1. A **confusion matrix** is a summary of the performance of a (supervised) classifier over a set of development ("test") data, by counting the various instances:

		Actual			
		a	*b*	*c*	*d*
Classified	*a*	10	2	3	1
	b	2	5	3	1
	c	1	3	7	1
	d	3	0	3	5

(i). Calculate the classification **accuracy** of the system. Find the **error rate** for the system.

$$\text{Acc} = \frac{\#\text{ of correctly classified}}{\text{total }\#\text{ of instances}}$$

$$= \frac{10+5+7+5}{\text{sum of the table} = 50}$$

$$= \frac{27}{50}$$

$$= 54\%$$

$$ER = 1 - \text{Acc} = 46\%$$

(ii). Calculate the **precision**, **recall** and **F-score** (where $\beta = 1$), for class *d*. (Why can't we do this for the whole system? How can we consider the whole system?)

$$\text{Precision} = \frac{\#\text{ of instances "correctly" classified as } d}{\#\text{ of instances classified as } d} = \frac{TP}{TP+FP} = \frac{5}{5+6} = \frac{5}{11}$$

$$\text{Recall} = \frac{\#\text{ of instances "correctly" classified as } d}{\#\text{ of instances truly } d} = \frac{TP}{TP+FN} = \frac{5}{5+3} = \frac{5}{8}$$

		Actual			
		a	*b*	*c*	*d*
Classified	*a*	10	2	3	1
	b	2	5	3	1
	c	1	3	7	1
	d	3	0	3	5

FN=3

FP=6

TP=5

$$\beta = 1 : F_1 = \frac{2P \cdot R}{P+R} = \frac{2}{\frac{1}{P}+\frac{1}{R}} \quad \text{(harmonic mean)}$$

$$= \frac{2 \cdot \frac{5}{11} \cdot \frac{5}{8}}{\frac{5}{11} + \frac{5}{8}}$$

$$= 53\%$$

2. For the following dataset:

ID	*Outl*	*Temp*	*Humi*	*Wind*	PLAY
TRAINING INSTANCES					
A	s	h	n	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	m	n	T	?
H	?	h	?	F	?

(i). Classify the test instances using the method of **0-R**.

0-R : Majority class classifier

train : 3Y , 3N (choose either!)

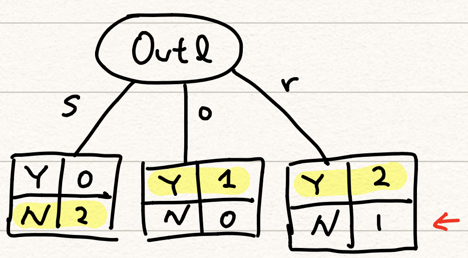
=> Let's choose "N"

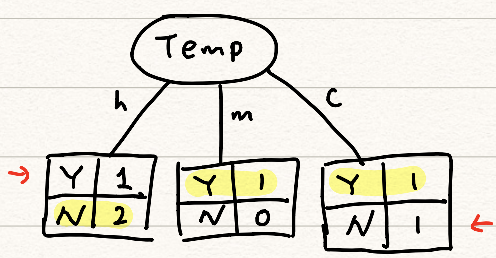
=> Instances : G → N

H → N

(ii). *[OPTIONAL]* Classify the test instances using the method of **1-R**. (Assume Outl = s for H)

1-R: Classify instances based on one attribute





ER = $\frac{1}{6}$ ER = $\frac{2}{6}$

Choose attribute with lowest ER : Assume "Outl" is best

=> Instances : G : Outl = o → Y

H : Outl = s → N

ID	Outl	Temp	Humi	Wind	PLAY
		TRAINING INSTANCES			
A	s	h	n	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	m	n	T	?
H	?	h	?	F	?

