

(1)

ELEC 515 Fall 2026 Assignment 2 Solutions

1. a) Let T = top and B = bottom

$$I(T; B) = H(B) - H(B|T)$$

since $p(T = \text{heads}) = p(T = \text{tails}) = \frac{1}{2}$

$$H(B) = \log_2 2 = 1 \text{ bit}$$

$$H(B|T) = 0 \quad \text{so} \quad I(T; B) = 1 \text{ bit}$$

b) Let F be the side of the cube facing usThere are 4 possibilities for the top T
which are equally probable

$$\text{Thus} \quad I(T; F) = H(T) - H(T|F)$$

$$= \log_2 6 - \log_2 4$$

$$= \log_2 3 - 1 = 1.585 - 1 = .585 \text{ bit}$$

since T has a uniform distribution on

$$\{1, 2, \dots, 6\}$$

2. Course Notes Problem 1.2

$$X = \{x_0, x_1, x_2\} \quad p(x_0) = p(x_1) = \frac{1}{4} \quad p(x_2) = \frac{1}{2}$$

a) channel transition matrices

$$X \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \quad X \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Y Z

$$b) H(X) = - \sum_{i=0}^2 p(x_i) \log p(x_i) = 1.5 \text{ bits}$$

X	Y	Z	$p(x_i)$	$p(y_j)$	$p(z_k)$	$p(y_j, z_k)$
x_0	y_0	z_0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$
x_1	y_1	z_0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$
x_2	y_0	z_1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$
x_2	y_1	z_1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{4}$

$$p(y_0) = p(x_0) + \frac{1}{2} p(x_2) = \frac{1}{4} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2}$$

$$p(y_1) = p(x_1) + \frac{1}{2} p(x_2) = \frac{1}{4} + \frac{1}{2} + \frac{1}{2} = \frac{1}{2}$$

$$H(Y) = 1 \text{ bit}$$

$$p(z_0) = p(x_0) + p(x_1) = \frac{1}{2} \quad p(z_1) = p(x_2) = \frac{1}{2}$$

$$H(Z) = 1 \text{ bit}$$

$$H(YZ) = - \sum_{i=0}^1 \sum_{j=0}^1 p(y_i, z_j) \log p(y_i, z_j)$$

$$= 2 \text{ bits} = H(Y) + H(Z)$$

∴ no statistical dependence between Y and Z

c) calculate $I(X; Y)$, $I(X; Z)$

$$I(X; Y|Z), I(X; YZ)$$

$$\begin{aligned}
 p(x_0, y_0) &= \frac{1}{4} \\
 p(x_0, y_1) &= 0 \\
 p(x_1, y_0) &= 0 \\
 p(x_1, y_1) &= \frac{1}{4} \\
 p(x_2, y_0) &= \frac{1}{4} \\
 p(x_2, y_1) &= \frac{1}{4}
 \end{aligned}$$

$$\begin{aligned}
 p(x_0, z_0) &= \frac{1}{4} \\
 p(x_0, z_1) &= 0 \\
 p(x_1, z_0) &= \frac{1}{4} \\
 p(x_1, z_1) &= 0 \\
 p(x_2, z_0) &= 0 \\
 p(x_2, z_1) &= \frac{1}{2}
 \end{aligned}$$

$$I(X; Y) = \sum_{i=0}^2 \sum_{j=0}^1 p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i) p(y_j)}$$

$$= \frac{1}{4} \log 2 + \frac{1}{4} \log 2 + \frac{1}{4} \log 1 + \frac{1}{4} \log 1 = \frac{1}{2} \text{ bit}$$

$$I(X; Z) = \sum_{i=0}^2 \sum_{k=0}^1 p(x_i, z_k) \log \frac{p(x_i, z_k)}{p(x_i) p(z_k)}$$

$$= \frac{1}{4} \log 2 + \frac{1}{4} \log 2 + \frac{1}{2} \log 2 = 1 \text{ bit}$$

