08.10.2024

Statistical Methods in AI (CS7.403)

Lecture-17: Decision Tree Learning

Ravi Kiran (ravi.kiran@iiit.ac.in)

https://ravika.github.io

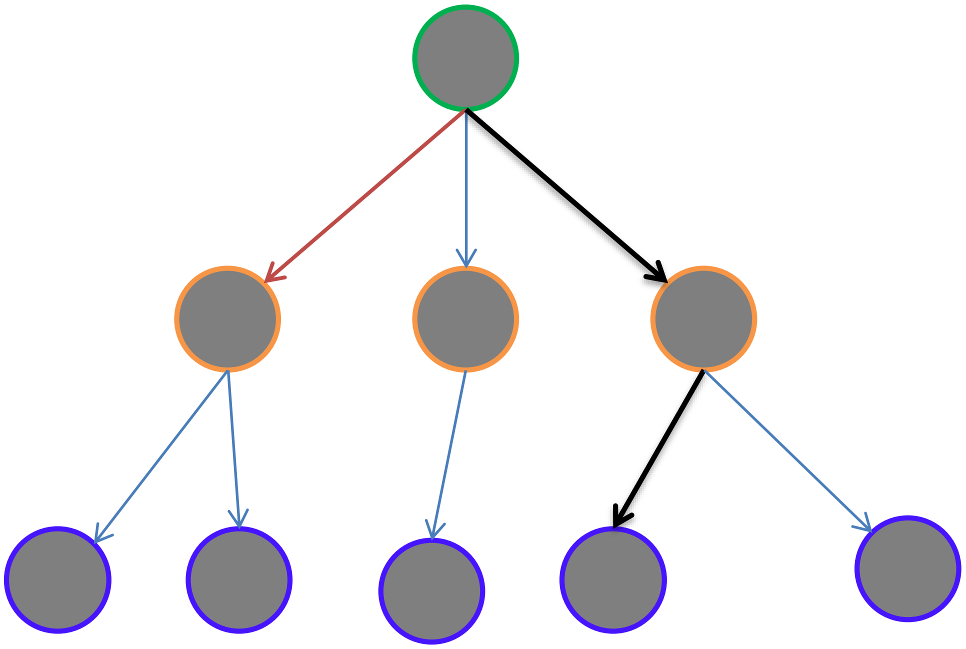
@vikataravi

Center for Visual Information Technology (CVIT)

