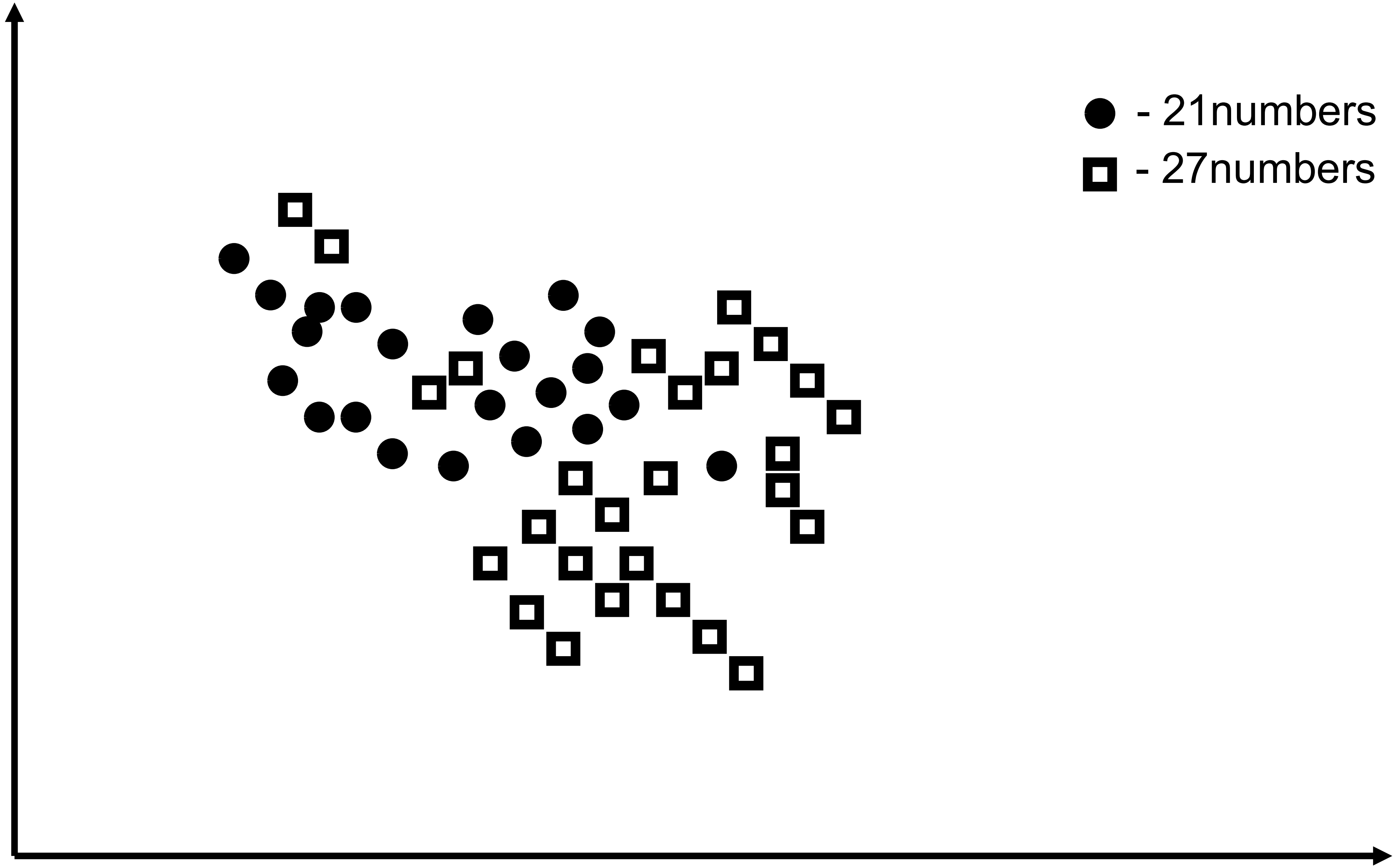
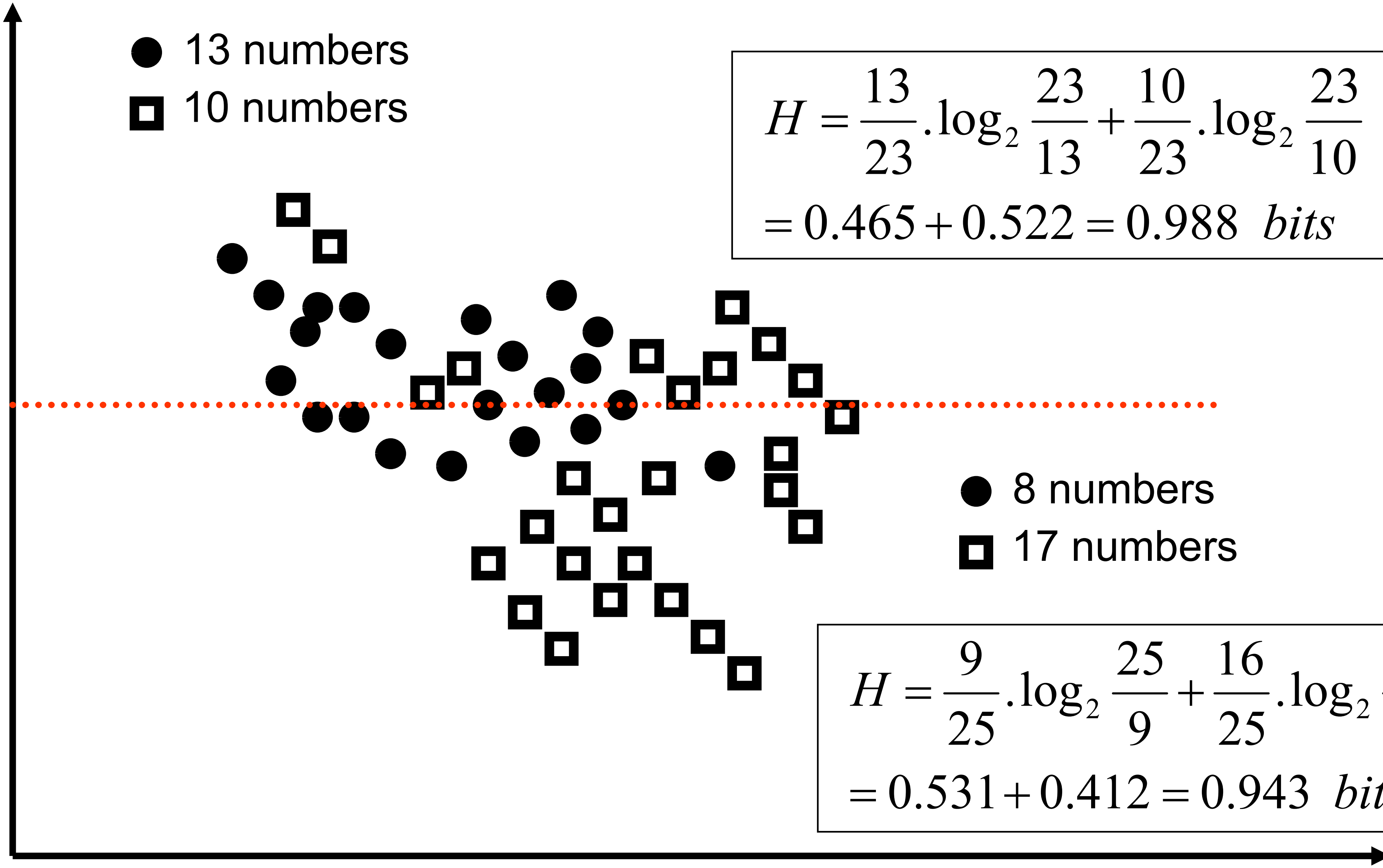
Decision Tree

What is entropy?

