1. For the following dataset; Classify the test instances using the ID3 *Decision Tree* method:

ID	*Outl*	*Temp*	*Humi*	*Wind*	PLAY
TRAINING INSTANCES					
A	s	h	h	F	N
B	s	h	h	T	N
C	o	h	h	F	Y
D	r	m	h	F	Y
E	r	c	n	F	Y
F	r	c	n	T	N
TEST INSTANCES					
G	o	c	n	T	?
H	s	m	h	F	?

(i). Using the **Information Gain** as a splitting criterion

(ii). Using the **Gain Ratio** as a splitting criterion

(i) IG: At each level of DT, choose attribute with largest IG

weight : fraction of instance at child node i

$$IG(A|R) = \underbrace{H(R)}_{\text{entropy of parent node}} - \underbrace{\sum_{i \in A} P(A=i)\,H(A=i)}_{\text{weighted average entropy across child nodes}}$$

(Mean Information : MI)

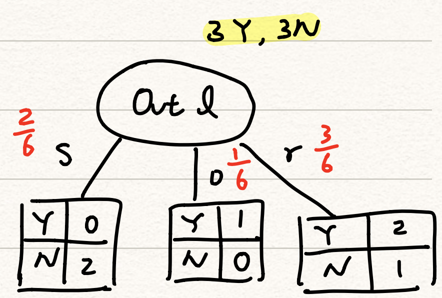
H: entropy (impurity) of a node

$$H(X) = -\sum_i p(x_i)\log_2 p(x_i)$$

Root node: 3 Y, 3 N

$$H(R) = -\Big\{\underbrace{\tfrac{1}{2}\log_2\tfrac{1}{2}}_{Y} + \underbrace{\tfrac{1}{2}\log_2\tfrac{1}{2}}_{N}\Big\} = 1$$

Example: For Outl:



$$H(Outl=s) = -\{0\log_2 0 + 1\log_2 1\} = 0$$

$$H(Outl=o) = -\{1\log_2 1 + 0\log_2 0\} = 0$$

$$H(Outl=r) = -\{\tfrac{2}{3}\log_2\tfrac{2}{3} + \tfrac{1}{3}\log_2\tfrac{1}{3}\} = 0.9183$$

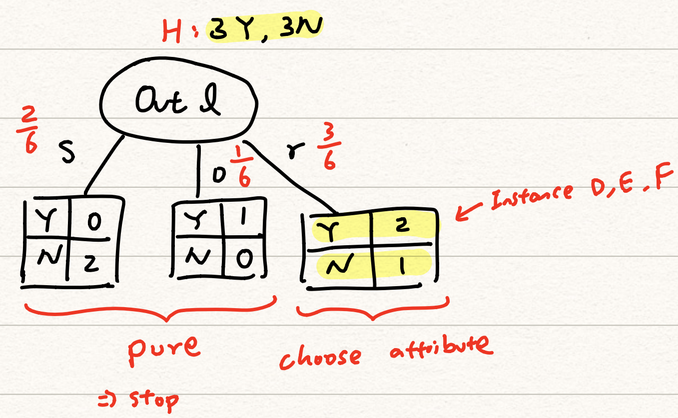
$$MI(Outl) = \tfrac{2}{6}\times 0 + \tfrac{1}{6}\times 0 + \tfrac{3}{6}\times 0.9183 = 0.4592$$

$IG(Outl) = H(R) - MI(Outl) = 1 - 0.4592 = 0.5408$

will get useless classifier

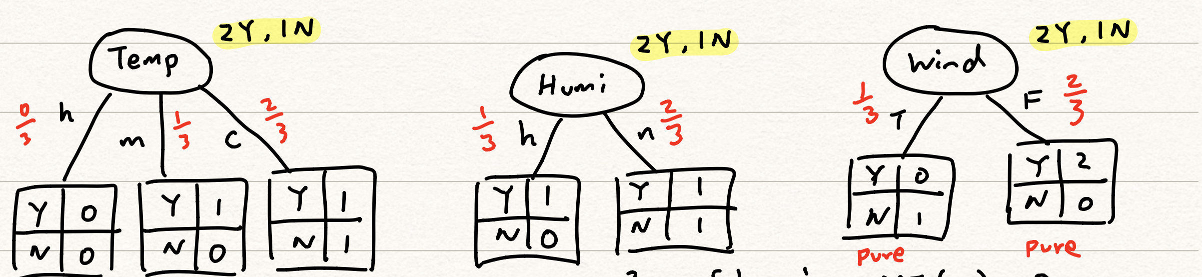
	R	Outl			Temp			H		Wind		~~ID~~					
		s	o	r	h	m	c	h	n	T	F	A	B	C	D	E	F
Y	3	0	1	2	1	1	1	2	1	0	3	0	0	1	1	1	0
N	3	2	0	1	2	0	1	2	1	2	1	1	1	0	0	0	1
Total	6	2	1	3	3	1	2	4	2	2	4	1	1	1	1	1	1
P(Y)	1/2	0	1	2/3	1/3	1	1/2	1/2	1/2	0	3/4	0	0	1	1	1	0
P(N)	1/2	1	0	1/3	2/3	0	1/2	1/2	1/2	1	1/4	1	1	0	0	0	1
H	1	0	0	0.9183	0.9183	0	1	1	1	0	0.8112	0	0	0	0	0	0
MI		0.4592			0.7924			1		0.5408		0					
IG		0.5408			0.2076			0		0.4592		1					
SI		1.459			1.459			0.9183		0.9183		2.585					
GR		0.3707			0.1423			0		0.5001		0.3868					