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

$$H(X|Z) = - \sum_{i=0}^2 \sum_{k=0}^1 p(x_i, z_k) \log \frac{p(x_i, z_k)}{p(z_k)}$$

$$= - \frac{1}{4} \log \frac{1}{2} - \frac{1}{4} \log \frac{1}{2} - \frac{1}{2} \log 1 = \frac{1}{2} \text{ bit}$$

$$H(X|YZ) = - \sum_{i=0}^2 \sum_{j=0}^1 \sum_{k=0}^1 p(x_i, y_j, z_k) \log \frac{p(x_i, y_j, z_k)}{p(y_j, z_k)}$$

$$p(x_0, y_0, z_0) = \frac{1}{4} = p(x_0)$$

$$p(x_2, y_0, z_1) = \frac{1}{4} = \frac{1}{2} p(x_2)$$

$$p(x_1, y_1, z_0) = \frac{1}{4} = p(x_1)$$

$$p(x_2, y_1, z_1) = \frac{1}{4} = \frac{1}{2} p(x_2) \quad \text{all others are 0}$$

$$H(X|YZ) = -\frac{1}{4} \log 1 - \frac{1}{4} \log 1 - \frac{1}{4} \log 1 - \frac{1}{4} \log 1 = 0$$

$$\therefore I(X; Y|Z) = H(X|Z) - H(X|YZ) = \frac{1}{2} \text{ bit}$$

$$I(X; YZ) = \sum_{i=0}^2 \sum_{j=0}^1 \sum_{k=0}^1 p(x_i, y_j, z_k) \log \frac{p(x_i, y_j, z_k)}{p(x_i) p(y_j, z_k)}$$

$$= \frac{1}{4} \log 4 + \frac{1}{4} \log 2 + \frac{1}{4} \log 4 + \frac{1}{4} \log 2$$

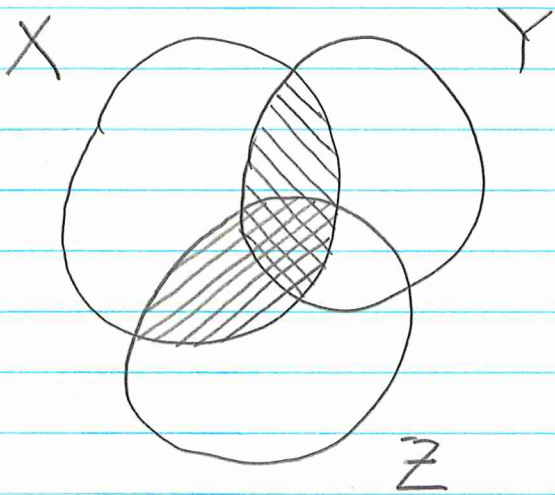
$$= \frac{3}{2} \text{ bits}$$

(6)

d) interpretation

$$H(X) = \frac{3}{2} \text{ bits} \quad I(X; Y) = \frac{1}{2} \text{ bit} \quad I(X; Z) = 1 \text{ bit}$$

$$I(X; Y|Z) = \frac{1}{2} \text{ bit} \quad I(X; YZ) = \frac{3}{2} \text{ bits}$$



since $I(X; Y|Z) = I(X; Y)$

knowing Z provides no additional information between X and Y

$$\therefore I(X; Y; Z) = 0$$

since $I(X; YZ) = H(X)$

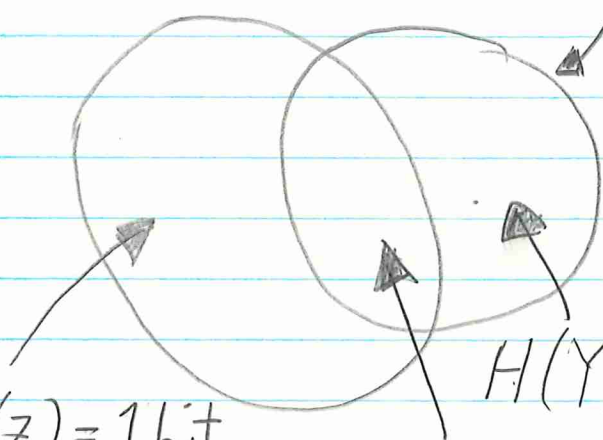
knowing Y and Z provides complete information about X

