CS3244 Tutorial 3

Zhuang Jianning
A0214561M
T13

TM3.1

a) $C = A \wedge \sim B$

A	B	$C = A \wedge \sim B$
T	T	F
T	F	T
F	T	F
F	F	F

$H(C|A) = \frac{2}{4} B(\frac{1}{2}) + \frac{2}{4} B(\frac{0}{2})$

$= \frac{2}{4} + 0 = \frac{1}{2}$

$Gain(C, A) = B(\frac{1}{4}) - H(C|A)$

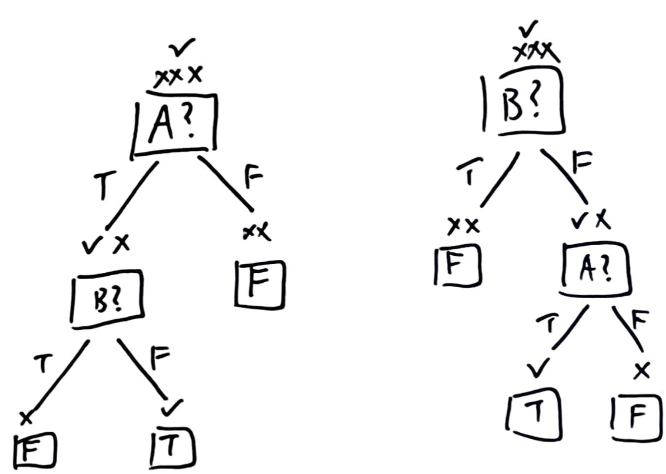
$= -(\frac{1}{4}\log\frac{1}{4} + \frac{3}{4}\log\frac{3}{4}) - \frac{1}{2}$

$= 0.31125$

$H(C|B) = \frac{2}{4} B(\frac{0}{2}) + \frac{2}{4} B(\frac{1}{2})$

$= \frac{1}{2}$

$Gain(C, B) = 0.31125$



same size

A and B have the same Gain $\Rightarrow$ choose either as root

b) $D = A \vee (B \wedge C)$

A	B	C	$A \vee (B \wedge C)$
T	T	T	T
T	T	F	T
T	F	T	T
T	F	F	T
F	T	T	T
F	T	F	F
F	F	T	F
F	F	F	F

$H(D|A) = \frac{4}{8} B(\frac{4}{4}) + \frac{4}{8} B(\frac{1}{4})$

$= 0 + \frac{4}{8} \cdot -(\frac{1}{4}\log\frac{1}{4} + \frac{3}{4}\log\frac{3}{4})$

$= 0.406$

$H(D|B) = \frac{4}{8} B(\frac{3}{4}) + \frac{4}{8} B(\frac{2}{4})$

$= 0.406 + \frac{4}{8}$

$= 0.906$

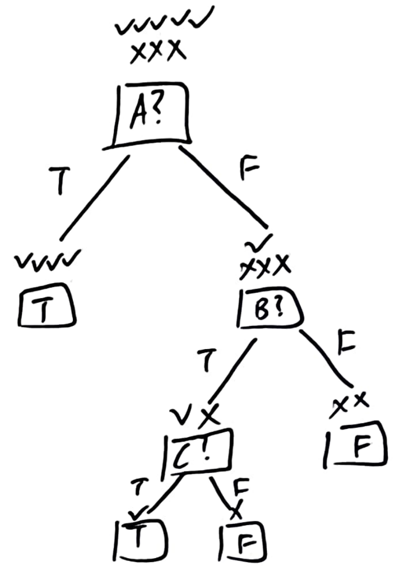
$H(D|C) = \frac{4}{8} B(\frac{3}{4}) + \frac{4}{8} B(\frac{2}{4})$