largest



parent : $H(Outl = r) = 0.9183$

=> Find IG for Temp, Humi, Wind



$MI(T) = \frac{2}{3} \times -\left\{\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right\} = \frac{2}{3}$

$IG(T) = 0.9183 - \frac{2}{3} = 0.2516$

$MI(w) = \frac{2}{3} \times -\left\{\frac{1}{2}\log_2\frac{1}{2} + \frac{1}{2}\log_2\frac{1}{2}\right\} = 0.2516$

$IG(w) = 0.2516$

$MI(w) = 0$

(largest)

$IG(w) = 0.9183 - 0 = 0.9183$

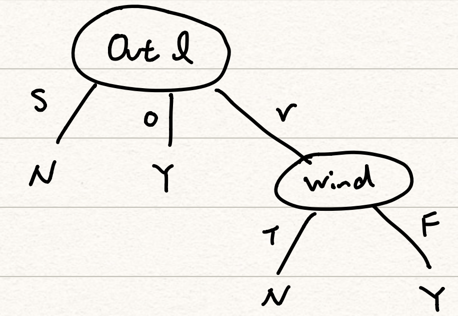
1. For the following dataset; Classify the test instances using the ID3 *Decision Tree* method:

ID	*Outl*	*Temp*	*Humi*	*Wind*	PLAY
TRAINING INSTANCES					
A	s	h	h	F	N
B	s	h	h	T	N
C	o	h	h	F	Y
D	r	m	h	F	Y
E	r	c	n	F	Y
F	r	c	n	T	N
TEST INSTANCES					
G	o	c	n	T	?
H	s	m	h	F	?

(i). Using the **Information Gain** as a splitting criterion

(ii). Using the **Gain Ratio** as a splitting criterion

DT :



Classify G: Outl = o $\Rightarrow$ Y

Classify H: Outl = s $\Rightarrow$ N

(ii) Choose attribute with largest GR.

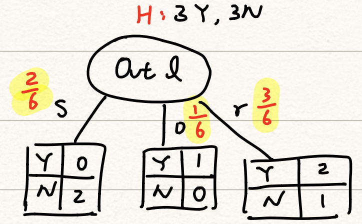
$$GR(A) = \frac{IG(A)}{SI(A)}$$ split info (entropy)

H : 3Y, 3N

$$SI(Outl) = -\left\{\frac{2}{6}\log_2\frac{2}{6} + \frac{1}{6}\log_2\frac{1}{6} + \frac{3}{6}\log_2\frac{3}{6}\right\}$$

$= 1.459$

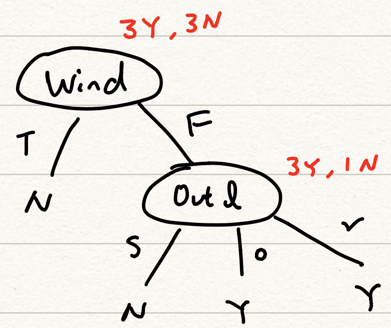
$$\Rightarrow GR(Outl) = \frac{0.5408}{1.459} = 0.3707$$



SI : dist of instances across child nodes

	R	Outl			Temp			H		Wind		~~ID~~					
		s	o	r	h	m	c	h	n	T	F	A	B	C	D	E	F
Y	3	0	1	2	1	1	1	2	1	0	3	0	0	1	1	1	0
N	3	2	0	1	2	0	1	2	1	2	1	1	1	0	0	0	1
Total	6	2	1	3	3	1	2	4	2	2	4	1	1	1	1	1	1
P(Y)	1/2	0	1	2/3	1/3	1	1/2	1/2	1/2	0	3/4	0	0	1	1	1	0
P(N)	1/2	1	0	1/3	2/3	0	1/2	1/2	1/2	1	1/4	1	1	0	0	0	1
H	1	0	0	0.9183	0.9183	0	1	1	1	0	0.8112	0	0	0	0	0	0
MI		0.4592			0.7924			1		0.5408		0					
IG		0.5408			0.2076			0		0.4592		1					
SI		1.459			1.459			0.9183		0.9183		2.585					
GR		0.3707			0.1423			0		0.5001		0.3868					

Skip to DT:



Test Instances					
G	o	c	n	T	?
H	s	m	h	F	?

