(X)} = \frac{I(X;Y)}{H(X)}$$

b) since $0 \le H(Y|X) \le H(Y) = H(X)$

$$0 \le \rho \le 1$$

c) $\rho = 0$ iff $I(X;Y) = 0$ (X and Y are independent)

d) $\rho = 1$ iff $H(Y|X) = 0$

(X and Y have a one-to-one relationship)

2. FGD discrete RVs

Prove the validity of the following

a) $I(GD;F) \geq I(D;F)$

$I(GD;F) = I(D;F) + I(G;F|D)$

Since $I(G;F|D) \geq 0$

it must be that $I(GD;F) \geq I(D;F)$

equality holds iff conditioned on D G and F are independent

b) $H(GD|F) \geq H(D|F)$

$H(GD|F) = H(D|F) + H(G|DF)$

since $H(G|DF) \geq 0$ it must be that

$H(GD|F) \geq H(D|F)$

equality holds iff G is completely determined by DF

c) $I(G;F|D) \geq I(D;F|G) + I(F;G) - I(D;F)$

Rearranging gives

$I(G;F|D) + I(D;F) \geq I(D;F|G) + I(F;G)$

both sides are equal to $I(DG;F)$

thus the inequality is satisfied with equality

d) $H(FDG) - H(DF) \leq H(GD) - H(D)$

simplifying gives

$H(G|DF) \leq H(G|D)$

given $H(G|D) = H(G|DF) + I(G;F|D)$

then $H(G|DF) \leq H(G|DF) + I(G;F|D)$

since $I(G;F|D) \geq 0$ the inequality is true

equality occurs when $I(G;F|D) = 0$

G and F are independent given D

3. The following equiprobable messages are used over a BSC with crossover probability p

$Y = 0000$ is received

a) $$I(M_1;0) = \log\left[\frac{p(\text{1st digit}=0 \mid M_1 \text{ transmitted})}{p(\text{1st digit}=0)}\right]$$

$$= \log \frac{1-p}{1/2} = \log[2(1-p)]$$

b) note that the first 3 bits transmitted are independent and equiprobable

since the channel is memoryless, the first 3 output bits are also independent and equiprobable

Thus

$$I(M_1;010) = \log[2(1-p)]$$

and

$$I(M_1;0100) = \log[2(1-p)]$$

Averaging over all codewords, the probability of receiving

$$Y = 0000$$

is

$$\frac{1}{8}\left[(1-p)^4 + 6(1-p)^2 p^2 + p^4\right]$$

Thus $P(0 \mid 000) = (1-p)^4 + 6(1-p)^2 p^2 + p^4$

$$I(M_1; 0 \mid 000) = \log\left[\frac{1-p}{(1-p)^4 + 6(1-p)^2 p^2 + p^4}\right]$$

$$= \log\left[\frac{p(\text{last digit} = 0 \mid M_1 \text{ transmitted})}{p(\text{last digit} = 0 \mid 000)}\right]$$

4a) Decision tree for accepting a job offer

4 attributes
- Salary (S)
- Commute (C)
- Worklife Balance (W)
- Remote Working (R)

the output is Accept Job (A)

22 training inputs $p(yes) = \frac{12}{22}$ $p(no) = \frac{10}{22}$

$$H(A) = -\frac{12}{22}\log\frac{12}{22} - \frac{10}{22}\log\frac{10}{22}$$

$$= .994 \text{ bit}$$

for Salary (S)

8 high	7 yes	1 no
8 medium	4 yes	4 no
6 low	1 yes	5 no

$$H(A|S) = -\frac{8}{22}\left(\frac{7}{8}\log\frac{7}{8} + \frac{1}{8}\log\frac{1}{8}\right) - \frac{8}{22}\left(\frac{4}{8}\log\frac{4}{8} + \frac{4}{8}\log\frac{4}{8}\right)$$

$$-\frac{6}{22}\left(\frac{1}{6}\log\frac{1}{6} + \frac{5}{6}\log\frac{5}{6}\right)$$

$$= .198 + .364 + .177 = .739 \text{ bit}$$

$I(A;S) = H(A) - H(A|S) = .994 - .739 = .255$ bit

for Commute (C)