- Measure of disorder
- Use it to classify the objects

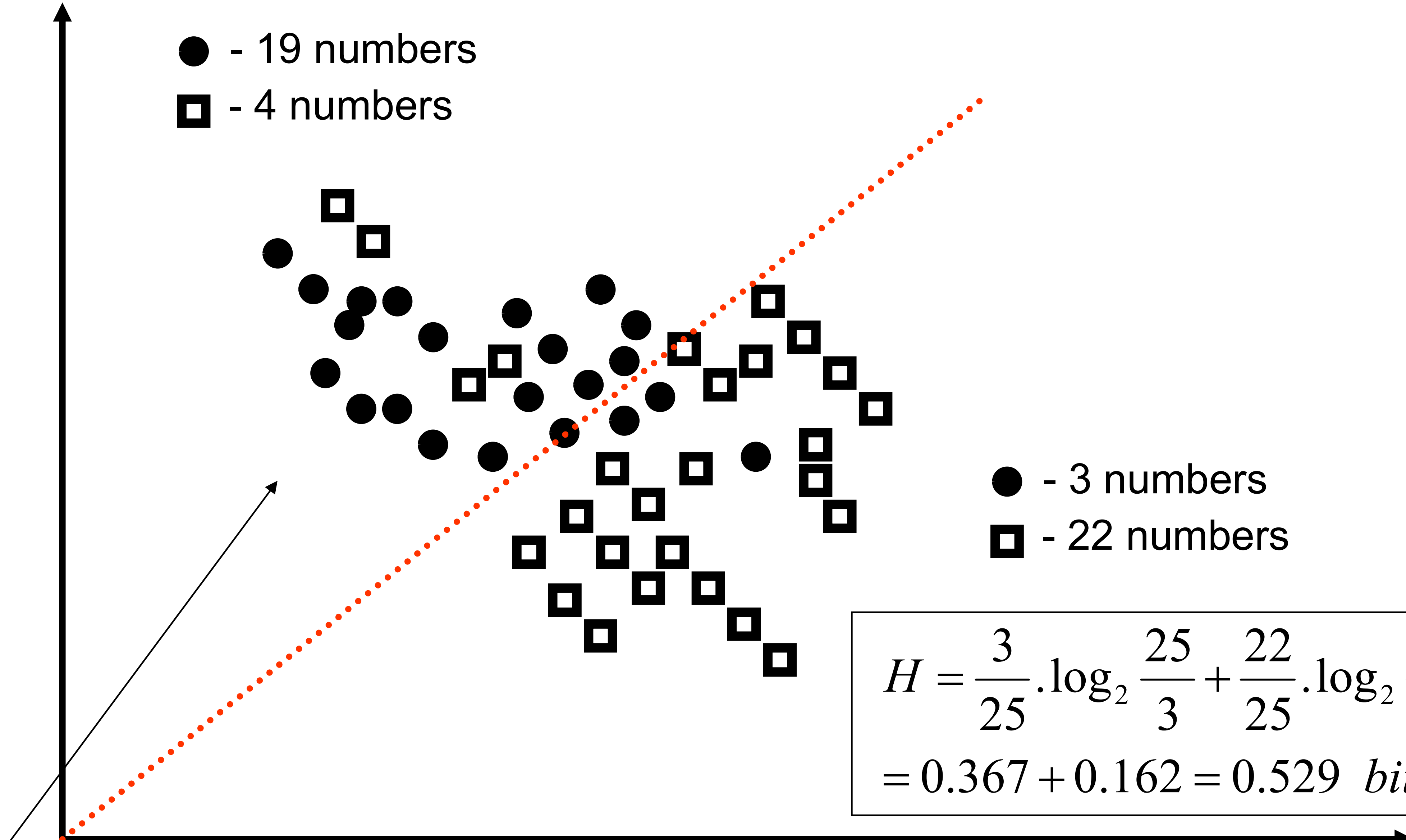


$$H=\frac{21}{48}.\log_2\frac{48}{21}+\frac{27}{48}.\log_2\frac{48}{27}$$
$$=0.522+0.467=0.989\ \ bits$$



$$H = \frac{13}{23}.\log_2 \frac{23}{13} + \frac{10}{23}.\log_2 \frac{23}{10}$$
$$= 0.465 + 0.522 = 0.988 \ \ bits$$

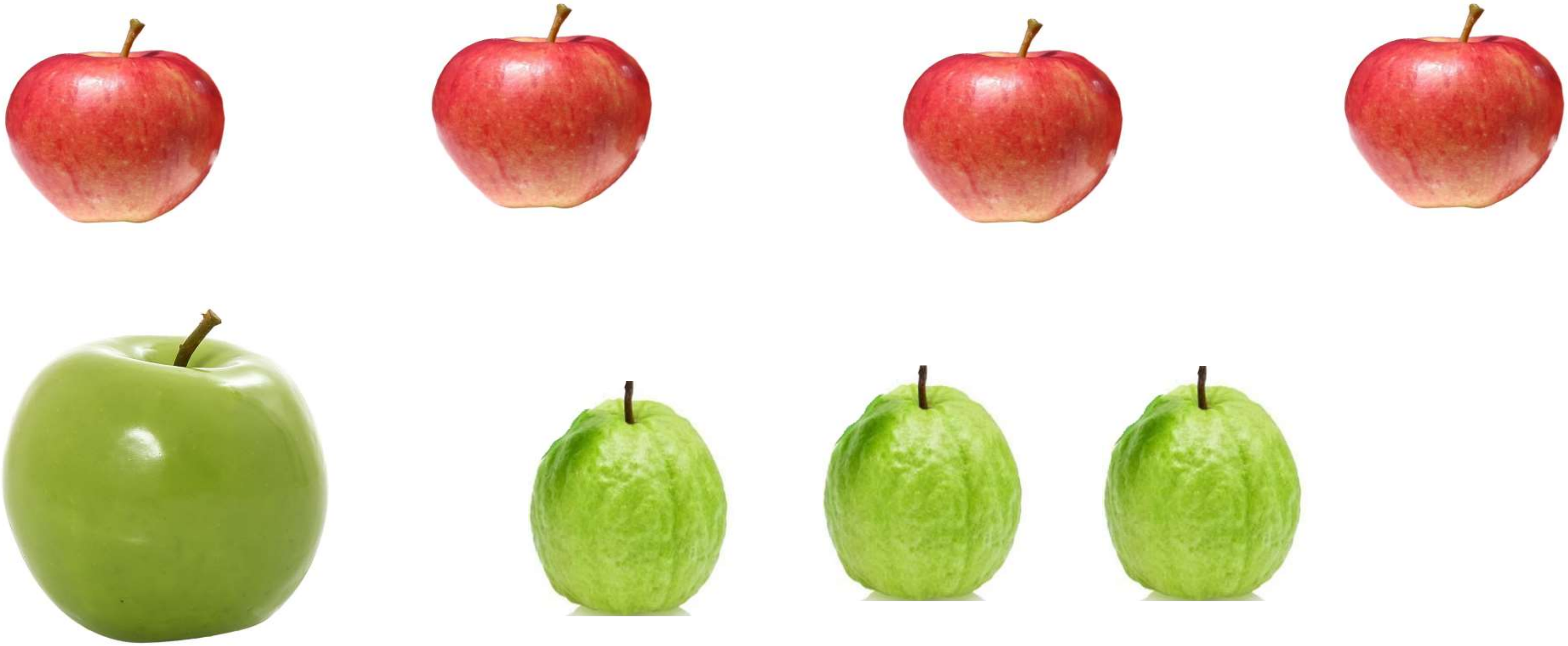
$$H = \frac{9}{25}.\log_2 \frac{25}{9} + \frac{16}{25}.\log_2 \frac{25}{16}$$
$$= 0.531 + 0.412 = 0.943 \ \ bits$$



$$H = \frac{3}{25}.\log_2 \frac{25}{3} + \frac{22}{25}.\log_2 \frac{25}{22}$$
$$= 0.367 + 0.162 = 0.529 \ \ bits$$

$$H = \frac{19}{23}.\log_2 \frac{23}{19} + \frac{4}{23}.\log_2 \frac{23}{4}$$
$$= 0.228 + 0.439 = 0.667 \ \ bits$$

What is entropy?