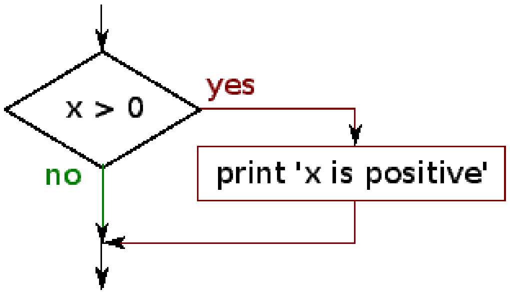
IIIT Hyderabad

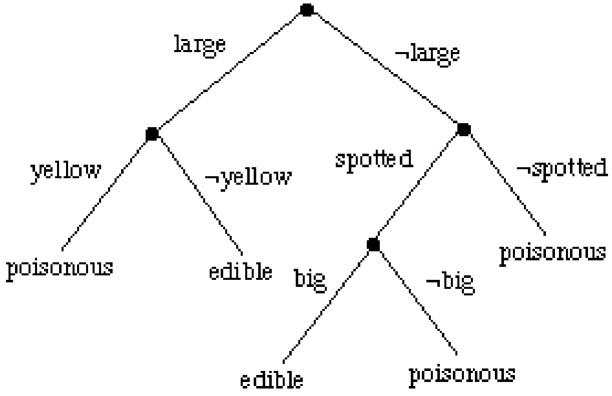
Trees

- ❑ Node
- ❑ Root
- ❑ Leaf
- ❑ Edge/Branch
- ❑ Path
- ❑ Depth

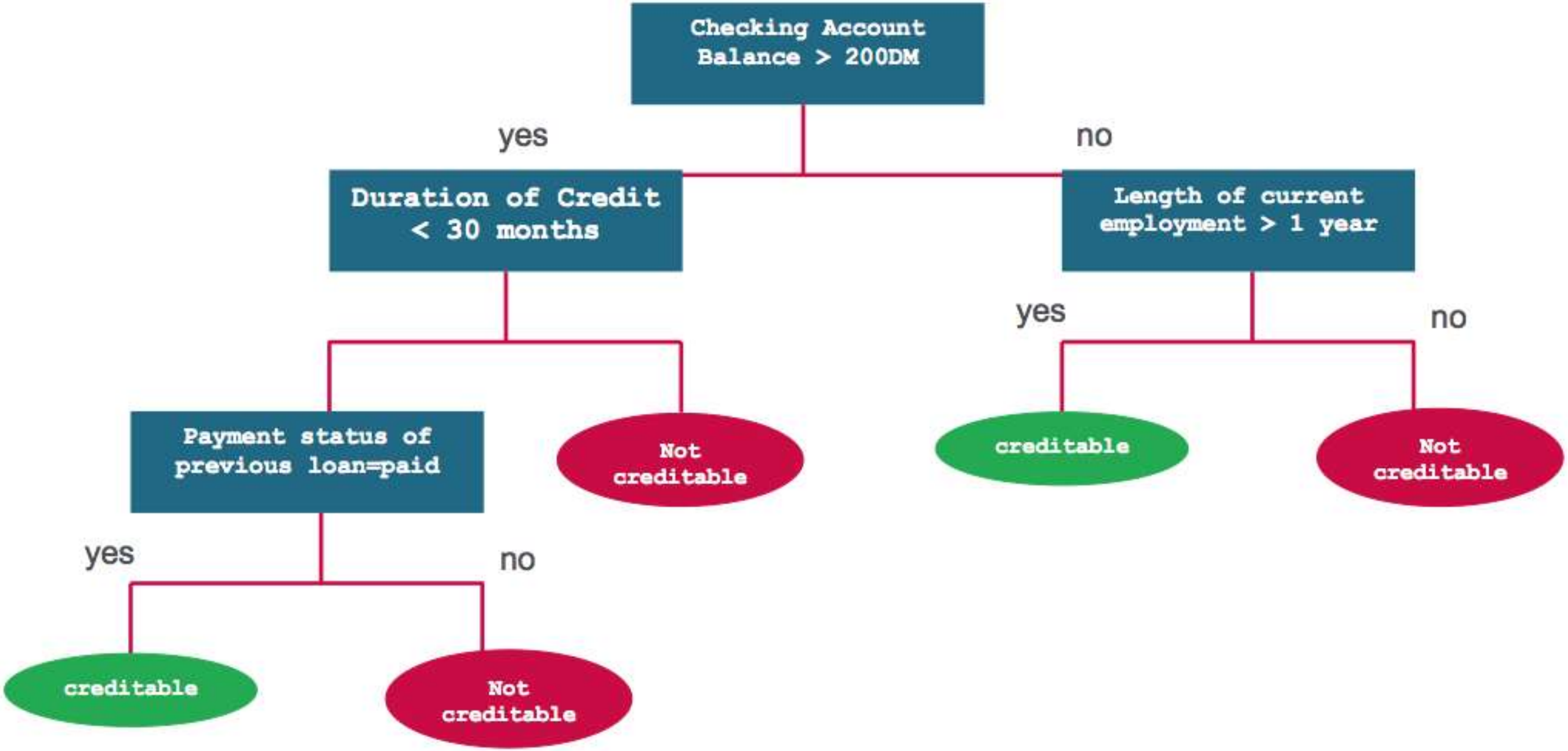


Hand-crafted, fixed trees





Credit Approval



Credit Approval (Raw Data)

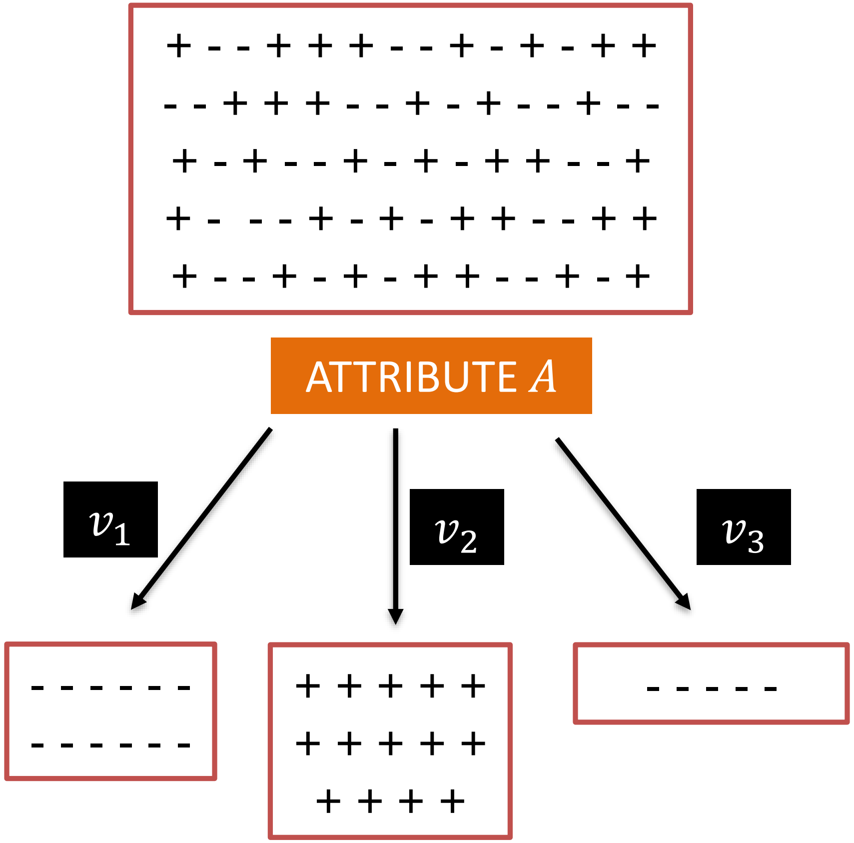
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
64	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15	class
65	a	20.42	0.835	u	g	q	v	1.585	t	t	1	f	g	0	0	+
66	b	26.67	4.25	u	g	cc	v	4.29	t	t	1	t	g	120	0	+
67	b	34.17	1.54	u	g	cc	v	1.54	t	t	1	t	g	520	50000	+
68	a	36	1	u	g	c	v	2	t	t	11	f	g	0	456	+
69	b	25.5	0.375	u	g	m	v	0.25	t	t	3	f	g	260	15108	+
70	b	19.42	6.5	u	g	w	h	1.46	t	t	7	f	g	80	2954	+
71	b	35.17	25.125	u	g	x	h	1.625	t	t	1	t	g	515	500	+
72	b	32.33	7.5	u	g	e	bb	1.585	t	f	0	t	s	420	0	-
73	b	34.83	4	u	g	d	bb	12.5	t	f	0	t	g		0	-
74	a	38.58	5	u	g	cc	v	13.5	t	f	0	t	g	980	0	-
75	b	44.25	0.5	u	g	m	v	10.75	t	f	0	f	s	400	0	-
76	b	44.83	7	y	p	c	v	1.625	f	f	0	f	g	160	2	-
77	b	20.67	5.29	u	g	q	v	0.375	t	t	1	f	g	160	0	-
78	b	34.08	6.5	u	g	aa	v	0.125	t	f	0	t	g	443	0	-