3. Given the above dataset, build a Naïve Bayes model for the given training instances. (6 training instances)

NB : probabilistic classification model

⇒ Calculate prior & conditional probs

Prior probs :

$P(Y) = \frac{1}{2}$ $P(N) = \frac{1}{2}$

Conditional :

$P(Outl = s \mid Y) = 0$ $P(Outl = s \mid N) = \frac{2}{3}$

$P(Outl = o \mid Y) = \frac{1}{3}$ $P(Outl = o \mid N) = 0$

$P(Outl = r \mid Y) = \frac{2}{3}$ $P(Outl = r \mid N) = \frac{1}{3}$

$P(Temp = h \mid N) = \frac{2}{3}$ $P(Temp = m \mid N) = 0$ $P(Temp = c \mid N) = \frac{1}{3}$

$P(Temp = h \mid Y) = \frac{1}{3}$ $P(Temp = m \mid Y) = \frac{1}{3}$ $P(Temp = c \mid Y) = \frac{1}{3}$

$P(Humi = n \mid N) = \frac{2}{3}$ $P(Humi = h \mid N) = \frac{1}{3}$

$P(Humi = n \mid Y) = \frac{1}{3}$ $P(Humi = h \mid Y) = \frac{2}{3}$

$P(Wind = T \mid N) = \frac{2}{3}$ $P(Wind = F \mid N) = \frac{1}{3}$

$P(Wind = T \mid Y) = 0$ $P(Wind = F \mid Y) = 1$

4. Using the Naïve Bayes model that you developed in question 4, classify the given test instances.

(i). No smoothing.

$$\hat{c} = \underset{c}{\operatorname{argmax}}\ P(c \mid \underbrace{t_1, \dots, t_K}_{\text{attributes}})$$

$$= \underset{c}{\operatorname{argmax}}\ \frac{P(t_1, \dots, t_K \mid c)\, P(c)}{P(t_1, \dots, t_K)}$$

same for all classes (ignore in "argmax")

$$= \underset{c}{\operatorname{argmax}}\ P(t_1, \dots, t_K \mid c)\, P(c)$$

$$= \underset{c}{\operatorname{argmax}}\ \underbrace{\left\{ \prod_{i=1}^{K} P(t_i \mid c) \right\}}_{\text{NB assumption: attributes are conditionally independent}} \cdot P(c)$$

NB assumption: attributes are conditionally independent

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

$P(Outl = s \mid N) = \frac{2}{3}$	$P(Outl = o \mid N) = 0$	$P(Outl = r \mid N) = \frac{1}{3}$
$P(Outl = s \mid Y) = 0$	$P(Outl = o \mid Y) = \frac{1}{3}$	$P(Outl = r \mid Y) = \frac{2}{3}$
$P(Temp = h \mid N) = \frac{2}{3}$	$P(Temp = m \mid N) = 0$	$P(Temp = c \mid N) = \frac{1}{3}$
$P(Temp = h \mid Y) = \frac{1}{3}$	$P(Temp = m \mid Y) = \frac{1}{3}$	$P(Temp = c \mid Y) = \frac{1}{3}$
$P(Humi = n \mid N) = \frac{2}{3}$	$P(Humi = h \mid N) = \frac{1}{3}$	$P(Y) = \frac{1}{2}$
$P(Humi = n \mid Y) = \frac{1}{3}$	$P(Humi = h \mid Y) = \frac{2}{3}$	$P(N) = \frac{1}{2}$
$P(Wind = T \mid N) = \frac{2}{3}$	$P(Wind = F \mid N) = \frac{1}{3}$	
$P(Wind = T \mid Y) = 0$	$P(Wind = F \mid Y) = 1$	

Instance G:

- Class Y:

$P(Outl = o \mid Y)\, P(Temp = m \mid Y)\, P(Humi = n \mid Y)$

$\times P(Wind = T \mid Y)\, P(Y)$

$= \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times 0 \times \frac{1}{2} = 0$

- Class N:

$P(\cdot\cdot) \dots P(N)$

$= 0 \times 0 \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{2}$

$= 0$

$\Rightarrow$ Choose either label for G

(e.g. Y)

Instance H:

missing values $\Rightarrow$ use what we have.

- Class Y:

$P(Temp = h \mid Y)\, P(wind = F \mid Y)\, P(Y)$

$= \frac{1}{3} \times 1 \times \frac{1}{2} = \frac{1}{6}$ ✓

- Class N:

$P(Temp = h \mid N)\, P(wind = F \mid N)\, P(N)$

$= \frac{2}{3} \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{9}$

$\Rightarrow$ Classify instance H as Y.

(ii). Using the "epsilon" (ε) smoothing method.