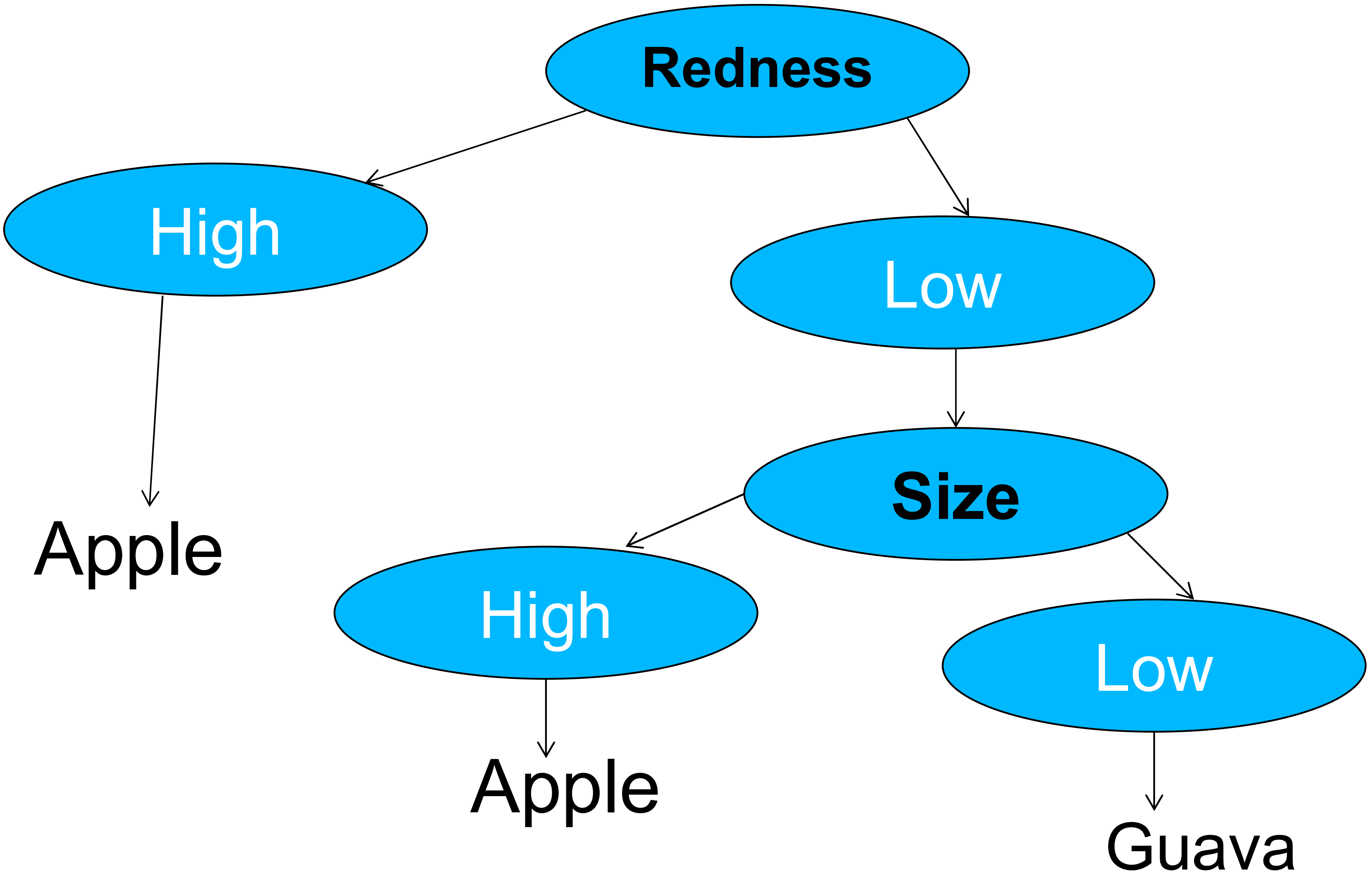
Observations

(1) Green apples are bigger in size compared to red apples

(2) Guavas are smaller than green apples but are equal to red apples

Redness	Size	Fruit	
High	Low	Apple	
Low	High	Apple	
Low	Low	Guava	

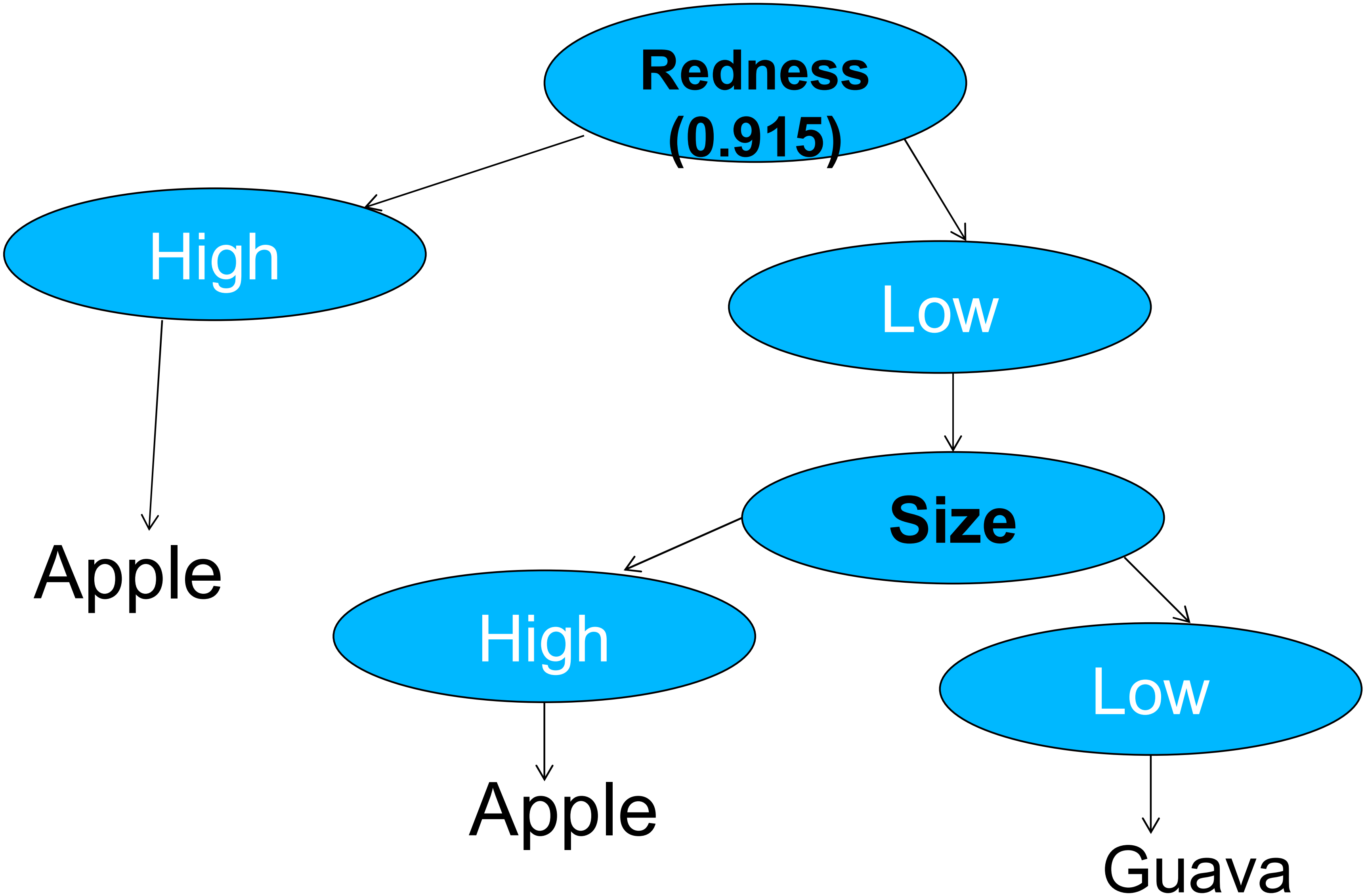
Building decision tree



Entropy of redness

- Two outcomes
  - Low redness
  - High redness
- P(low redness) = 2/3=0.67
- P(high redness) = 1/3=0.33
- H=-0.67*log(0.67) – 0.33*log(0.33)
  =0.915 bits