(7)

Y provides $\frac{1}{2}$ bit of information about X

Z provides 1 bit of information about X

$$H(Y) = 1 \text{ bit}$$



$$I(X; Z) = H(X|Y) = H(Z) = 1 \text{ bit}$$

$$H(Z|X) = 0$$

$$H(Y|X) = \frac{1}{2} \text{ bit}$$

$$I(X; Y) = \frac{1}{2} \text{ bit}$$

3. Let ABC be a joint ensemble of three RVs

Prove the validity of the following

$$a) I(AB;C) \geq I(A;C)$$

$$I(AB;C) = I(A;C) + I(B;C|A)$$

$$\text{since } I(B;C|A) \geq 0$$

$$\text{it must be that } I(AB;C) \geq I(A;C)$$

Equality holds iff conditioned on A

B and C are independent

$$b) H(AB|C) \geq H(A|C)$$

$$H(AB|C) = H(A|C) + H(B|AC)$$

since $H(B|AC) \geq 0$ it must be that

$$H(AB|C) \geq H(A|C) \text{ with equality iff}$$

B is uniquely determined by A and C

(9)

$$c) I(A; C|B) \geq I(C; B|A) - I(C; B) + I(A; C)$$

rearranging gives

$$I(A; C|B) + I(C; B) \geq I(C; B|A) + I(A; C)$$

clearly both sides are equal to $I(AB; C)$

so the inequality is always satisfied with equality

$$d) H(ABC) - H(AB) \leq H(AC) - H(A)$$

$$\text{or } H(C|AB) \leq H(C|A)$$

$$\text{Given } H(C|A) = H(C|AB) + I(C; B|A)$$

$$\text{then } H(C|AB) \leq H(C|AB) + I(C; B|A)$$

since $I(C; B|A) \geq 0$ the inequality is true

Equality occurs iff given A

B and C are independent

4 a) $X \rightarrow 5X$ is a one-to-one mapping so

$$H(5X) = H(X)$$

b) by the data processing inequality

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

c) conditioning cannot increase entropy so

$$H(X_0 | X_1) \geq H(X_0 | X_1, X_2)$$

d) $H(XY) \leq H(X) + H(Y)$ with equality iff

$$I(X; Y) = 0$$

therefore
$$\frac{H(XY)}{H(X) + H(Y)} \leq 1$$

5. Decision tree for buying a product prediction

4 attributes

Age Group (A)
Income Level (I)
Student (S)
Credit Rating (C)

the output is Buy (B)

15 training inputs

$$p(\text{yes}) = \frac{9}{15} \quad p(\text{no}) = \frac{6}{15}$$

$$H(B) = -\frac{9}{15} \log \frac{9}{15} - \frac{6}{15} \log \frac{6}{15} = .971 \text{ bit}$$

for Age Group A

5 young	3 yes	2 no
5 middle	3 yes	2 no
5 senior	3 yes	2 no

$$\begin{aligned} H(B|A) &= -\frac{5}{15} \left(\frac{3}{5} \log \frac{3}{5} + \frac{2}{5} \log \frac{2}{5} \right) \\ &\quad - \frac{5}{15} \left(\frac{3}{5} \log \frac{3}{5} + \frac{2}{5} \log \frac{2}{5} \right) \\ &\quad - \frac{5}{15} \left(\frac{3}{5} \log \frac{3}{5} + \frac{2}{5} \log \frac{2}{5} \right) \end{aligned}$$

$$= .3237 + .3237 + .3237 = .971 \text{ bit}$$

$$I(B; A) = H(B) - H(B|A) = .971 - .971 = 0 \text{ bits}$$

for Income Level (I)