$= 0.906$

since $H(D|A)$ has lowest entropy, it will result in highest Gain

choose A as root

$$B\left(\tfrac{1}{4}\right) = -\left(\tfrac{1}{4}\log\tfrac{1}{4} + \tfrac{3}{4}\log\tfrac{3}{4}\right)$$
$$= 0.81125$$



$$H(D\mid B) = \tfrac{2}{4}B\left(\tfrac{1}{2}\right) + \tfrac{2}{4}B\left(\tfrac{0}{2}\right)$$
$$= \tfrac{2}{4} + 0 = \tfrac{1}{2}$$

$$\mathrm{Gain}(D,B) = B\left(\tfrac{1}{4}\right) - H(D\mid B)$$
$$= 0.31125$$

$$H(D\mid C) = \tfrac{2}{4}B\left(\tfrac{1}{2}\right) + \tfrac{2}{4}B\left(\tfrac{0}{2}\right)$$
$$= \tfrac{1}{2}$$

$$\mathrm{Gain}(D,C) = 0.31125$$

B and C have same Gain ⇒ choose either as root of subtree

c) $E = (A \wedge B) \vee (C \wedge D)$

$$H(E\mid A) = \tfrac{8}{16}B\left(\tfrac{5}{8}\right) + \tfrac{8}{16}B\left(\tfrac{3}{8}\right)$$
$$= 0.883$$

$$H(E\mid B) = \tfrac{8}{16}B\left(\tfrac{5}{8}\right) + \tfrac{8}{16}B\left(\tfrac{3}{8}\right)$$
$$= 0.883$$

$$H(E\mid C) = \tfrac{8}{16}B\left(\tfrac{5}{8}\right) + \tfrac{8}{16}B\left(\tfrac{3}{8}\right)$$
$$= 0.883$$

$$H(E\mid D) = \tfrac{8}{16}B\left(\tfrac{5}{8}\right) + \tfrac{8}{16}B\left(\tfrac{3}{8}\right)$$
$$= 0.883$$

A	B	C	D	$(A\wedge B)\vee(C\wedge D)$
T	T	T	T	T
T	T	T	F	T
T	T	F	T	T
T	T	F	F	T
T	F	T	T	T
T	F	T	F	F
T	F	F	T	F
T	F	F	F	F
F	T	T	T	T
F	T	T	F	F
F	T	F	T	F
F	T	F	F	F
F	F	T	T	T
F	F	T	F	F
F	F	F	T	F
F	F	F	F	F

All attributes have same Gain ⇒ symmetry!

WLOG, choose A to be root

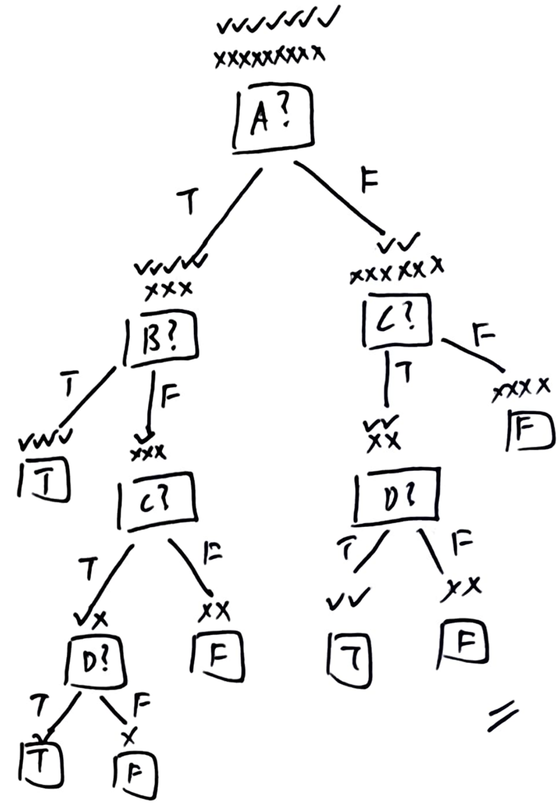
subtree for A=T

$$H(E|B) = \frac{4}{8}B\left(\frac{4}{4}\right) + \frac{4}{8}B\left(\frac{1}{4}\right)$$
$$= 0 + 0.406 = 0.406$$

$$H(E|C) = \frac{4}{8}B\left(\frac{3}{4}\right) + \frac{4}{8}B\left(\frac{2}{4}\right)$$
$$= 0.406 + \frac{1}{2} = 0.906$$

$$H(E|D) = \frac{4}{8}B\left(\frac{3}{4}\right) + \frac{4}{8}B\left(\frac{2}{4}\right) = 0.906$$