Building decision tree



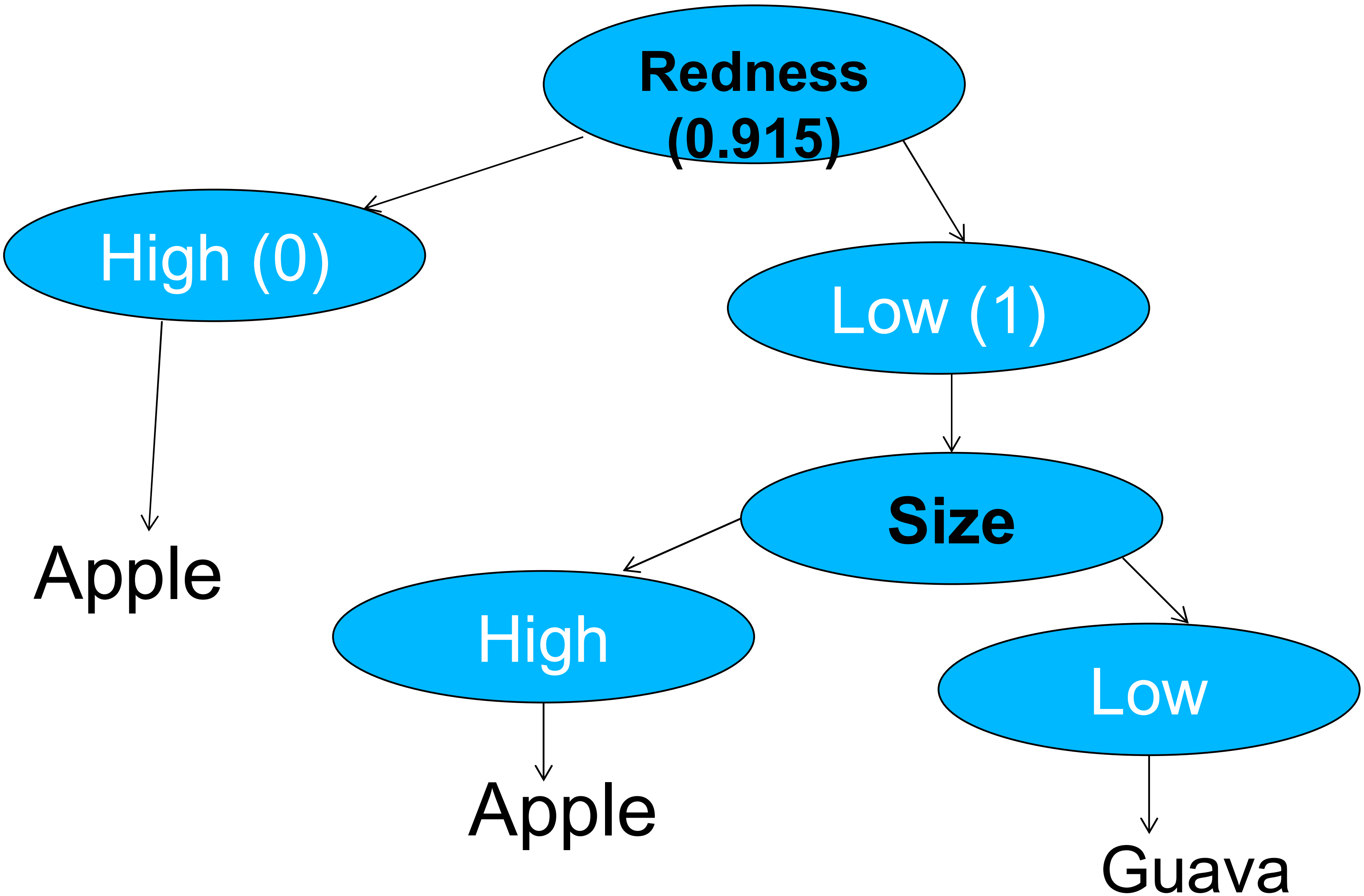
Entropy of High redness

- Only one object i.e. apple
- H = 0

Entropy of Low redness

- Could be either apple or guava
- $H = -0.5*\log(0.5) - 0.5*\log(0.5) = 1$ bit

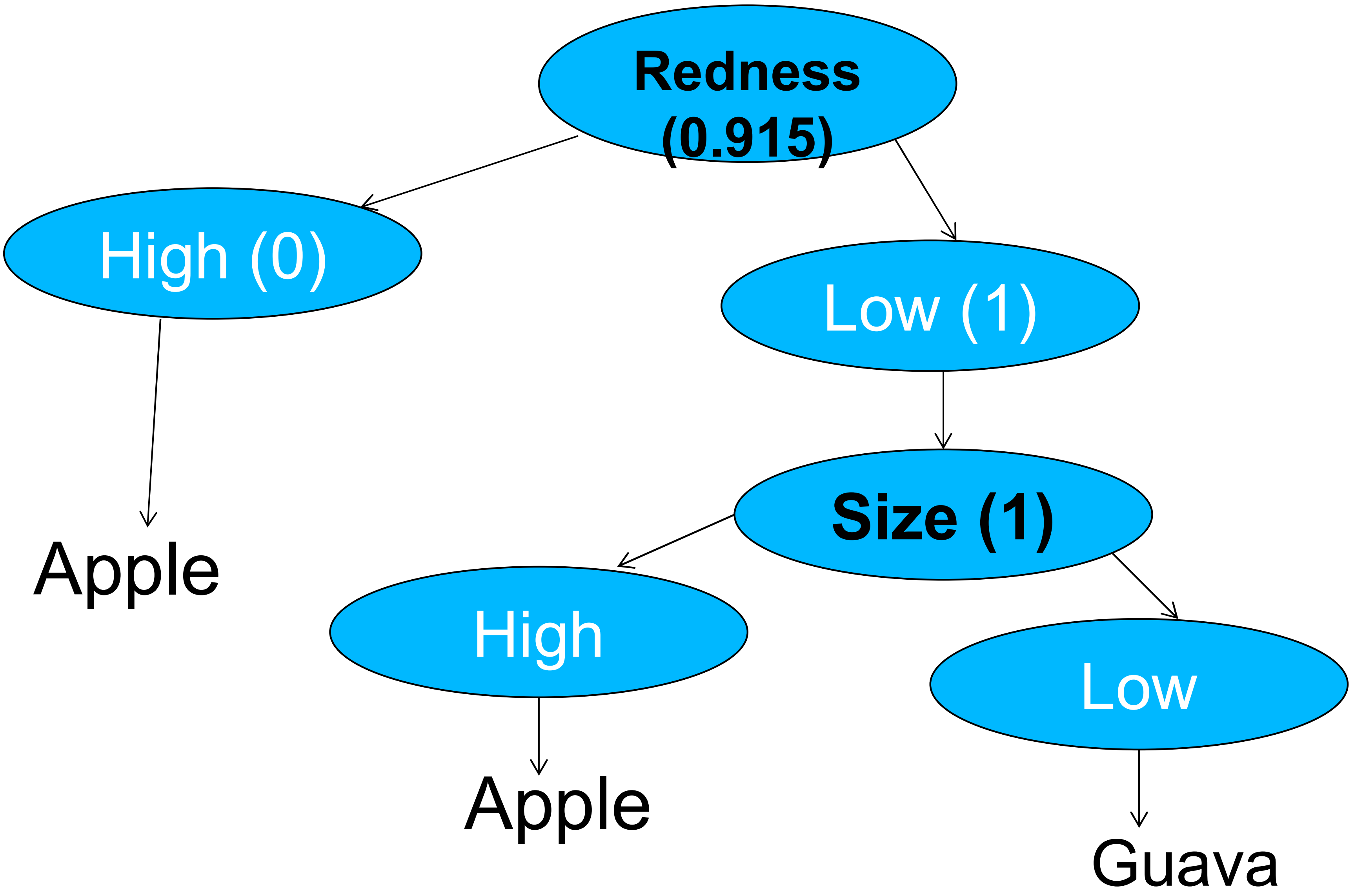
Building decision tree



Entropy of size with low redness

- Two outcomes
  - Low size
  - High size
- P(low size) = 1/2=0.5
- P(high size) = 1/2=0.5
- $H=-0.5*\log(0.5) - 0.5*\log(0.5)$
  $=1$ bit

Building decision tree



Entropy of High size (with low redness)

- Only one object i.e. apple
- H = 0

Entropy of Low size (with low redness)