8	short	6 yes	2 no
7	medium	4 yes	3 no
7	long	2 yes	5 no

$$H(A|C) = -\frac{8}{22}\left(\frac{6}{8}\log\frac{6}{8} + \frac{2}{8}\log\frac{2}{8}\right) - \frac{7}{22}\left(\frac{4}{7}\log\frac{4}{7} + \frac{3}{7}\log\frac{3}{7}\right)$$
$$- \frac{7}{22}\left(\frac{2}{7}\log\frac{2}{7} + \frac{5}{7}\log\frac{5}{7}\right)$$
$$= .883 \text{ bit}$$

$I(A;C) = H(A) - H(C|A) = .994 - .883 = .111$ bit

for Worklife Balance (W)

9	good	6 yes	3 no
7	average	5 yes	2 no
6	poor	1 yes	5 no

$$H(A|W) = -\frac{9}{22}\left(\frac{6}{9}\log\frac{6}{9} + \frac{3}{9}\log\frac{3}{9}\right) - \frac{7}{22}\left(\frac{5}{7}\log\frac{5}{7} + \frac{2}{7}\log\frac{2}{7}\right)$$
$$- \frac{6}{22}\left(\frac{1}{6}\log\frac{1}{6} + \frac{5}{6}\log\frac{5}{6}\right)$$

$H(A|W) = .828$ bit

$I(A;W) = H(A) - H(A|W) = .994 - .828 = .166$ bit

for Remote Working (R)

12 yes   8 yes 4 no
10 no   4 yes 6 no

$$H(A|R) = -\frac{12}{22}\left(\frac{8}{12}\log\frac{8}{12} + \frac{4}{12}\log\frac{4}{12}\right) - \frac{10}{22}\left(\frac{4}{10}\log\frac{4}{10} + \frac{6}{10}\log\frac{6}{10}\right)$$

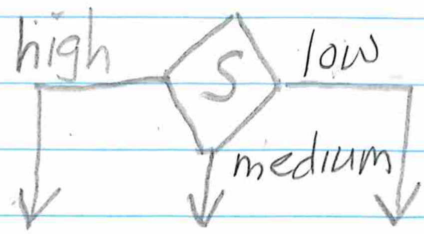
$$= .942 \text{ bit}$$

$I(A;R) = H(A) - H(A|R) = .994 - .942 = .052$ bit

Summary

$I(A;S) = .255$ bit
$I(A;C) = .111$ bit
$I(A;W) = .166$ bit
$I(A;R) = .052$ bit

S provides the largest information gain so use this for the root node



branch for $S = high$ 7 yes   1 no

$$H(A) = -\frac{7}{8}\log\frac{7}{8} - \frac{1}{8}\log\frac{1}{8} = .544 \text{ bit}$$

for Commute (C)

3	short	3 yes 0 no
2	medium	2 yes 0 no
3	long	2 yes 1 no

$H(A|C) = .344$ bit $I(A;C) = .544 - .344 = .200$ bit

for Worklife Balance (W)

3	good	3 yes 0 no
3	average	3 yes 0 no
2	poor	1 yes 1 no

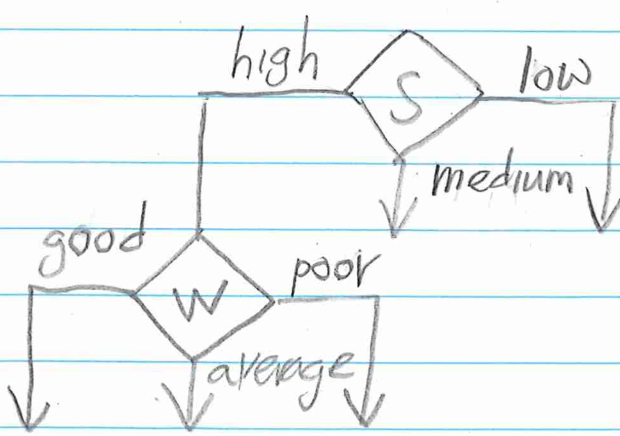
$H(A|W) = .250$ bit $I(A;W) = .544 - .250 = .294$ bit for Remote Working (R)