outlook	temp	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

Probability, Information, Entropy

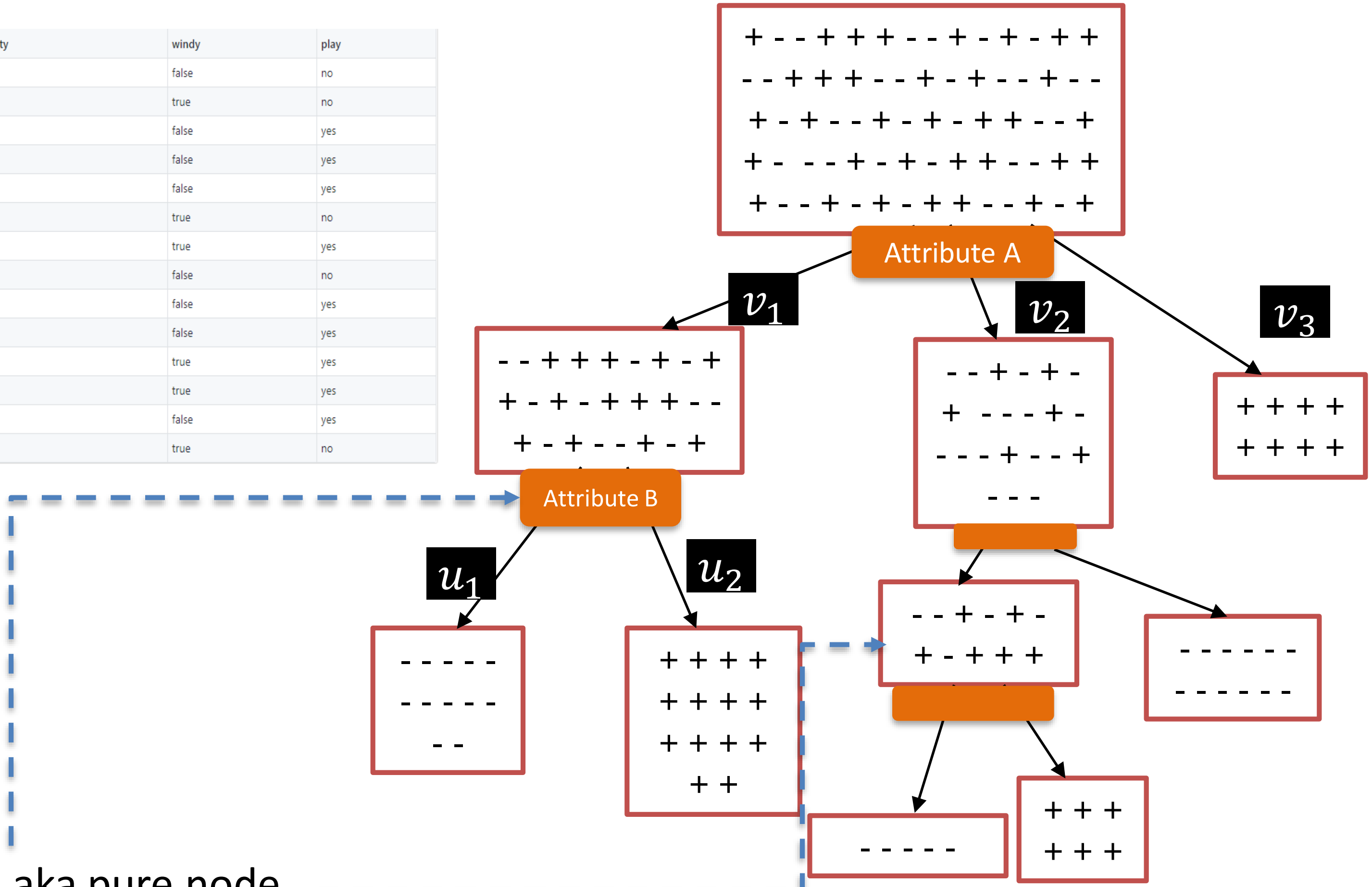
outlook	temp	humidity	windy	play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes
rainy	mild	high	true	no