- Only one object i.e. Guava
- H = 0

One more example – make use of information gain

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Find H of E

Total = 16

8 positives

8 negatives

P(+ve) = P(-ve)=0.5

H(E) = -0.5*log(0.5) -0.5*log(0.5) = 1 bit

Starting
(+ as well as)
H = 1
-

Based on A, B, C, D – we classify

- Randomly choose some number for A say 5
- No. of A ≥ 5= 12
- No. of A < 5= 4

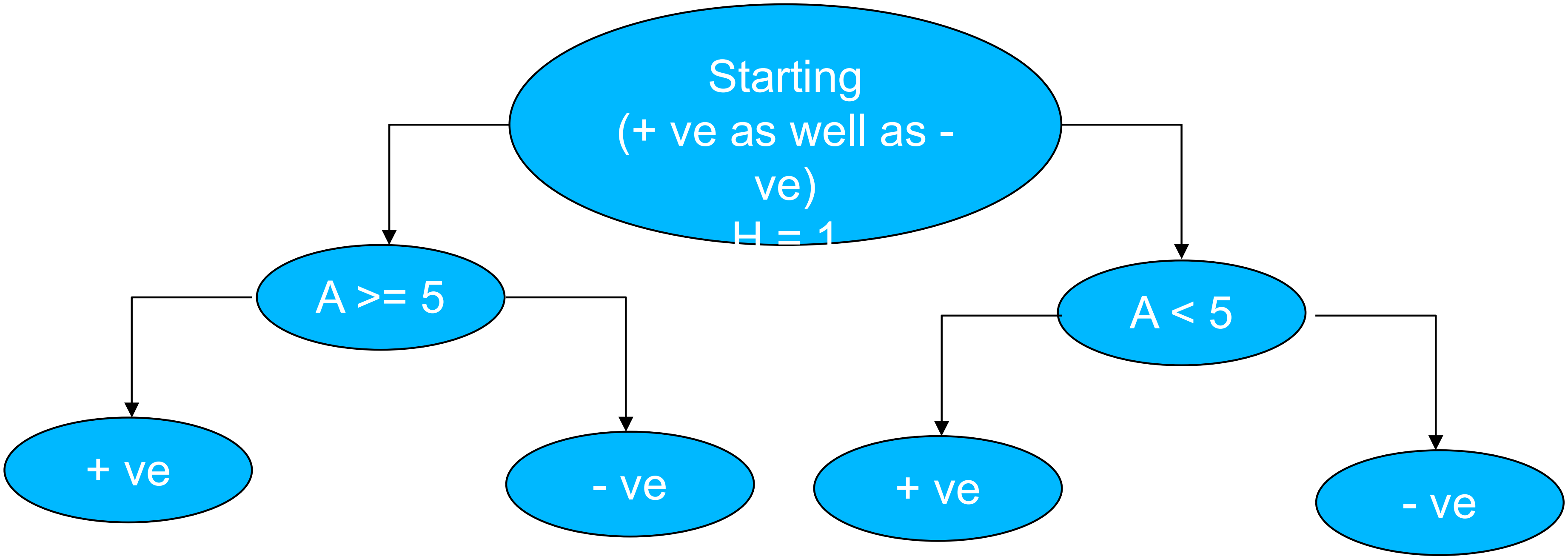
	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Entropy of ($A \geq 5$)

- No. of $A \geq 5$= 12
- Some of them +ve and some of them –ve
- 5 positive
- 7 negative
- P(+ve with $A \geq 5$) = 5/12
- P(-ve with $A \geq 5$) = 7/12
- $H(A \geq 5)$=0.98 bits

Entropy of ($A < 5$)

- No. of $A < 5$= 4
- Some of them +ve and some of them –ve
- 3 positive
- 1 negative
- P(+ve with $A < 5$) = 3/4
- P(-ve with $A < 5$) = 1/4
- $H(A<5)$=0.81 bits


Starting
(+ ve as well as - ve)
H = 1
A >= 5
A < 5
+ ve
- ve
+ ve
- ve

Entropy of A

- Average of $H(A \geq 5)$ and $H(A < 5)$
- $P(A \geq 5)*H(A \geq 5) + P(A<5)*H(A<5)$
- $=12/16*0.98+4/16*0.81 = 0.938$ bits

- Information Gain=H(E) – H(A)

$=1-0.938=0.062$ bits

Based on A, B, C, D – we classify

- Randomly choose some number for B say 3
- No. of B ≥ 3= 12
- No. of B < 3= 4

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Entropy of $(B \geq 3)$

- No. of $B \geq 3$= 12
- Some of them +ve and some of them –ve
- 8 positive
- 4 negative
- P(+ve with $B \geq 3$) = 8/12
- P(-ve with $B \geq 3$) = 4/12
- $H(B \geq 3)=0.915$ bits

Entropy of $(B < 3)$

- No. of $B < 3$= 4
- Some of them +ve and some of them –ve
- 0 positive
- 4 negative
- P(+ve with $B < 3$) = 0
- P(-ve with $B < 3$) = 1
- $H(B<3)=0$

Entropy of B

- Average of $H(B \geq 3)$ and $H(B < 3)$
- $P(B \geq 3)*H(B \geq 3) + P(B<3)*H(B<3)$
- $=12/16*0.915+4/16*0 = 0.687$ bits

- Information Gain=H(E) – H(B)

$=1-0.687=0.313$ bits

Based on A, B, C, D – we classify

- Randomly choose some number for C say 4.2
- No. of C ≥ 4.2= 6
- No. of C< 4.2 = 10

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Entropy of ($C \geq 5$)

- No. of $C \geq 4.2$= 6
- Some of them +ve and some of them –ve
- 0 positive
- 6 negative
- P(+ve with $C \geq 4.2$) = 0/6
- P(-ve with $C \geq 4.2$) = 6/6
- $H(C \geq 4.2)=0$

Entropy of ($C < 5$)

- No. of $C < 4.2$ = 10
- Some of them +ve and some of them –ve
- 8 positive
- 2 negative
- P(+ve with $C < 4.2$) = 8/10
- P(-ve with $C < 4.2$) = 2/10
- $H(C < 4.2)=0.723$ bits

Entropy of C

- Average of $H(C \geq 4.2)$ and $H(C < 4.2)$
- $P(C \geq 4.2)*H(C \geq 4.2) + P(C< 4.2)*H(C< 4.2)$
- $=6/16*0+10/16*0.723 = 0.452$ bits

- Information Gain=H(E) – H(C)

$=1-0.452=0.548$ bits

Based on A, B, C, D – we classify

- Randomly choose some number for D say 1.4
- No. of D ≥ 1.4= 5
- No. of D< 1.4 = 11

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Entropy of ($D \geq 1.4$)

- No. of $D \geq 1.4$= 5
- Some of them +ve and some of them –ve
- 0 positive
- 5 negative
- P(+ve with $D \geq 1.4$) = 0/5
- P(-ve with $D \geq 1.4$) = 5/5
- $H(D \geq 1.4)=0$

Entropy of ($D< 1.4$)

- No. of $D< 1.4$ = 11
- Some of them +ve and some of them –ve
- 8 positive
- 3 negative
- P(+ve with $D<1.4$) = 8/11
- P(-ve with $D<1.4$) = 3/11
- $H(D< 1.4)=0.846$ bits

Entropy of D

- Average of H(D ≥ 1.4) and H(D < 1.4)
- P(D ≥ 4.2)*H(D≥1.4) + P(D<1.4)*H(D<1.4)
- =5/16*0+11/16*0.846 = 0.582 bits

- Information Gain=H(E) – H(D)