4 yes 4 yes 0 no
4 no 3 yes 1 no

$H(A|R) = .406$ bit $I(A;R) = .544 - .406 = .038$ bit

Summary $I(A;C) = .200$ bit
$I(A;W) = .294$ bit
$I(A;R) = .038$ bit

W provides the largest information gain



branch for S = medium 4 yes 4 no

$$H(A) = -\frac{4}{8}\log\frac{4}{8} - \frac{4}{8}\log\frac{4}{8} = 1.0 \text{ bit}$$

for Commute (C)

3 short	2 yes 1 no
3 medium	2 yes 1 no
2 long	0 yes 2 no

$H(A|C) = .689$ bit $I(A;C) = 1 - .689 = .311$ bit

for Worklife Balance (W)

3 good	2 yes 1 no
3 average	2 yes 1 no
2 poor	0 yes 2 no

$H(A|W) = .689$ bit $I(A;W) = 1 - .689 = .311$ bit

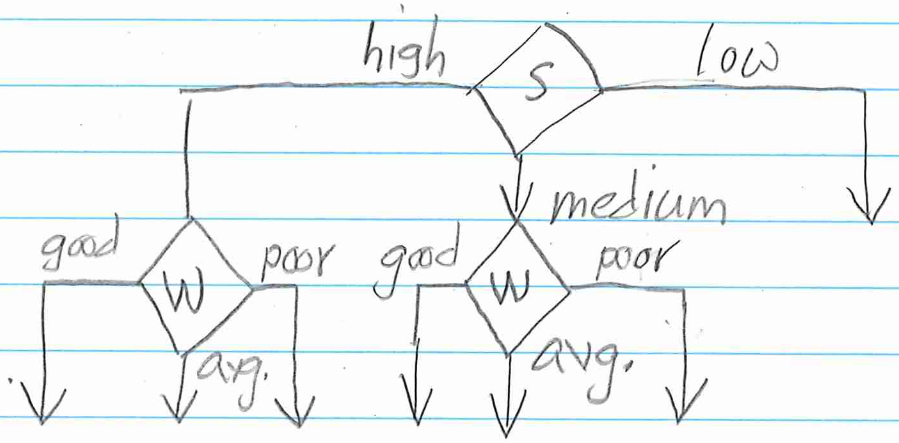
for Remote Working (R)

5 yes	3 yes 2 no
3 no	1 yes 2 no

$H(A|R) = .938$ bit $\quad I(A;R) = 1 - .938 = .062$ bit

Summary
$I(A;C) = .311$ bit
$I(A;W) = .311$ bit
$I(A;R) = .062$ bit

C and W provide the largest information gain
Choose W



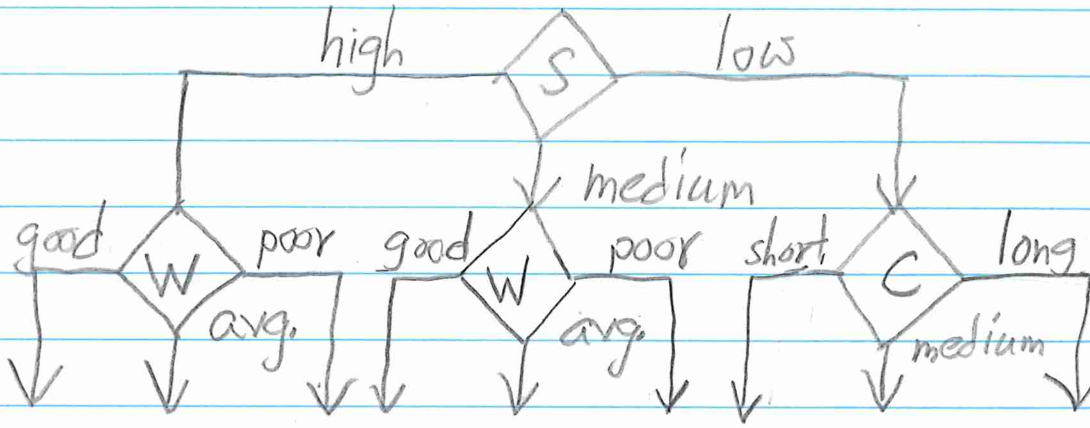
branch for S = low 1 yes 5 no

$$H(A) = -\frac{1}{6}\log\frac{1}{6} - \frac{5}{6}\log\frac{5}{6} = .650 \text{ bit}$$

for Commute (C) $H(A|C) = .333$ bit
for Worklife Balance $H(A|W) = .459$ bit
for Remote Working $H(A|R) = .459$ bit