B has smallest entropy ⇒ highest Gain
⇒ next for subtree under A=T



subtree for A=F

$$H(E|B) = \frac{4}{8}B\left(\frac{1}{4}\right) + \frac{4}{8}B\left(\frac{1}{4}\right)$$
$$= 0.81125$$

$$H(E|C) = \frac{4}{8}B\left(\frac{2}{4}\right) + \frac{4}{8}B\left(\frac{0}{4}\right)$$
$$= 0.5$$

$$H(E|D) = \frac{4}{8}B\left(\frac{2}{4}\right) + \frac{4}{8}B\left(\frac{0}{4}\right)$$
$$= 0.5$$

wlog, choose C to be
next ~~Br~~ subtree under A=F

subtree for A=T, B=F

$$H(E|C) = \frac{2}{4}B\left(\frac{1}{2}\right) + \frac{2}{4}B\left(\frac{0}{2}\right)$$
$$= 0.5$$

$$H(E|D) = \frac{2}{4}B\left(\frac{1}{2}\right) + \frac{2}{4}B\left(\frac{0}{2}\right)$$

wlog, choose C
as next node

subtree for A=F, C=T

$$H(E|B) = \frac{2}{4}B\left(\frac{1}{2}\right) + \frac{2}{4}B\left(\frac{1}{2}\right)$$
$$= 1$$

$$H(E|D) = \frac{2}{4}B\left(\frac{2}{2}\right) + \frac{2}{4}B\left(\frac{0}{2}\right)$$
$$= 0$$

D as
next node

BL4

$$-\left(\frac{7}{12}\log\frac{7}{12} + \frac{5}{12}\log\frac{5}{12}\right) = 0.98$$

a) $\text{Gain}(\text{Decision}, \text{Income}) = B\left(\frac{7}{12}\right) - H(\text{Decision} \mid \text{Income})$

$$H(\text{Decision} \mid \text{Income}) = \frac{6}{12}B\left(\frac{3}{6}\right) + \frac{4}{12}B\left(\frac{3}{4}\right) + \frac{2}{12}B\left(\frac{1}{2}\right)$$

$$= \frac{6}{12} + \frac{2}{12} + \frac{4}{12} \cdot -\left(\frac{3}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{1}{4}\right)$$

$$= 0.937$$

$$\text{Gain}(\text{Decision}, \text{Income}) = 0.98 - 0.937$$
$$= 0.043$$

$$H(\text{Decision} \mid \text{Credit History}) = \frac{3}{12}B\left(\frac{0}{3}\right) + \frac{6}{12}B\left(\frac{4}{6}\right) + \frac{3}{12}B\left(\frac{3}{3}\right)$$

$$= 0 + 0 + \frac{6}{12} \cdot -\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right)$$

$$= 0.459$$

$$\text{Gain}(\text{Decision}, \text{CreditHistory}) = 0.98 - 0.459$$
$$= 0.521$$

$$H(\text{Decision} \mid \text{Debt}) = \frac{8}{12}B\left(\frac{5}{8}\right) + \frac{4}{12}B\left(\frac{2}{4}\right)$$

$$= \frac{4}{12} + \frac{8}{12} \cdot -\left(\frac{5}{8}\log\frac{5}{8} + \frac{3}{8}\log\frac{3}{8}\right)$$