5	high	2 yes	3 no
6	medium	5 yes	1 no
4	low	2 yes	2 no

$$H(B|I) = -\frac{5}{15} \left(\frac{2}{5} \log \frac{2}{5} + \frac{3}{5} \log \frac{3}{5} \right) \\ - \frac{6}{15} \left(\frac{5}{6} \log \frac{5}{6} + \frac{1}{6} \log \frac{1}{6} \right) \\ - \frac{4}{15} \left(\frac{2}{4} \log \frac{2}{4} + \frac{2}{4} \log \frac{2}{4} \right)$$

$$= .3237 + .2600 + .2667 = .850 \text{ bit}$$

$$I(B; I) = H(B) - H(B|I) = .971 - .850 = .121 \text{ bit}$$

for Student (S)

6	yes	6 yes	0 no
9	no	3 yes	6 no

$$H(B|S) = -\frac{6}{15} \left(\frac{6}{6} \log \frac{6}{6} + \frac{0}{6} \log \frac{0}{6} \right) \\ - \frac{9}{15} \left(\frac{3}{9} \log \frac{3}{9} + \frac{6}{9} \log \frac{6}{9} \right)$$

$$= 0 + .5510 = .551 \text{ bit}$$

$$I(B;S) = H(B) - H(B|S) = .971 - .551 = .420 \text{ bit}$$

for Credit Rating (C)

8 good	6 yes	2 no
7 poor	3 yes	4 no

$$H(B|C) = -\frac{8}{15} \left(\frac{6}{8} \log \frac{6}{8} + \frac{2}{8} \log \frac{2}{8} \right) - \frac{7}{15} \left(\frac{3}{7} \log \frac{3}{7} + \frac{4}{7} \log \frac{4}{7} \right)$$

$$= .4327 + .4598 = .892 \text{ bit}$$

$$I(B;C) = H(B) - H(B|C) = .971 - .892 = .078 \text{ bit}$$

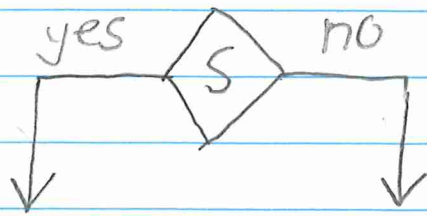
Summary

$I(B;A) = 0 \text{ bits}$
$I(B;I) = .121 \text{ bit}$
$I(B;S) = .420 \text{ bit}$
$I(B;C) = .078 \text{ bit}$

S provides the largest information gain

therefore this is selected as the

root node



branch for $S = \text{yes}$ 6 training inputs

6 yes 0 no

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

Student = yes $\rightarrow$ Buy = yes

branch for $S = \text{no}$ 9 training inputs

3 yes 6 no

$$H(B) = -\frac{3}{9} \log \frac{3}{9} - \frac{6}{9} \log \frac{6}{9} = .918 \text{ bit}$$

for A

3 young	1 yes	2 no
3 middle	1 yes	2 no
3 senior	1 yes	2 no

$$\begin{aligned}H(B|A) &= -\frac{3}{9}\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right) \\&= -\frac{3}{9}\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right) \\&= -\frac{3}{9}\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right) \\&= .918 \text{ bit}\end{aligned}$$

$$I(B;A) = H(B) - H(B|A) = .918 - .918 = 0 \text{ bits}$$

for C

5 good	3 yes	2 no
4 poor	0 yes	4 no

$$\begin{aligned}H(B|C) &= -\frac{5}{9}\left(\frac{3}{5}\log\frac{3}{5} + \frac{2}{5}\log\frac{2}{5}\right) \\&\quad -\frac{4}{9}\left(\frac{0}{4}\log\frac{0}{4} + \frac{4}{4}\log\frac{4}{4}\right) \\&= .5394 = .539 \text{ bit}\end{aligned}$$

$$I(B;C) = H(B) - H(B|C) = .918 - .539 = .379 \text{ bit}$$

for I

4 high	1 yes	3 no
3 medium	2 yes	1 no
2 low	0 yes	2 no

$$H(B|I) = -\frac{4}{9} \left(\frac{1}{4} \log \frac{1}{4} + \frac{3}{4} \log \frac{3}{4} \right) \\ - \frac{3}{9} \left(\frac{2}{3} \log \frac{2}{3} + \frac{1}{3} \log \frac{1}{3} \right) \\ - \frac{2}{9} \left(\frac{0}{2} \log \frac{0}{2} + \frac{2}{2} \log \frac{2}{2} \right)$$