$I(A;C) = .650 - .333 = .317$ bit
$I(A;W) = .650 - .459 = .191$ bit
$I(A;R) = .650 - .459 = .191$ bit

C provides the largest information gain so choose C



branch for S=high W=good 3 yes 0 no

this is a pure (leaf) node so no split is required

branch for S=high W=average 3 yes 0 no

this is a pure (leaf) node so no split is required branch for S = high W = poor 1 yes 1 no

$H(A) = 1$ bit

for Commute (C)

1 short 1 yes 0 no
0 medium 0 yes 0 no
1 long 0 yes 1 no

$H(A|C) = 0$ bit $I(A;C) = 1-0 = 1$ bit

for Remote Working (R)

1 yes 1 yes 0 no
1 no 0 yes 1 no

$H(A|R) = 0$ bit $I(A;R) = 1-0 = 1$ bits

C and R provide the largest information gain so choose R branch for S= medium W = good 2 yes 1 no

$H(A) = .918$

for Commute (C)

1 short 1 yes 0 no
1 medium 1 yes 0 no
1 long 0 yes 1 no

$H(A|C) = 0$ bit $I(A;C) = .918 - 0 = .918$ bit

for Remote Working (R)

2 yes 2 yes 0 no
1 no 0 yes 1 no

$H(A|R) = 0$ bit $I(A;R) = .918 - 0 = .918$ bit

C and R provide the largest information gain so choose R branch for S= medium W= average 2yes 1 no

$H(A) = .918$ bit

for Commute (C)

1 short 1 yes 0 no
1 medium 1 yes 0 no
1 long 0 yes 1 no

$H(A|C) = 0$ bit $I(A;C) = .918 - 0 = .918$ bit

for Remote Working (R)

2 yes 1 yes 1 no
1 no 1 yes 0 no

$H(A|R) = .333$ bit $I(A;R) = .918 - .333 = .585$ bit

C provides the largest information gain
so choose C

branch for S= medium W = poor 0 yes 2 no

pure node so no split is required

branch for S=low C= short 1 yes 1 no

$H(A) = 1$ bit

for Worklife Balance (W)