=1–0.582=0.418 bits

Order A, B, C, D based on information gain

IG of A = 0.062 bits

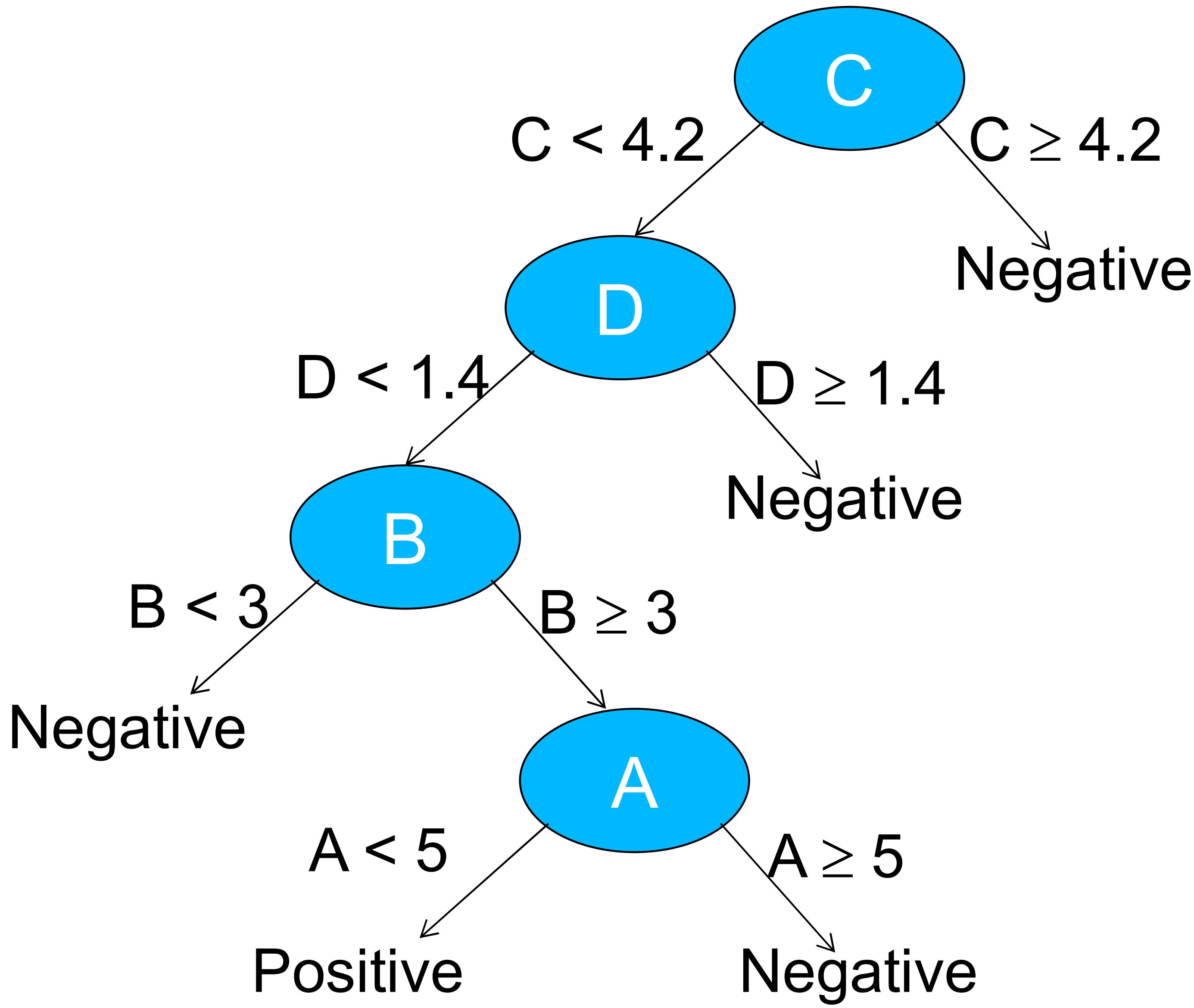
IG of B = 0.313 bits

IG of C = 0.548 bits

IG of D = 0.418 bits

Rank them

C, D, B, A


C
C < 4.2
C ≥ 4.2
Negative
D
D < 1.4
D ≥ 1.4
Negative
B
B < 3
B ≥ 3
Negative
A
A < 5
A ≥ 5
Positive
Negative

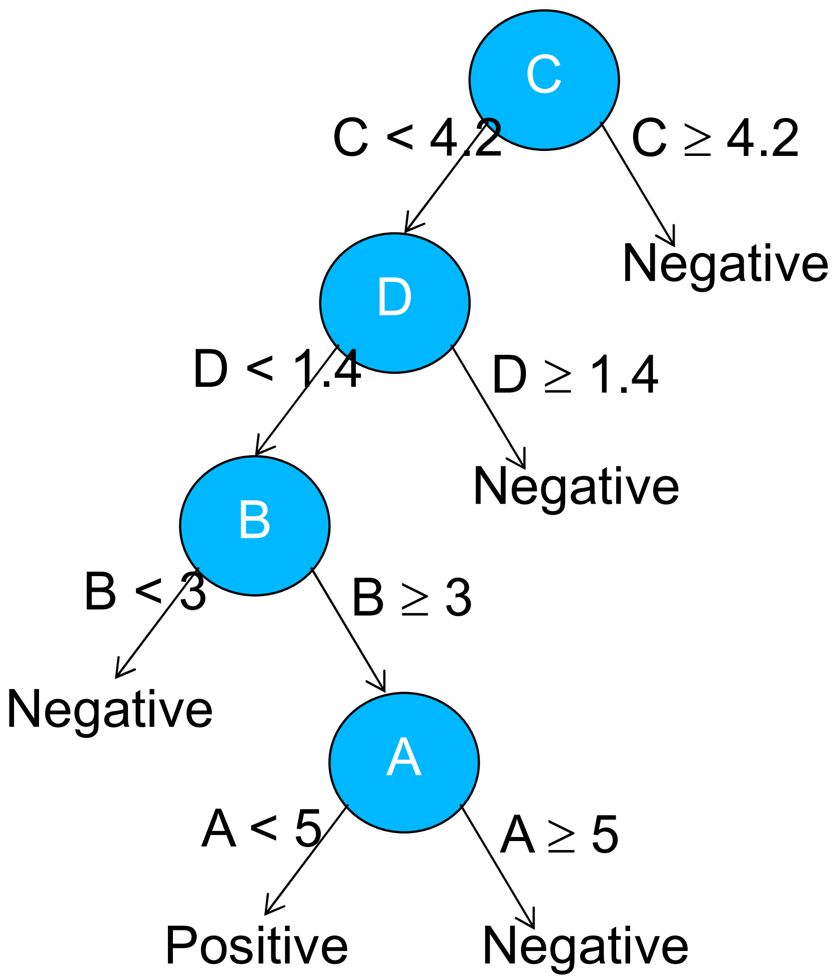
Decision Tree algorithm

- Target
- Target entropy
- Variables
- Variable entropy
- Information gain
- Ranking
- Drawing tree

Rule based classification

What is rule?

If ***something*** Then ***something***



Decision Tree ≡ Rule-based

If C >=4.2

then conclude negative

If C < 4.2 & D >=1.4

then conclude negative

If C < 4.2 & D < 1.4 & B >=3

then conclude negative

If C < 4.2 & D < 1.4 & B < 3

then conclude negative

...

...

Not Unique

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Find H of E
Total = 16
8 positives
8 negatives
P(+ve) = P(-ve)=0.5

H(E) = -0.5*log(0.5)
-0.5*log(0.5) = 1 bit

Based on A, B, C, D – we classify

- Randomly choose some number for A say 6.5
- No. of A ≥ 6.5= 3
- No. of A < 6.5= 13

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Entropy of ($A \geq 6.5$)

- No. of $A \geq 6.5$= 3
- Some of them +ve and some of them –ve
- 0 positive
- 3 negative
- P(+ve with $A \geq 6.5$) = 0
- P(-ve with $A \geq 6.5$) = 1
- $H(A \geq 6.5)=0$

Entropy of ($A < 6.5$)

- No. of $A < 6.5$= 13
- Some of them +ve and some of them –ve
- 8 positive
- 5 negative
- P(+ve with $A < 6.5$) = 8/13
- P(-ve with $A < 6.5$) = 5/13
- $H(A<6.5)=0.96$ bits

Entropy of A

- Average of $H(A \geq 6.5)$ and $H(A < 6.5)$
- $P(A \geq 6.5)*H(A \geq 6.5) + P(A<6.5)*H(A<6.5)$
- $=3/16*0+13/16*0.96 = 0.78$ bits

- Information Gain=H(E) – H(A)

$=1–0.78=0.22$ bits

Based on A, B, C, D – we classify

- Randomly choose some number for B say 3
- No. of B ≥ 3= 12
- No. of B < 3= 4

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Entropy of ($B \geq 3$)

- No. of $B \geq 3$= 12
- Some of them +ve and some of them –ve
- 8 positive
- 4 negative
- P(+ve with $B \geq 3$) = 8/12
- P(-ve with $B \geq 3$) = 4/12
- $H(B \geq 3)$=0.915 bits

Entropy of ($B < 3$)

- No. of $B < 3$= 4
- Some of them +ve and some of them –ve
- 0 positive
- 4 negative
- P(+ve with $B < 3$) = 0
- P(-ve with $B < 3$) = 1
- $H(B<3)=0$

Entropy of B

- Average of $H(B \geq 3)$ and $H(B < 3)$
- $P(B \geq 3)*H(B \geq 3) + P(B<3)*H(B<3)$
- $=12/16*0.915+4/16*0 = 0.687$ bits