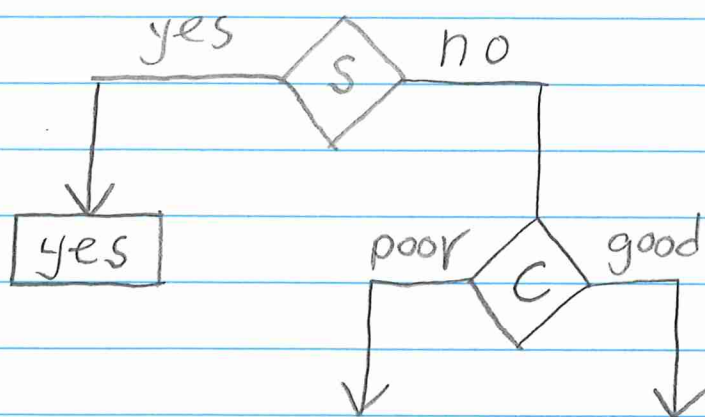
$$= .3606 + .3081 + 0 = .667 \text{ bit}$$

$$I(B;I) = H(B) - H(B|I) = .918 - .667 = .251 \text{ bit}$$

Summary

$$I(B;A) = 0 \text{ bits}$$
$$I(B;C) = .379 \text{ bit}$$
$$I(B;I) = .251 \text{ bit}$$

C provides the greatest information gain
so use this as the rule



branch for $C = \text{poor}$ 4 training inputs

0 yes 4 no

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

branch for $C = \text{good}$

3 yes 2 no

$$H(B) = -\frac{3}{5} \log \frac{3}{5} - \frac{2}{5} \log \frac{2}{5} = .971 \text{ bit}$$

the remaining attributes are A and I

for A

2 young	1 yes	1 no
2 middle	1 yes	1 no
1 senior	1 yes	1 no

$$H(B|A) = -\frac{2}{5} \left(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2} \right) \\ - \frac{2}{5} \left(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2} \right) \\ - \frac{1}{5} (1 \log 1 + 0 \log 0)$$

$$= .4000 + .4000 + 0 = .800 \text{ bit}$$

$$I(B;A) = H(B) - H(B|A) = .971 - .800 = .171 \text{ bit}$$

for I

2 high	1 yes	1 no
2 medium	2 yes	0 no
1 low	0 yes	1 no

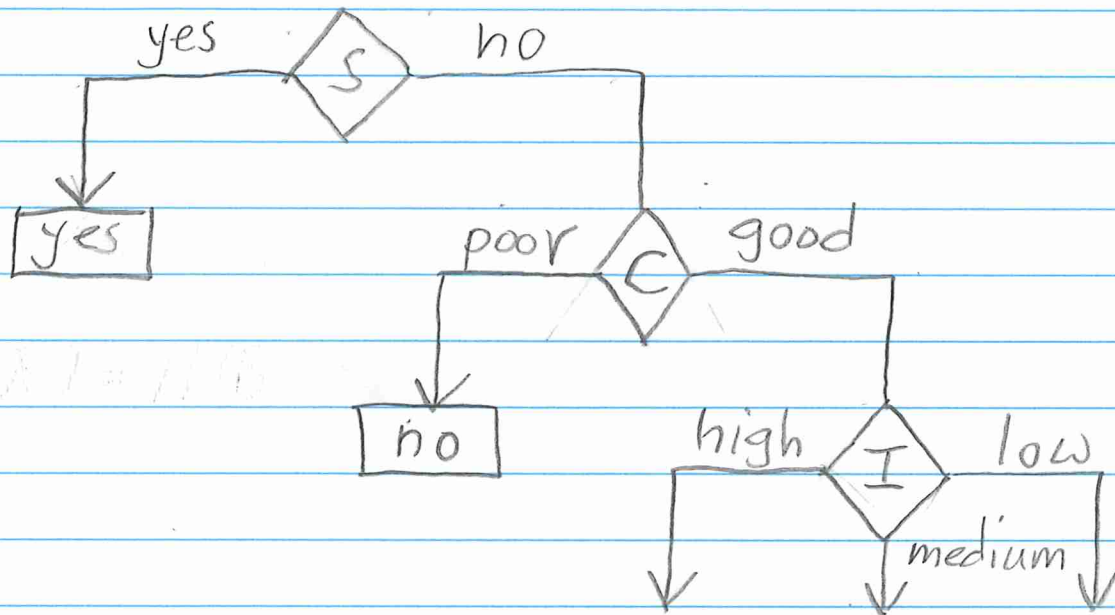
$$H(B|I) = -\frac{2}{5} \log \left(\frac{1}{2} \log \frac{1}{2} + \frac{1}{2} \log \frac{1}{2} \right) + \frac{3}{5} (0)$$