1	good	1 yes 0 no
1	average	0 yes 1 no
0	poor	0 yes 0 no

$H(A|W) = 0$ bit $I(A;W) = 1-0 = 1$ bit

for Remote Working (R) 1 yes 1 no

1	yes	1 yes 0 no
1	no	0 yes 1 no

$H(A|R) = 0$ bit $I(A;R) = 1-0 = 1$ bit

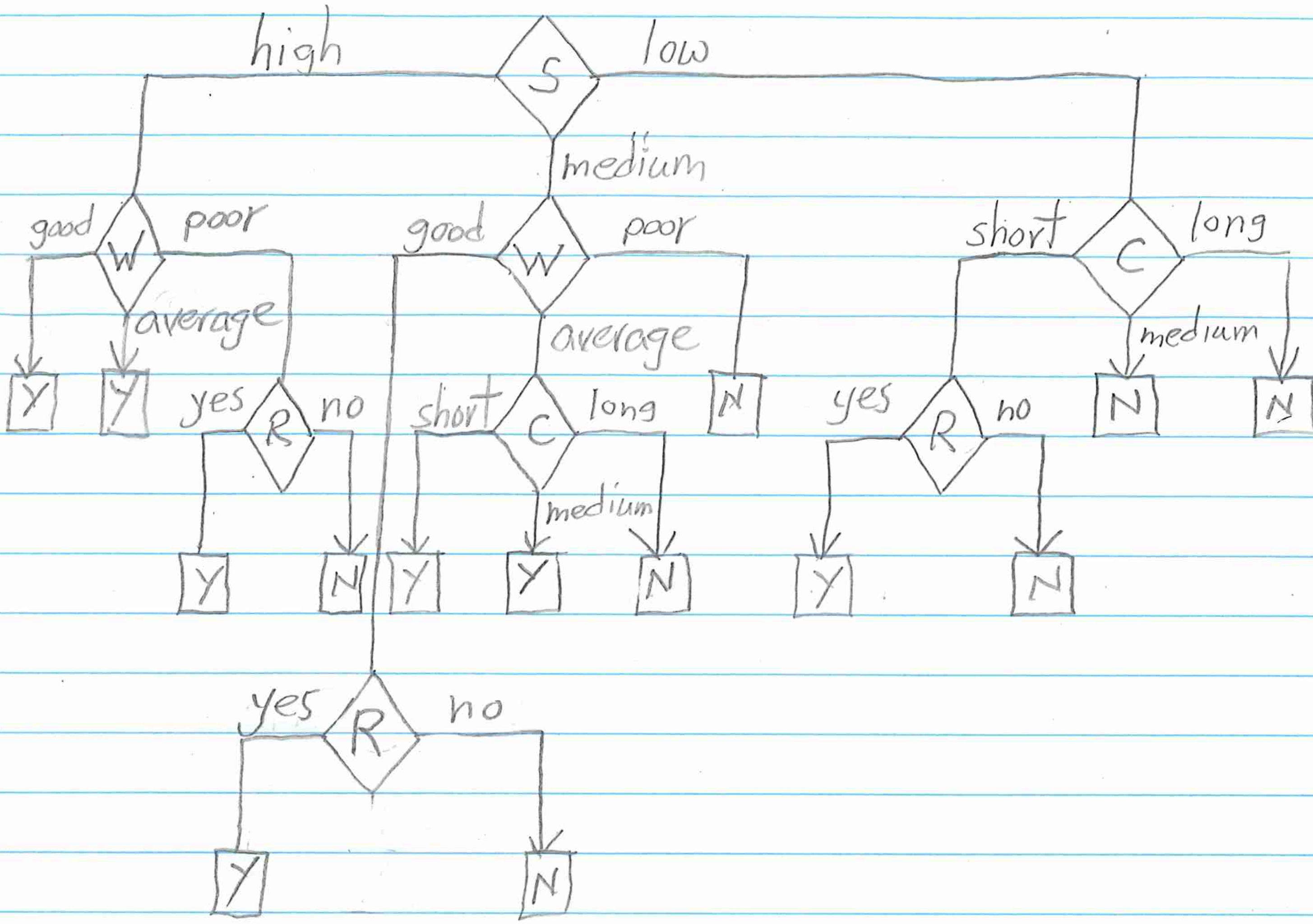
R and W provide the largest information gain so choose R branch for S=low C=medium 0 yes 2 no

this is a pure (leaf) node so no split is required

branch for S=low C=long 0 yes 2 no

this is a pure (leaf) node so no split is required

the branches for S=medium W=average and C=short, medium, and long are all pure (leaf) nodes so no split is required



high
S
low
medium
good
W
poor
average
Y
Y
yes
R
no
Y
N
good
W
poor
average
short
C
long
medium
Y
Y
N
N
short
C
long
medium
yes
R
no
Y
N
N
N
yes
R
no
Y
N

b) test inputs

23	yes
24	yes
25	no
26	no
27	no
28	no

5. $h(p) = -p\log_2 p - (1-p)\log_2(1-p)$ in bits

or in nats

$= -p\ln p - (1-p)\ln(1-p)$

Considering the shape of

$-p\ln p$ and $-(1-p)\ln(1-p)$

the areas under the curves are the same

thus the average is

$$-2\int_0^1 p\ln p \, dp$$

using integration by parts

$$= -2\left[\frac{p^2}{2}\ln p - \frac{p^2}{4}\right]_0^1 = \frac{1}{2} \text{ nats}$$

$\therefore$ the average entropy is $\frac{1}{2}\log_2 e = .721$ bits let $u = \ln p \rightarrow du = \frac{1}{p}\,dp$