An ideal attribute



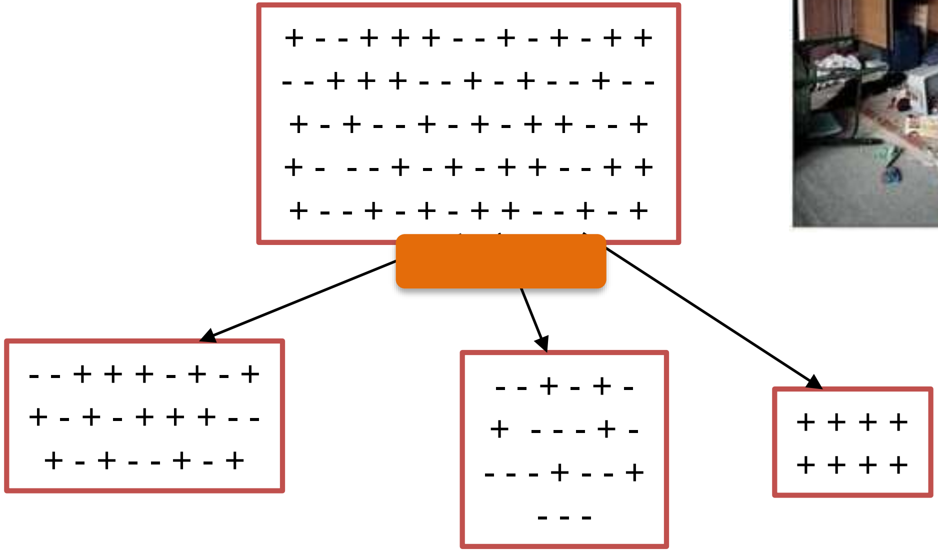
Decision Tree

1	outlook	temp	humidity	windy	play
2	sunny	hot	high	false	no
3	sunny	hot	high	true	no
4	overcast	hot	high	false	yes
5	rainy	mild	high	false	yes
6	rainy	cool	normal	false	yes
7	rainy	cool	normal	true	no
8	overcast	cool	normal	true	yes
9	sunny	mild	high	false	no
10	sunny	cool	normal	false	yes
11	rainy	mild	normal	false	yes
12	sunny	mild	normal	true	yes
13	overcast	mild	high	true	yes
14	overcast	hot	normal	false	yes
15	rainy	mild	high	true	no



Ideal attribute aka pure node

How much ‘impurity’ does this attribute decrease ?



PROBABILITY

What is the probability of a ball falling in a black pocket on a roulette wheel?

Probability is the chance (or likelihood) of an event happening. We can describe probability using words from a probability scale:

← less likely | more likely →

no chance | poor chance | even chance | good chance | certain

What does 'random' mean?
If you put your hand into a bag of marbles and took one out without looking, you picked it at **random**. You didn't choose that one - you got it by chance.

We can also describe probability using fractions.

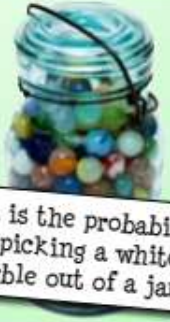
The probability of a flipped coin landing on heads is one out of two. $\frac{1}{2}$

The probability of this spinner landing on '3' is three out of six. $\frac{3}{6}$

The probability of rolling a '2' on a die is one out of six. $\frac{1}{6}$

What is the probability of rolling two sixes?

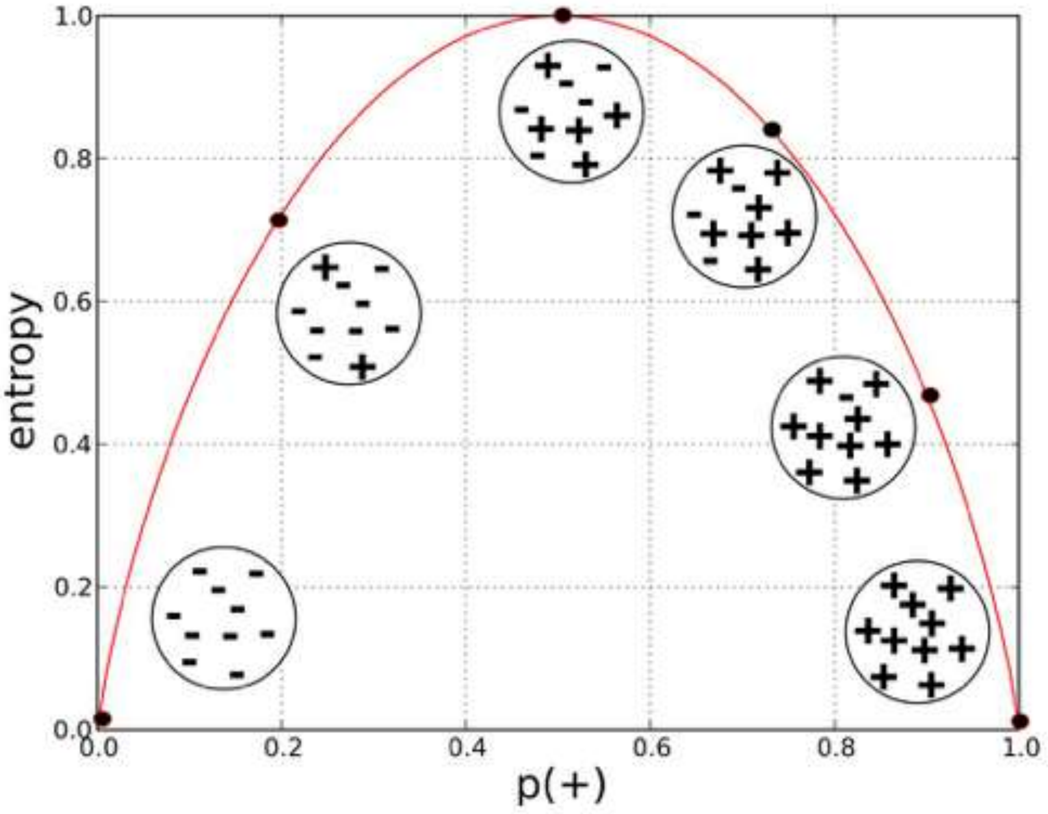
What is the probability of a flipped coin landing on tails?



