1) Consider the following data table:

F_1	F_2	F_3	O
a	a	a	$+$
c	b	c	$+$
c	a	c	$+$
b	a	a	$-$
a	b	c	$-$
b	b	c	$-$

a) Determine the whole decision tree using ID3 (**information gain**), taking "O" as the target. Show all steps.

$$E\left(\frac{3}{3+3}, \frac{3}{3+3}\right) = E\left(\frac{1}{2}, \frac{1}{2}\right)$$

$$= -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = 1$$

Solution:

Before anything, we need to compute the entropy we start with:

$$E_{start} = E\left(\frac{3}{3+3}, \frac{3}{3+3}\right) = 1bit$$

Let us test attribute F_1:

$$\begin{array}{ccc}
F_1 = a & F_1 = b & F_1 = c \\
\downarrow & \downarrow & \downarrow \\
\#\{O=+\} = 1 & \#\{O=+\} = 0 & \#\{O=+\} = 2 \\
\#\{O=-\} = 1 & \#\{O=-\} = 2 & \#\{O=-\} = 0 \\
\downarrow & \downarrow & \downarrow \\
E\left(\frac{1}{1+1}, \frac{1}{1+1}\right) & E\left(\frac{2}{2+0}\right) & E\left(\frac{2}{2+0}\right)
\end{array}$$

$E(1) = 0$

$$E_{F_1} = \frac{1}{3}(1) + \frac{1}{3}(0) + \frac{1}{3}(0) = \frac{1}{3}$$

Total entropy after partitioning by F_1 is equal to the weighted average of each partition's entropy: $E_{F_1} = \frac{2}{6}E\left(\frac{1}{1+1}, \frac{1}{1+1}\right) + \frac{2}{6}E\left(\frac{2}{2+0}\right) + \frac{2}{6}E\left(\frac{2}{2+0}\right) \approx 0.33bit.$

So, the information gain is given by subtracting the entropy at begining from the remaining entropy after F_1: $G(F_1) = E_{start} - E_{F_1} \approx 0.66bit.$

Repeating for F2 and F3:

Let us test the next attribute, namely F_2:

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

$$\begin{array}{cc}
F_2 = a & F_2 = b \\
\downarrow & \downarrow \\
\#\{O=+\} = 2 & \#\{O=+\} = 1 \\
\#\{O=-\} = 1 & \#\{O=-\} = 2 \\
\downarrow & \downarrow \\
E\left(\frac{2}{2+1}, \frac{1}{2+1}\right) & E\left(\frac{1}{1+2}, \frac{2}{1+2}\right)
\end{array}$$

Total entropy after partitioning by F_2 is equal to the weighted average of each partition's entropy: $E_{F_2} = \frac{3}{6}E\left(\frac{2}{2+1}, \frac{1}{2+1}\right) + \frac{3}{6}E\left(\frac{1}{1+2}, \frac{2}{1+2}\right) = 0.9183bit.$

So, the information gain is given by subtracting the entropy at begining from the remaining entropy after F_2: $G(F_2) = E_{start} - E_{F_2} = 0.0817bit$

Total entropy after partitioning by F_3 is equal to the weighted average of each partition's entropy: $E_{F_3} = \frac{2}{6}E\left(\frac{1}{1+1}, \frac{1}{1+1}\right) + \frac{4}{6}E\left(\frac{2}{2+2}, \frac{2}{2+2}\right) = 1.0bit.$

So, the information gain is given by subtracting the entropy at begining from the remaining entropy after F_3: $G(F_3) = E_{start} - E_{F_3} = 0bit$

Since F_1 provides the highest gain, we use it as the first node in the tree and

Finishing the first step, we can notice that when choosing F1=a or F2=b, there is only one option for the final result in either case, meaning there is no uncertainty.