$$= .400 \text{ bit}$$

$$I(B;I) = H(B) - H(B|I) = .971 - .400 = .571 \text{ bit}$$

Summary $I(B;A) = .171$ bit
 $I(B;I) = .571$ bit

I provides the greatest information gain
so use this as the next rule



branch for $I = \text{medium}$

2 yes 0 no

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

branch for $I = \text{low}$

0 yes 1 no

no further split is required

branch for $I = \text{high}$

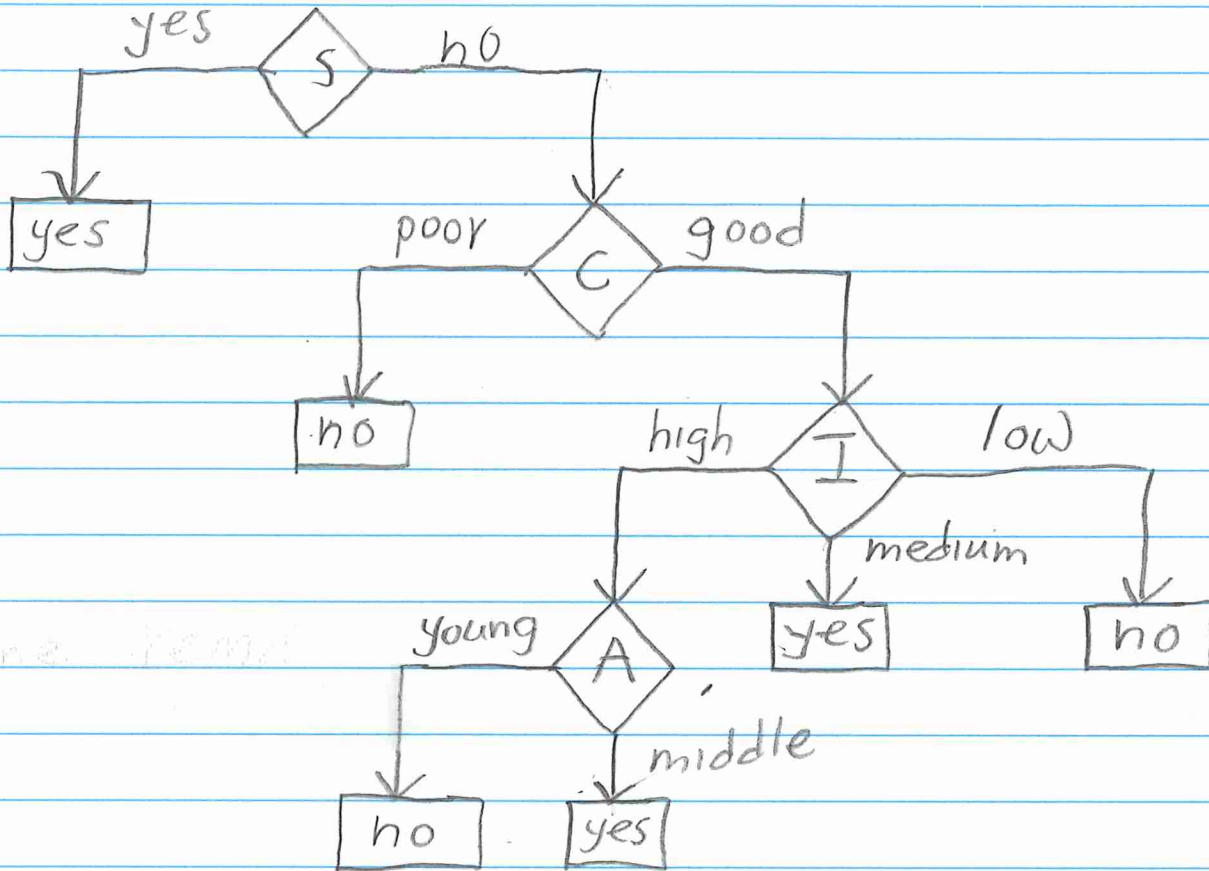
1 yes 1 no

one remaining attribute A so split on this

young
middle

0	yes	1	no
1	yes	0	no

the final tree



the result

Classification of the six test inputs

1	no
2	yes
3	no
4	yes
5	yes
6	yes

6. Let X be an RV with probabilities

$$p(x_i) \quad i=1, 2, \dots, N$$

without loss of generality, assume

$$p(x_1) > p(x_2)$$

Let Y be the RV with probabilities

$$p(y_1) = p(x_1) - \epsilon$$

$$p(y_2) = p(x_2) + \epsilon$$

$$p(y_i) = p(x_i) \quad i=3, \dots, N$$

where $0 < \epsilon < \frac{p(x_1) - p(x_2)}{2}$

we must show that $H(Y) > H(X)$

$$\begin{aligned} H(X) - H(Y) &= -p(x_1) \log p(x_1) - p(x_2) \log p(x_2) \\ &\quad + (p(x_1) - \epsilon) \log (p(x_1) - \epsilon) + (p(x_2) + \epsilon) \log (p(x_2) + \epsilon) \end{aligned}$$