What is the probability of picking a white marble out of a jar?

What is the probability of picking the Queen of Spades from a deck of cards?

Entropy (2 class)



https://towardsdatascience.com/entropy-how-decision-trees-make-decisions-2946b9c18c8

Step-1: Compute impurity score of training label distribution

Day	Temperature	Outlook	Humidity	Windy	Play Golf?
07-05	hot	sunny	high	false	no
07-06	hot	sunny	high	true	no
07-07	hot	overcast	high	false	yes
07-09	cool	rain	normal	false	yes
07-10	cool	overcast	normal	true	yes
07-12	mild	sunny	high	false	no
07-14	cool	sunny	normal	false	yes
07-15	mild	rain	normal	false	yes
07-20	mild	sunny	normal	true	yes
07-21	mild	overcast	high	true	yes
07-22	hot	overcast	normal	false	yes
07-23	mild	rain	high	true	no
07-26	cool	rain	normal	true	no
07-30	mild	rain	high	false	yes

Entropy: $i(V) = -(q \log q + (1-q) \log(1-q))$

$$E(S) = -(\frac{9}{14}log(\frac{9}{14}) + \frac{5}{14}log(\frac{5}{14})) = 0.94$$

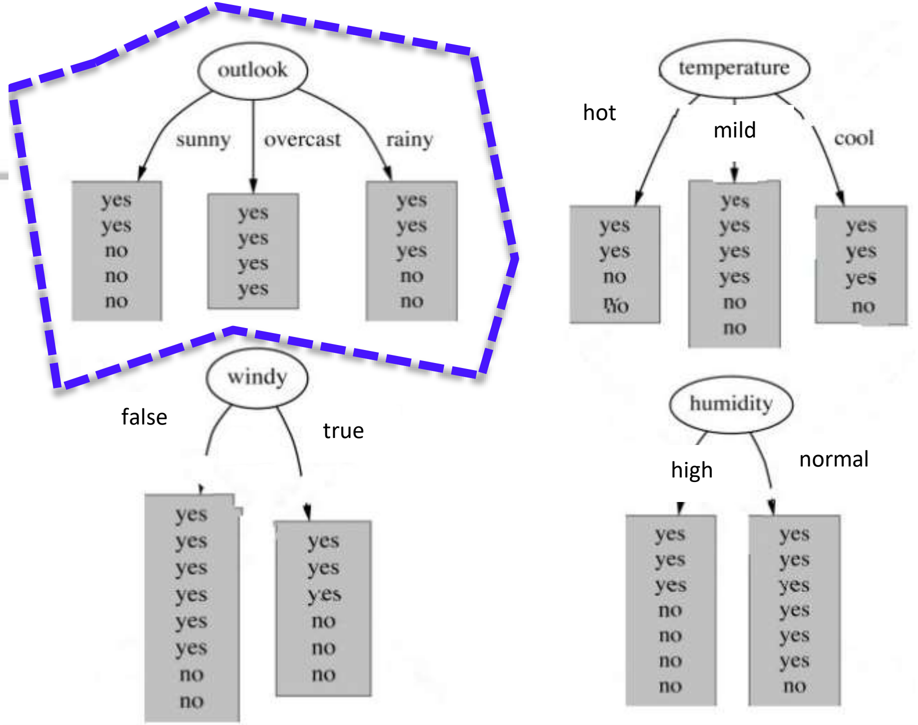
Step-2: Compute impurity score for each unique value of candidate attributes

Example: Attribute Outlook

Entropy: $i(V) = -(q \log q + (1-q)\log(1-q))$

- Outlook = rainy   3 examples yes, 2 examples no

$$E(\text{Outlook}=\text{sunny}) = -\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right) = 0.971$$



Step-2: Compute impurity score for each unique value of candidate attributes

Example: Attribute Outlook

Entropy: $i(V) = -(q \log q + (1-q)\log(1-q))$

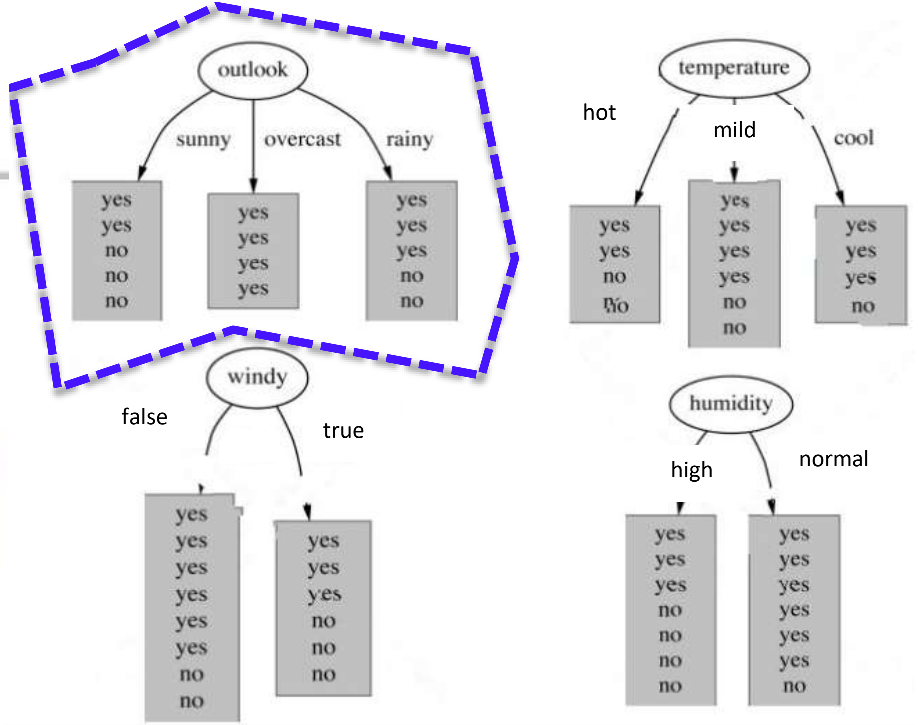
- **Outlook = rainy** 3 examples yes, 2 examples no

$$E(\text{Outlook}=\text{sunny}) = -\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right) = 0.971$$