$dv = p\,dp \rightarrow v = \frac{p^2}{2}$

we have

$$\int u\,dv = uv - \int v\,du$$

$$= -2\left[\frac{p^2}{2}\ln p\Big|_0^1 - \int_0^1 \frac{p^2}{2}\frac{1}{p}\,dp\right]$$

$$= -\left[p^2 \ln p\Big|_0^1 - \int_0^1 p\,dp\right]$$

$$= -\left[p^2 \ln p\Big|_0^1 - \frac{p^2}{2}\Big|_0^1\right]$$

$$= \quad 0 + \frac{1}{2} = \frac{1}{2} \text{ nats}$$

or $\frac{1}{2}\log_2 e \approx .721$ bits

6. Binary Erasure Channel

the channel forward transition matrix is

$$\begin{bmatrix} 1-p & 0 \\ p & p \\ 0 & 1-p \end{bmatrix}$$

Calculate $I(X;Y)$ in terms of ω

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

$$p(y=0) = (1-p)\omega$$
$$p(y=1) = (1-p)\bar{\omega}$$
$$p(y=e) = p\omega + p\bar{\omega} = p$$
$$H(Y) = -(1-p)\omega\log((1-p)\omega) - (1-p)\bar{\omega}\log((1-p)\bar{\omega}) - p\log p$$

$$H(Y) = -(1-p)\omega \log(1-p) - (1-p)\omega \log \omega$$
$$- (1-p)\bar{\omega} \log(1-p) - (1-p)\bar{\omega} \log \bar{\omega}$$
$$- p \log p$$

$$H(Y) = -(1-p)\log(1-p) - p\log p$$
$$-(1-p)\left[\omega \log \omega + \bar{\omega} \log \bar{\omega}\right]$$

$$H(Y|X) = -\sum_i p(x_i) \sum_j p(y_j|x_i) \log p(y_j|x_i)$$

$$= -\omega\left[(1-p)\log(1-p) + p\log p\right]$$
$$- \bar{\omega}\left[p \log p + (1-p)\log(1-p)\right]$$
$$= -(1-p)\log(1-p) - p\log p$$

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

$$= -(1-p)\left[\omega \log \omega + \bar{\omega} \log \bar{\omega}\right]$$

b) since p is a channel parameter that cannot be changed $I(X;Y)$ is maximized by maximizing

$$-\omega \log \omega - \bar{\omega} \log \bar{\omega}$$

which occurs at $\omega = \frac{1}{2}$

$\therefore$ the maximum is

$$I(X;Y)_{max} = (1-p)$$

7. Random variable X outcomes $\{a, b, c, d\}$

$p(X) = 5/8, 1/8, 1/8, 1/8$

$q(X) = 3/8, 1/4, 1/4, 1/8$

$H(p) = -\frac{5}{8}\log\frac{5}{8} - \frac{3}{8}\log\frac{1}{8} = .424 + 1.125 = 1.549$ bits

$H(q) = -\frac{3}{8}\log\frac{3}{8} - \frac{1}{2}\log\frac{1}{4} - \frac{1}{8}\log\frac{1}{8} = .531 + 1 + .375 = 1.906$ bits

$$D(p\|q) = \sum_i p(x_i)\log\left[\frac{p(x_i)}{q(x_i)}\right]$$

$$= \frac{5}{8}\log\frac{5}{3} + \frac{1}{8}\log\frac{1}{2} + \frac{1}{8}\log\frac{1}{2} + \frac{1}{8}\log 1$$

$$= .461 - .250 + 0 = .211 \text{ bit}$$

$$D(q\|p) = \sum_i q(x_i)\log\left[\frac{q(x_i)}{p(x_i)}\right]$$

$$= 3/8\log\frac{3}{5} + \frac{1}{4}\log 2 + \frac{1}{4}\log 2 + \frac{1}{8}\log 1$$

$$= -.276 + .250 + .250 + 0 = .224 \text{ bit}$$

$\therefore\ D(p\|q) \neq D(q\|p)$

b) let $p(x) = 5/8, 1/4, 1/8$

$q(x) = 1/8, 1/4, 5/8$

$$D(p\|q) = \frac{5}{8}\log 5 + \frac{1}{4}\log 1 + \frac{1}{8}\log 1/5$$

$$= 1.451 + 0 - .290 = 1.161 \text{ bits}$$

$$D(q\|p) = \frac{1}{8}\log 1/5 + 1/4 \log 1 + \frac{5}{8}\log 5 = D(p\|q)$$