$$= 0.970$$

$$\text{Gain}(\text{Decision}, \text{Debt}) = 0.98 - 0.970$$
$$= 0.01$$

CreditHistory results in highest Gain ⇒ chosen as root in Decision Tree

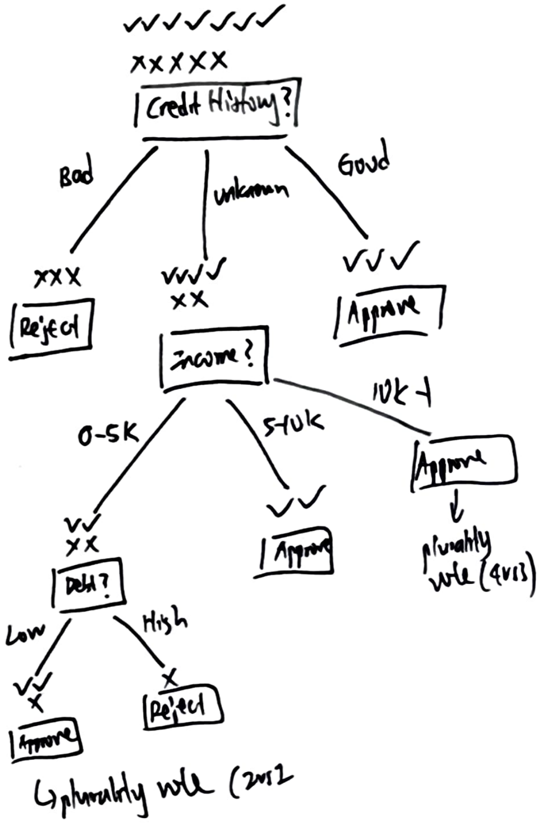
Bad → Reject

Good → Approve

unknown → requires subtree and further branching

(see DT in next page)

$$B\left(\frac{2}{6}\right) = -\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right)$$
$$= 0.918$$



$$H(\text{Decision} \mid \text{Income}) = \frac{4}{6}B\left(\frac{2}{4}\right) + \frac{2}{6}B\left(\frac{2}{2}\right)$$
$$= \frac{4}{6} + 0 = \frac{4}{6}$$

$$\text{Gain}(\text{Decision}, \text{Income}) = 0.918 - \frac{4}{6}$$
$$= 0.251$$

$$H(\text{Decision} \mid \text{Debt}) = \frac{4}{6}B\left(\frac{3}{4}\right) + \frac{2}{6}B\left(\frac{1}{2}\right)$$
$$= \frac{2}{6} + \frac{4}{6} \cdot -\left(\frac{3}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{1}{4}\right)$$
$$= 0.874$$

$$\text{Gain}(\text{Decision}, \text{Debt}) = 0.918 - 0.874$$
$$= 0.044$$

Income results in highest Gain ⇒ chosen as root of subtree under CreditHistory = Unknown

Income = 10K+ has no examples ⇒ plurality vote on parent examples (4✓ vs 3×) Approve ✓

Debt = low is not decisive but no more attributes left ⇒ plurality vote on current examples (2✓ vs 1×) Approve ✓

same attribute values but different Decision? ⇒ missing attributes?

b) 4K yearly income, good credit, high debt } immediately approve just based on good credit history

B2 5

a) $H(\text{Appealing} \mid \text{Taste}) = \frac{3}{10} B\left(\frac{0}{3}\right) + \frac{4}{10} B\left(\frac{2}{4}\right) + \frac{3}{10} B\left(\frac{3}{3}\right)$

$= 0 + \frac{4}{10} + 0$

$= 0.4$

$\text{Gain}(\text{Appealing}, \text{Taste}) = B\left(\frac{5}{10}\right) - H(\text{Appealing} \mid \text{Taste})$

$= 1 - 0.4$

$= 0.6$

b) DT with Taste? as root

Taste = salty ⇒ No

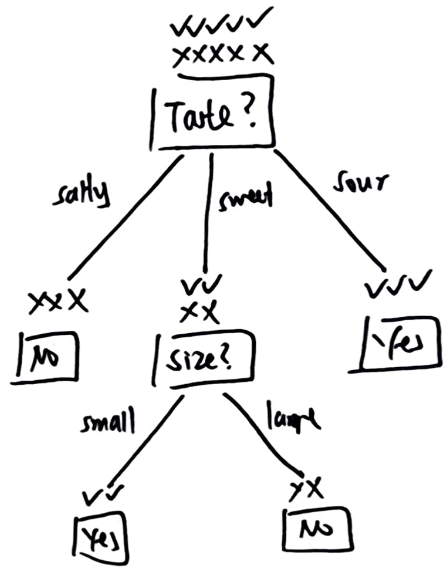
Taste = sour ⇒ Yes

Taste = sweet ⇒ 2✓ 2X

After splitting, all values for temperature is cold

Hence, no information gain can be made

Use size as next root for subtree



c) Temperature = Hot, Taste = salty, Size = small } immediately not appealing based on salty taste

Temperature = cold, Taste = sweet, Size = large } not appealing following path (taste = sweet ∧ size = large)