- Outlook = overcast: 4 examples yes, 0 examples no

$$E(\text{Outlook}=\text{overcast}) = -1\log(1) - 0\log(0) = 0$$

Note: this is normally undefined. Here: = 0



Step-2: Compute impurity score for each unique value of candidate attributes

Example: Attribute Outlook

Entropy: $i(V) = -(q \log q + (1-q)\log(1-q))$

- **Outlook = rainy** 3 examples yes, 2 examples no

$$E(\text{Outlook}=\text{sunny}) = -\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right) = 0.971$$

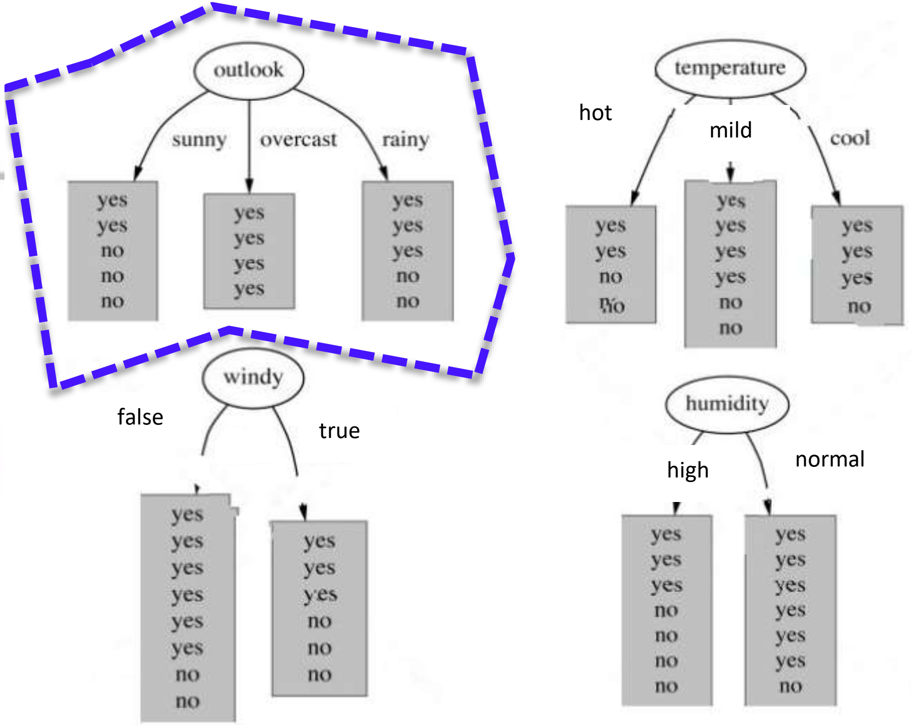
- Outlook = overcast: 4 examples yes, 0 examples no

$$E(\text{Outlook}=\text{overcast}) = -1\log(1) - 0\log(0) = 0$$

Note: this is normally undefined. Here: = 0

- **Outlook = sunny** 2 examples yes, 3 examples no

$$E(\text{Outlook}=\text{rainy}) = -\frac{3}{5}\log\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right) = 0.971$$



Step-3: Compute impurity score for candidate attribute

- **Outlook = rainy** 3 examples yes, 2 examples no

$$E(\text{Outlook}=\text{sunny}) = -\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right) = 0.971$$

- Outlook = overcast: 4 examples yes, 0 examples no

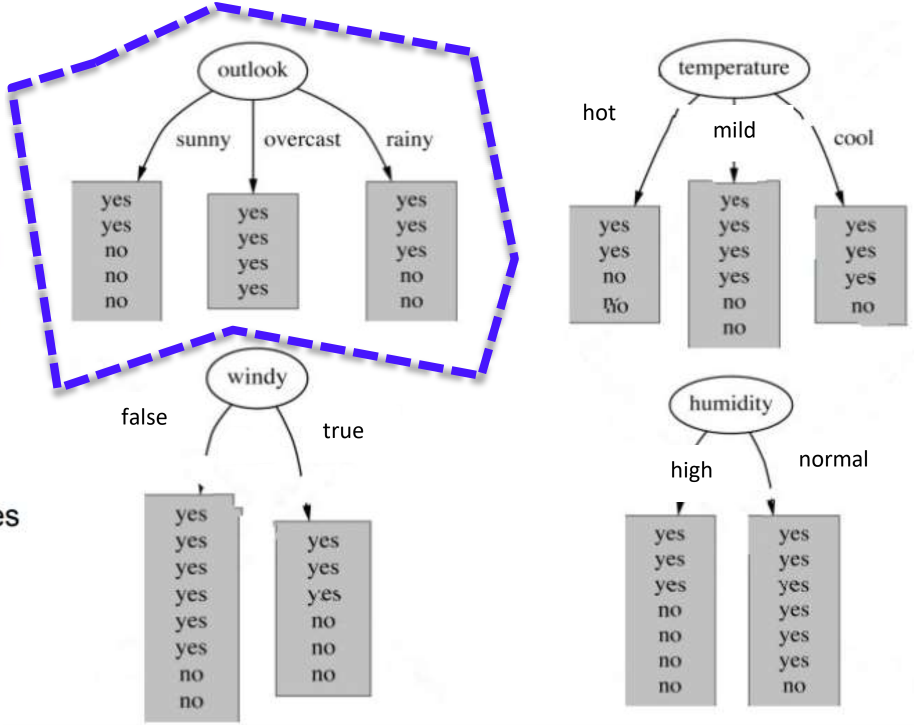
$$E(\text{Outlook}=\text{overcast}) = -1\log(1) - 0\log(0) = 0$$