Since F_1 provides the highest gain, we use it as the first node in the tree and get this break down of the dataset:

Always ends in "-"

Always ends in "+"

$F_1 = a$			$F_1 = b$			$F_1 = c$		
F_2	F_3	O	F_2	F_3	O	F_2	F_3	O
a	a	$+$	a	a	$-$	b	c	$+$
b	c	$-$	b	c	$-$	a	c	$+$
				Done!			*Done!*	

One of two options: $E\left(\frac{1}{2}, \frac{1}{2}\right) = 1$

The first partition ($F_1 = a$) still has uncertainty. For that reason, we will repeat the same process to decide the next attribute to test. Before anything, we need to compute the entropy we start with:

$$E_{start} = E\left(\frac{1}{1+1}, \frac{1}{1+1}\right) = 1bit$$

Computing the remaining two options when F1=a

Total entropy after partitioning by F_2 is equal to the weighted average of each partition's entropy: $E_{F_2} = \frac{1}{2}E\left(\frac{1}{1+0}\right) + \frac{1}{2}E\left(\frac{1}{0+1}\right) = 0bit$.

So, the information gain is given by subtracting the entropy at begining from the remaining entropy after F_2: $G(F_2) = E_{start} - E_{F_2} = 1bit$

Total entropy after partitioning by F_3 is equal to the weighted average of each partition's entropy: $E_{F_3} = \frac{1}{2}E\left(\frac{1}{1+0}\right) + \frac{1}{2}E\left(\frac{1}{0+1}\right) = 0.0bit$.

So, the information gain is given by subtracting the entropy at begining from the remaining entropy after F_3: $G(F_3) = E_{start} - E_{F_3} = 1bit$

Separating this last uncertainty in two more nodes, we reach the final result, with no uncertainties, which yields the final tree.

Since both attributes have the same gain. we can choose either. Let us go with F_2 and get the following partioning:

$F_1 = a$

$F_2 = a$			$F_2 = b$	
F_3	O		F_3	O
a	$+$		c	$-$
Done!			*Done!*	

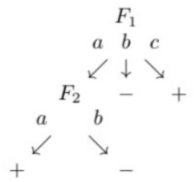
$F_1 = b$

F_2	F_3	O
a	a	$-$
b	c	$-$
Done!		

$F_1 = c$

F_2	F_3	O
b	c	$+$
a	c	$+$
Done!		

Since there is no more uncertainty, we needn't explore any further. Thus, we reach the final tree:



2) Consider the following data table:

F_1	F_2	F_3	O
d	a	b	m
c	a	b	n
c	a	a	y
d	a	a	y
c	b	a	f
c	b	b	f

a) Compute the first attribute to be tested using ID3.

Starting $E = E\left(\frac{1}{6}, \frac{1}{6}, \frac{1}{3}, \frac{1}{3}\right)$
$= 1.918$

F_1

$F_1 = d: \ E\left(\frac{1}{2}, \frac{1}{2}\right) = 1$

$F_1 = c: \ E\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right) = 1{,}5$

$E(F_1) = \frac{1}{3}(1) + \frac{2}{3}(1{,}5) = \frac{4}{3}$

$IG(F_1) = 1.918 - \frac{4}{3} = \boxed{0.585}$

F_2

$F_2 = a: \ E\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right) = 1{,}5$

$F_2 = b: \ E(1) = 0$

$E(F_2) = \frac{2}{3}(1{,}5) + \frac{1}{3}(0) = 1$

$IG(F_2) = \boxed{0.918}$

F_3

$F_3 = a$: $E\left(\frac{2}{3}, \frac{1}{3}\right) = 0.918$

$F_3 = b$: $E\left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) = 1.585$

$$E(F_3) = \frac{1}{2}(0.918) + \frac{1}{2}(1.585)$$
$$= 1.252$$
$$IG(F_3) = 1.918 - 1.252$$
$$= \boxed{0.666}$$

F2 yields the best information gain and is, therefore, the root of the tree.

b) Complete the tree started in the previous question. There is no need to perform all computations. What do you need to take into account?

$F_2 = a$

F_1	F_3	O
d	b	m
c	b	m
c	a	y
d	a	y

$F_2 = b$

F_1	F_3	O
c	a	f
c	b	f

Done! ✓

Starting $E = E\left(\frac{1}{2}, \frac{1}{2}\right) = 1$

- Testing F1 would lead to uncertainty regardless.
- Testing F3 allows us to be certain of one branch:

$F_2 = a$

$F_3 = a$

F_1	O
c	y
d	y

Done! ✓

$F_3 = b$

F_1	O
d	m
c	m

$F_2 = b$

F_1	F_3	O
c	a	f
c	b	f

Done! ✓