- Information Gain=H(E) – H(B)

$=1-0.687=0.313$ bits

Based on A, B, C, D – we classify

- Randomly choose some number for C say 4.2
- No. of C ≥ 4.2= 6
- No. of C< 4.2 = 10

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Entropy of ($C \geq 4.2$)

- No. of $C \geq 4.2$= 6
- Some of them +ve and some of them –ve
- 0 positive
- 6 negative
- P(+ve with $C \geq 4.2$) = 0/6
- P(-ve with $C \geq 4.2$) = 6/6
- $H(C \geq 4.2)=0$

Entropy of ($C < 4.2$)

- No. of $C< 4.2$ = 10
- Some of them +ve and some of them –ve
- 8 positive
- 2 negative
- P(+ve with $C <$ 4.2) = 8/10
- P(-ve with $C < 4.2$) = 2/10
- $H(C< 4.2)=0.723$ bits

Entropy of C

- Average of $H(C \geq 4.2)$ and $H(C < 4.2)$
- $P(C \geq 4.2)*H(C \geq 4.2) + P(C< 4.2)*H(C< 4.2)$
- $=6/16*0+10/16*0.723 = 0.452$ bits

- Information Gain=H(E) – H(C)

$=1-0.452=0.548$ bits

Based on A, B, C, D – we classify

- Randomly choose some number for D say 1.4
- No. of D ≥ 1.4= 5
- No. of D< 1.4 = 11

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Entropy of ($D \geq 1.4$)

- No. of $D \geq 1.4$= 5
- Some of them +ve and some of them –ve
- 0 positive
- 5 negative
- P(+ve with $D \geq 1.4$) = 0/5
- P(-ve with $D \geq 1.4$) = 5/5
- $H(D \geq 1.4)=0$

Entropy of ($D< 1.4$)

- No. of $D< 1.4$ = 11
- Some of them +ve and some of them –ve
- 8 positive
- 3 negative
- P(+ve with $D<1.4$) = 8/11
- P(-ve with $D<1.4$) = 3/11
- $H(D< 1.4)=0.846$ bits

Entropy of D

- Average of H(D ≥ 1.4) and H(D < 1.4)
- P(D ≥ 4.2)*H(D≥1.4) + P(D<1.4)*H(D<1.4)
- =5/16*0+11/16*0.846 = 0.582 bits

- Information Gain=H(E) – H(D)

=1–0.582=0.418 bits

Order A, B, C, D based on information gain

IG of A = 0.22 bits

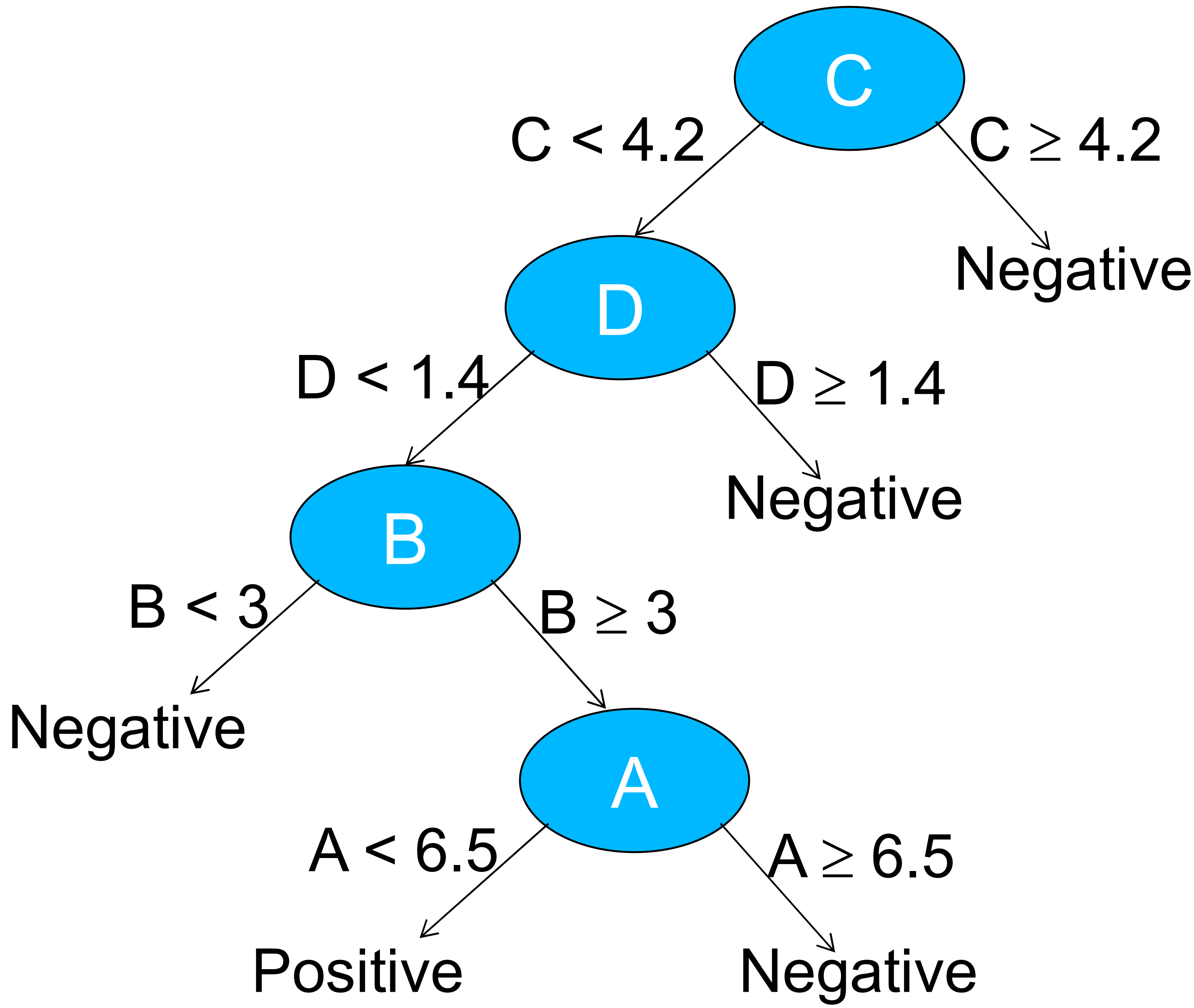
IG of B = 0.313 bits

IG of C = 0.548 bits

IG of D = 0.418 bits

Rank them

C, D, B, A


C
C < 4.2
C ≥ 4.2
Negative
D
D < 1.4
D ≥ 1.4
Negative
B
B < 3
B ≥ 3
Negative
A
A < 6.5
A ≥ 6.5
Positive
Negative

Not Unique

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Find H of E

Total = 16

8 positives

8 negatives

P(+ve) = P(-ve)=0.5

H(E) = -0.5*log(0.5)
 -0.5*log(0.5) = 1 bit

Based on A, B, C, D – we classify

- Randomly choose some number for A say 6.5
- No. of A ≥ 6.5= 3
- No. of A < 6.5= 13

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Entropy of (A $\geq$ 6.5)

- No. of A $\geq$ 6.5= 3
- Some of them +ve and some of them –ve
- 0 positive
- 3 negative
- P(+ve with A $\geq$ 6.5) = 0
- P(-ve with A $\geq$ 6.5) = 1
- H(A $\geq$ 6.5)=0

Entropy of (A $<$ 6.5)

- No. of A $<$ 6.5= 13
- Some of them +ve and some of them –ve
- 8 positive
- 5 negative
- P(+ve with A $<$ 6.5) = 8/13
- P(-ve with A $<$ 6.5) = 5/13
- H(A$<$6.5)=0.96 bits

Entropy of A

- Average of $H(A \geq 6.5)$ and $H(A < 6.5)$
- $P(A \geq 6.5) * H(A \geq 6.5) + P(A<6.5) * H(A<6.5)$
- $=3/16*0+13/16*0.96 = 0.78$ bits

- Information Gain=H(E) – H(A)

$=1-0.78=0.22$ bits

Based on A, B, C, D – we classify

- Randomly choose some number for B say 2.4
- No. of B ≥ 2.4= 15
- No. of B < 2.4= 1

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Entropy of ($B \geq 2.4$)