Note: this is normally undefined. Here: = 0

- **Outlook = sunny** 2 examples yes, 3 examples no

$$E(\text{Outlook}=\text{rainy}) = -\frac{3}{5}\log\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right) = 0.971$$

- Entropy only computes the quality of a single (sub-)set of examples
  - corresponds to a single value
- How can we compute the quality of the entire split?
  - corresponds to an entire attribute



Step-3: Compute impurity score for candidate attribute

- **Outlook = rainy** 3 examples yes, 2 examples no

$$E(\text{Outlook}=\text{sunny}) = -\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right) = 0.971$$

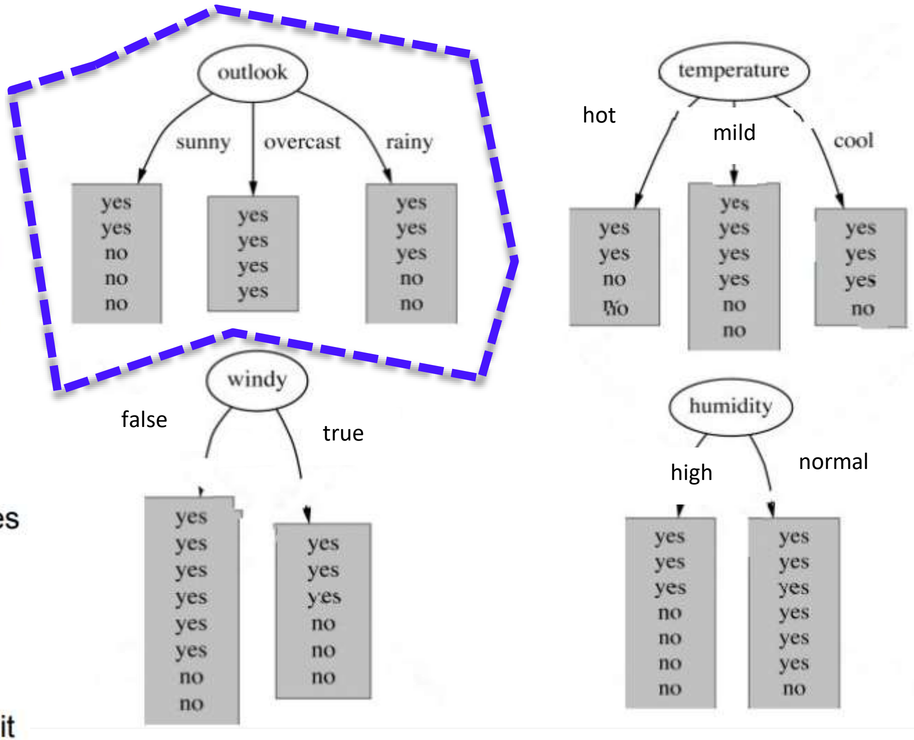
- Outlook = overcast: 4 examples yes, 0 examples no

$$E(\text{Outlook}=\text{overcast}) = -1\log(1) - 0\log(0) = 0$$

Note: this is normally undefined. Here: = 0

- **Outlook = sunny** 2 examples yes, 3 examples no

$$E(\text{Outlook}=\text{rainy}) = -\frac{3}{5}\log\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right) = 0.971$$



- Entropy only computes the quality of a single (sub-)set of examples
  - corresponds to a single value
- How can we compute the quality of the entire split?
  - corresponds to an entire attribute

Solution:

- Compute the weighted average over all sets resulting from the split
  - weighted by their size

$$I(S, A) = \sum_i \frac{|S_i|}{|S|} \cdot E(S_i)$$

Step-3: Compute impurity score for candidate attribute

- **Outlook = rainy** 3 examples yes, 2 examples no

$$E(\text{Outlook}=\text{sunny}) = -\frac{2}{5}\log\left(\frac{2}{5}\right) - \frac{3}{5}\log\left(\frac{3}{5}\right) = 0.971$$