Replace zero probabilities with ε (a small value, e.g. 10^{-6})

For instance G:

Y: $P(Outl=o|Y)\,P(Temp=m|Y)\,P(Humi=n|Y)$

$\times P(Wind=T|Y)\,P(Y)$

$= \frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} \times \varepsilon \times \frac{1}{2} = \frac{\varepsilon}{54}$

N: $P(\cdot\cdot) \cdots P(N)$

$= \varepsilon \times \varepsilon \times \frac{2}{3} \times \frac{2}{3} \times \frac{1}{2}$

$= \frac{2\varepsilon^2}{9}$

$\frac{\varepsilon}{54} > \frac{2\varepsilon^2}{9}$ $\Rightarrow$ Classify G as Y.

For instance H: no zero probabilities $\Rightarrow$ still Y.

add 1 smoothing

(iii). *[OPTIONAL]* Using "Laplace" smoothing ($\alpha = 1$)

For conditional probs:

$$p_i = \frac{x_i + \alpha}{N + \alpha d}$$ (was $p_i = \frac{x_i}{N}$)

possible values for this attribute

$\Rightarrow$ ensure $\Sigma_k P(A_i = k \mid \text{Class } c) = 1$

$\Rightarrow$ No zero conditional probs! (If $x_i = 0$, $p_i = \frac{\alpha}{N+\alpha d}$)

Instance G: s, o, r   h, m, c   n, h

Y: $P(Outl=o|Y)\,P(Temp=m|Y)\,P(Humi=n)$

$\times P(Wind=T)\,P(Y)$ T, F

$= \frac{1+1}{3+3} \times \frac{1+1}{3+3} \times \frac{1+1}{3+2} \times \frac{0+1}{3+2} \times \frac{1}{2}$

$= \frac{2}{6} \times \frac{2}{6} \times \frac{2}{5} \times \frac{1}{5} \times \frac{1}{2}$

$= 0.0044$

N : $P(\cdot\cdot) \dots P(N)$

$= \frac{0+1}{3+3} \times \frac{0+1}{3+3} \times \frac{2+1}{3+2} \times \frac{2+1}{3+2} \times \frac{1}{2}$

$= \frac{1}{6} \times \frac{1}{6} \times \frac{3}{5} \times \frac{3}{5} \times \frac{1}{2}$

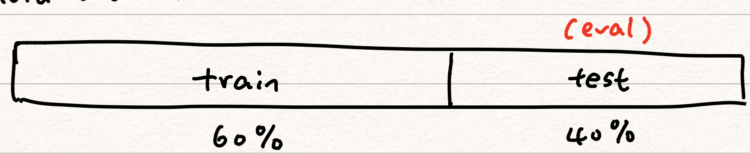
$= 0.005$ ✓

$\Rightarrow$ Classify instance G as N.

Similarly for H ...

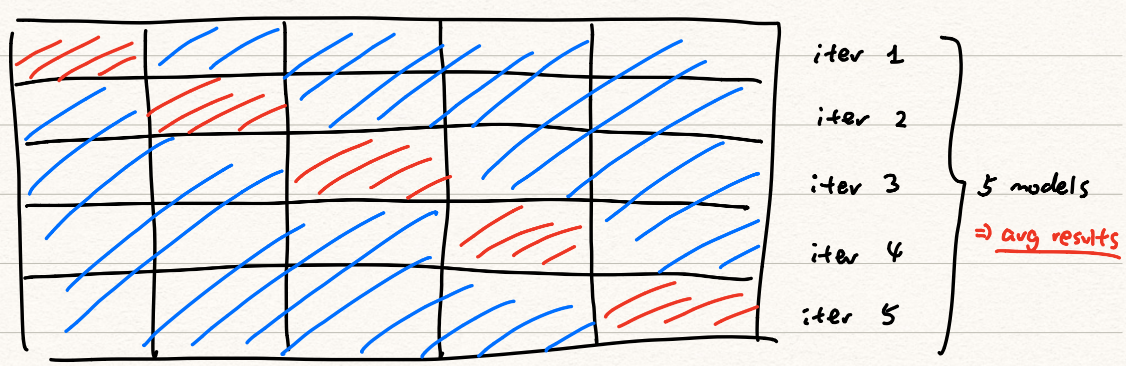
5. *[OPTIONAL]* How is **holdout** evaluation different to **cross-validation** evaluation? What are some reasons we would prefer one strategy over the other?

Hold-out : fixed



subject to some random variation in allocation of instances (to train & test)

Cross-validation : K-fold (K: 5 partitions)



for test

for trai

Each instance is used for testing.

(But CV takes longer)