$$H(X) - H(Y) = p(x_1) \log \left[\frac{p(x_1) - \epsilon}{p(x_1)} \right] + p(x_2) \log \left[\frac{p(x_2) + \epsilon}{p(x_2)} \right] \\ - \epsilon \log \left[\frac{p(x_1) - \epsilon}{p(x_2) + \epsilon} \right]$$

using the IT inequality

$$\log(x) \leq (x-1) \log e$$

$$H(X) - H(Y) \leq \log e [p(x_1) - \epsilon - p(x_1) + p(x_2) + \epsilon - p(x_2)] \\ - \epsilon \log \left[\frac{p(x_1) - \epsilon}{p(x_2) + \epsilon} \right] \\ \leq -\epsilon \log \left[\frac{p(x_1) - \epsilon}{p(x_2) + \epsilon} \right] < 0$$

$$\therefore H(X) - H(Y) < 0$$

$$\text{so } H(Y) > H(X)$$

7a) Two flips of the coin has outcomes

aa	ba	ca
ab	bb	cb
ac	bc	cc

with probabilities

$p(a)$	$p(b)p(a)$	$p(c)p(a)$
$p(a)p(b)$	$p^2(b)$	$p(c)p(b)$
$p(a)p(c)$	$p(b)p(c)$	$p^2(c)$

Since $p(a)p(b) = p(b)p(a)$

we can use

head	ab	tail	ba
	bc		cb
	ac		ca

which are equiprobable

the outcomes aa, bb, cc are ignored

b) The expected number of bits per flip is

$$\frac{1 - p(a)^2 - p(b)^2 - p(c)^2}{2}$$

$$= p(a)p(b) + p(b)p(c) + p(a)p(c)$$