Classify G: Wind = T ⇒ N

Classify H: Wind = F → Outl = S ⇒ N

2. Given the same dataset, we wished to perform feature selection on this dataset, where the class is PLAY:

ID	*Outl*	*Temp*	*Humi*	*Wind*	PLAY
A	s	h	h	F	N
B	s	h	h	T	N
C	o	h	h	F	Y
D	r	m	h	F	Y
E	r	c	n	F	Y
F	r	c	n	T	N

(i). Which of Humi and Wind has the greatest *Pointwise Mutual Information* for the class Y? What about N?

PMI :

$$PMI(A,C) = \log_2 \underbrace{\frac{P(A,C)}{P(A)\,P(C)}}$$

$\Rightarrow$ find A with largest PMI

$\sim 1 \Rightarrow$ Independent

$<< 1 \Rightarrow$ Negatively correlated

$>> 1 \Rightarrow$ occur together (positively correlated, if one occur, the other likely to occur)

For Humi (A) & Y (C) : (let Humi : h:T)

$$PMI(Humi{=}h, Play{=}Y) = \log_2 \frac{\frac{2}{6}}{\frac{4}{6} \cdot \frac{3}{6}} = \log_2 1 = 0 \quad \text{(uncorrelated)}$$

For Wind (A) & Y (C) :

$$PMI(Wind{=}T, Play{=}Y) = \log_2 \frac{\frac{0}{6}}{\frac{2}{6}\ \frac{3}{6}} = \log_2 0 = -\infty \quad \text{(neg correlated)}$$

$\Rightarrow$ Humi has better PMI for class Y.

Neg class (N) :

Humi : still uncorrelated

Wind : pos correlated

(per-class)

(ii). Which of the attributes has the greatest *Mutual Information* for the class, as a whole? (Note that we need to extend the lecture definition to handle non–binary attributes.)

$$MI(A,C) = \underbrace{\overbrace{\sum_{i \in \{a,\bar{a}\}} \sum_{j \in \{c,\bar{c}\}}}^{4 \text{ terms}} P(i,j)}_{\text{weighted-avg}} \underbrace{\log_2 \frac{P(i,j)}{P(i)\,P(j)}}_{\text{PMI}}$$

$$\Rightarrow \quad MI(A,C) = \underbrace{\sum_{i \in A}}_{\text{all values}} \sum_{j \in \{c,\bar{c}\}} P(i,j) \log_2 \frac{P(i,j)}{P(i)\,P(j)}$$

MI : Combine PMIs of the attribute occurring together with each class, as well as not occurring.

For Outl (a) & Play (c) :

MI(Outl)

$$MI(\text{Outl}, \text{Play}) = P(s,Y)\log_2\frac{P(s,Y)}{P(s)P(Y)} + P(o,Y)\log_2\frac{P(o,Y)}{P(o)P(Y)}$$
$$+ P(r,Y)\log_2\frac{P(r,Y)}{P(r)P(Y)} + P(s,N)\log_2\frac{P(s,N)}{P(s)P(N)}$$
$$+ P(o,N)\log_2\frac{P(o,N)}{P(o)P(N)} + P(r,N)\log_2\frac{P(r,N)}{P(r)P(N)}$$
$$= \frac{0}{6}\log_2\frac{\frac{0}{6}}{\frac{3}{6}\cdot\frac{3}{6}} + \frac{1}{6}\log_2\frac{\frac{1}{6}}{\frac{1}{6}\cdot\frac{3}{6}} + \cdots$$
$$= 0.541$$

For Temp (a) & Play (c) :

MI(Outl)

$$MI(\text{Outl}, \text{Play}) = P(h,Y)\log_2\frac{P(h,Y)}{P(h)P(Y)} + P(h,Y)\log_2\frac{P(h,Y)}{P(h)P(Y)}$$
$$+ P(m,Y)\log_2\frac{P(m,Y)}{P(m)P(Y)} + P(m,N)\log_2\frac{P(m,N)}{P(m)P(N)}$$
$$+ P(c,N)\log_2\frac{P(c,N)}{P(c)P(N)} + P(c,N)\log_2\frac{P(c,N)}{P(c)P(N)}$$
$$= 0.110$$

For Humi (a) & Play (c) : MI = 0
For Wind (a) & Play (c) : MI = 0.459

exercise

⇒ Outl is the best attribute (MI)

PMI : attribute value & a class value (c or $\bar{c}$)
MI : an attribute (consider all values) & a class (c & $\bar{c}$)