- No. of $B \geq 2.4$= 15
- Some of them +ve and some of them –ve
- 8 positive
- 7 negative
- P(+ve with $B \geq 2.4$) = 8/15
- P(-ve with $B \geq 2.4$) = 7/15
- $H(B \geq 2.4)=0.997$ bits

Entropy of ($B < 2.4$)

- No. of $B < 2.4$ = 1
- Some of them +ve and some of them –ve
- 0 positive
- 1 negative
- P(+ve with $B < 2.4$) = 0
- P(-ve with $B < 2.4$) = 1
- $H(B<2.4)=0$

Entropy of B

- Average of $H(B \geq 2.4)$ and $H(B < 2.4)$
- $P(B \geq 2.4)*H(B \geq 2.4) + P(B<2.4)*H(B<2.4)$
- $=15/16*0.915+1/16*0 = 0.857$ bits

- Information Gain=H(E) – H(B)

$=1-0.857=0.143$ bits

Based on A, B, C, D – we classify

- Randomly choose some number for C say 4.2
- No. of C ≥ 4.2= 6
- No. of C< 4.2 = 10

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Entropy of ($C \geq 4.2$)

- No. of $C \geq 4.2 = 6$
- Some of them +ve and some of them –ve
- 0 positive
- 6 negative
- P(+ve with $C \geq 4.2$) = 0/6
- P(-ve with $C \geq 4.2$) = 6/6
- $H(C \geq 4.2)=0$

Entropy of ($C < 4.2$)

- No. of $C < 4.2 = 10$
- Some of them +ve and some of them –ve
- 8 positive
- 2 negative
- P(+ve with $C < 4.2$) = 8/10
- P(-ve with $C < 4.2$) = 2/10
- $H(C < 4.2)=0.723$ bits

Entropy of C

- Average of H(C $\geq$ 4.2) and H(C < 4.2)
- P(C $\geq$ 4.2)*H(C $\geq$ 4.2) + P(C< 4.2)*H(C< 4.2)
- =6/16*0+10/16*0.723 = 0.452 bits

- Information Gain=H(E) – H(C)

=1–0.452=0.548 bits

Based on A, B, C, D – we classify

- Randomly choose some number for D say 1.4
- No. of D ≥ 1.4= 5
- No. of D< 1.4 = 11

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	4.9	2.4	3.3	1	negative

Entropy of (D ≥ 1.4)

- No. of D ≥ 1.4= 5
- Some of them +ve and some of them –ve
- 0 positive
- 5 negative
- P(+ve with D ≥ 1.4) = 0/5
- P(-ve with D ≥ 1.4) = 5/5
- H(D ≥ 1.4)=0

Entropy of (D< 1.4)

- No. of D< 1.4 = 11
- Some of them +ve and some of them –ve
- 8 positive
- 3 negative
- P(+ve with D<1.4) = 8/11
- P(-ve with D<1.4) = 3/11
- H(D< 1.4)=0.846 bits

Entropy of D

- Average of $H(D \geq 1.4)$ and $H(D < 1.4)$
- $P(D \geq 4.2) * H(D \geq 1.4) + P(D < 1.4) * H(D < 1.4)$
- $= 5/16 * 0 + 11/16 * 0.846 = 0.582$ bits

- Information Gain $= H(E) - H(D)$

$= 1 - 0.582 = 0.418$ bits

Order A, B, C, D based on information gain

IG of A = 0.22 bits

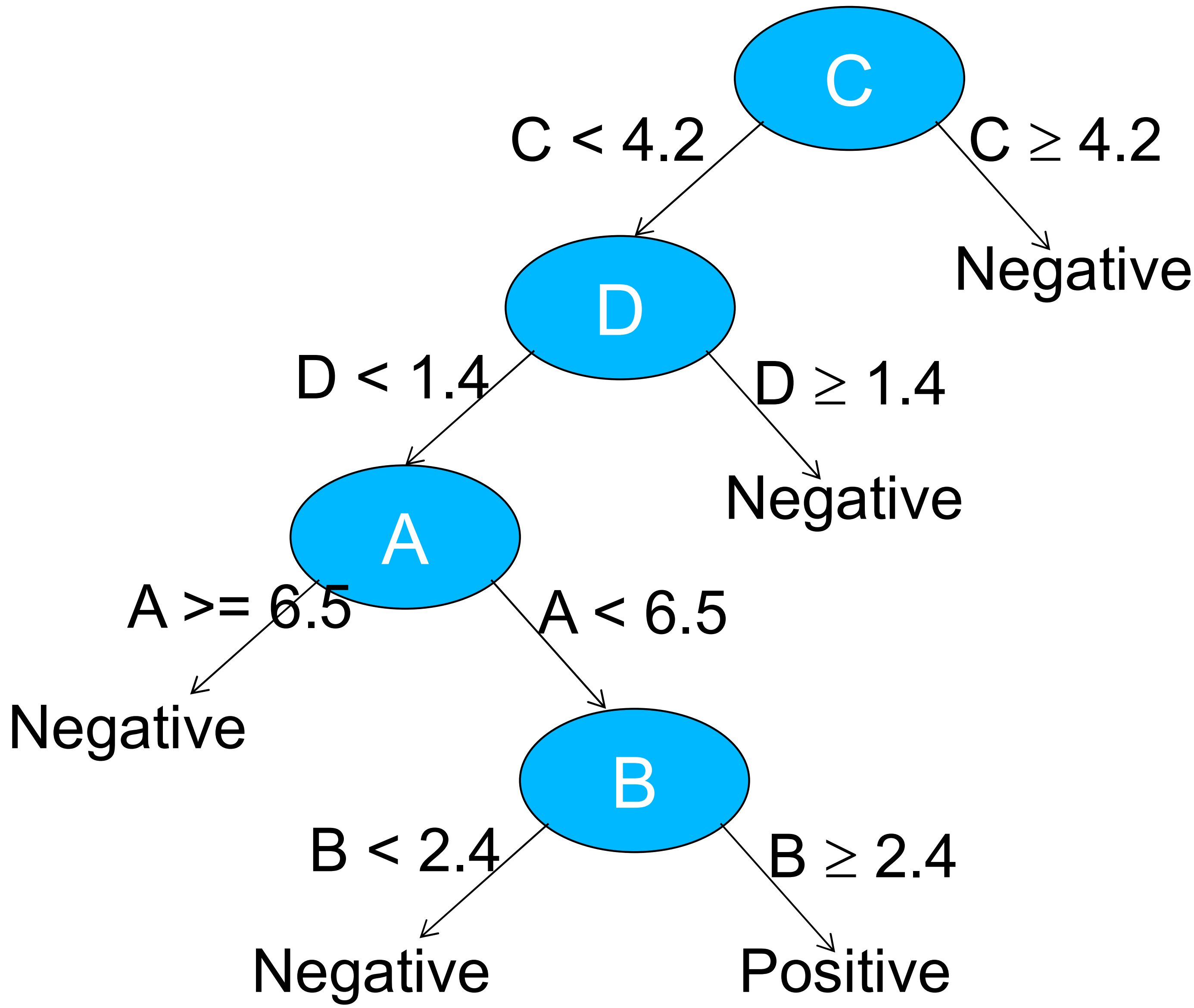
IG of B = 0.143 bits

IG of C = 0.548 bits

IG of D = 0.418 bits

Rank them

C, D, A, B


C
C < 4.2
C ≥ 4.2
Negative
D
D < 1.4
D ≥ 1.4
Negative
A
A >= 6.5
A < 6.5
Negative
B
B < 2.4
B ≥ 2.4
Negative
Positive

What'll happen here?

	A	B	C	D	E
1	4.8	3.4	1.9	0.2	positive
2	5	3	1.6	0.2	positive
3	5	3.4	1.6	0.4	positive
4	5.2	3.5	1.5	0.2	positive
5	5.2	3.4	1.4	0.2	positive
6	4.7	3.2	1.6	0.2	positive
7	4.8	3.1	1.6	0.2	positive
8	5.4	3.4	1.5	0.4	positive
9	7	3.2	4.7	1.4	negative
10	6.4	3.2	4.5	1.5	negative
11	6.9	3.1	4.9	1.5	negative
12	5.5	2.3	4	1.3	negative
13	6.5	2.8	4.6	1.5	negative
14	5.7	2.8	4.5	1.3	negative
15	6.3	3.3	4.7	1.6	negative
16	6.3	3.3	1.5	1.6	negative