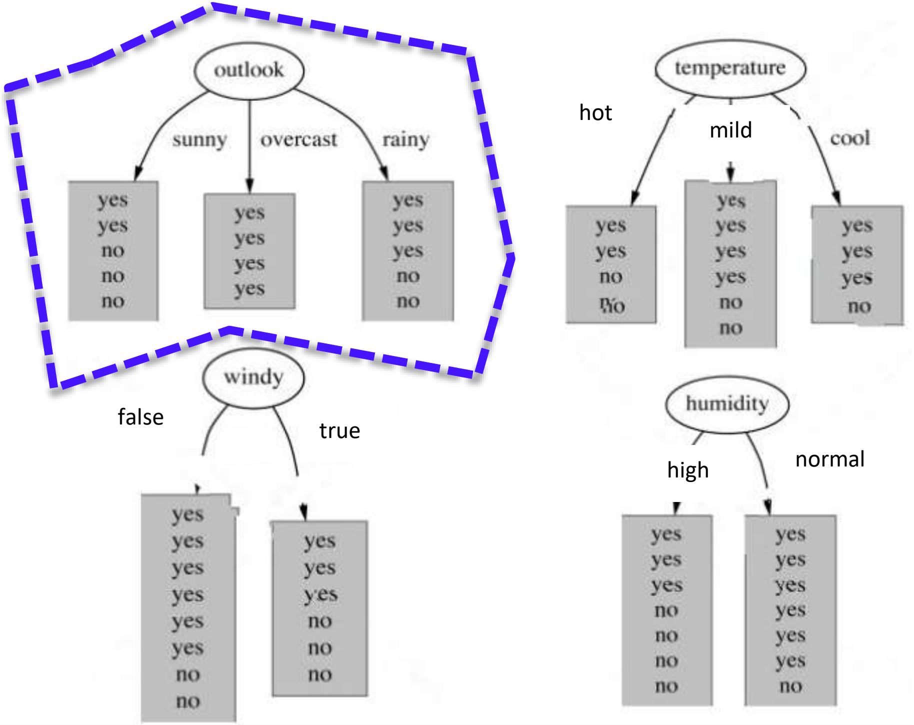
- Outlook = overcast: 4 examples yes, 0 examples no

$$E(\text{Outlook}=\text{overcast}) = -1\log(1) - 0\log(0) = 0$$

Note: this is normally undefined. Here: = 0

- **Outlook = sunny** 2 examples yes, 3 examples no

$$E(\text{Outlook}=\text{rainy}) = -\frac{3}{5}\log\left(\frac{3}{5}\right) - \frac{2}{5}\log\left(\frac{2}{5}\right) = 0.971$$



$$I(S, A) = \sum_i \frac{|S_i|}{|S|} \cdot E(S_i)$$

- Average entropy for attribute *Outlook*:

$$I(\text{Outlook}) = \frac{5}{14}\cdot 0.971 + \frac{4}{14}\cdot 0 + \frac{5}{14}\cdot 0.971 = 0.693$$

Step-4: Compute Information Gain (reduction in impurity score) provided by candidate attribute

$$I(S,A)=\sum_i \frac{|S_i|}{|S|}\cdot E(S_i)$$

- Average entropy for attribute *Outlook*:

$$I(\text{Outlook})=\frac{5}{14}\cdot 0.971+\frac{4}{14}\cdot 0+\frac{5}{14}\cdot 0.971=0.693$$

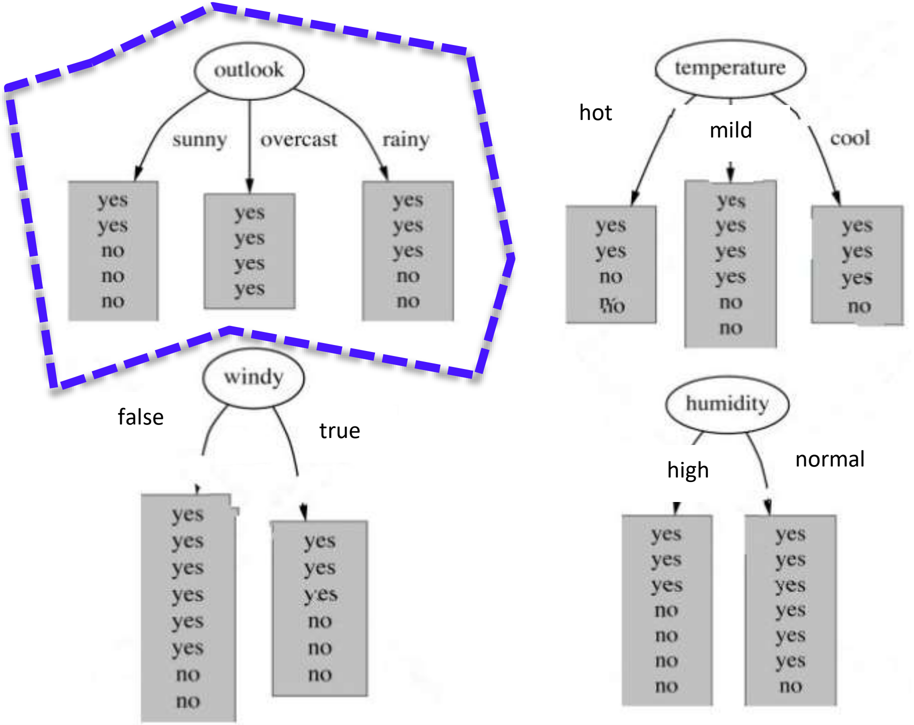
Entropy of root

$$E(S)=-(\frac{9}{14}log(\frac{9}{14})+\frac{5}{14}log(\frac{5}{14}))=0.94$$

Information Gain for Attribute A

$$Gain(S,A)=E(S)-I(S,A)=E(S)-\sum_i \frac{|S_i|}{|S|}\cdot E(S_i)$$

$$Gain(S,Outlook)=0.246$$



Step-5: Compare Information Gain provided by all candidates

Information Gain for Attribute A

$$Gain(S,A)=E(S)-I(S,A)=E(S)-\sum_i \frac{|S_i|}{|S|}\cdot E(S_i)$$

Gain (S, Humidity)

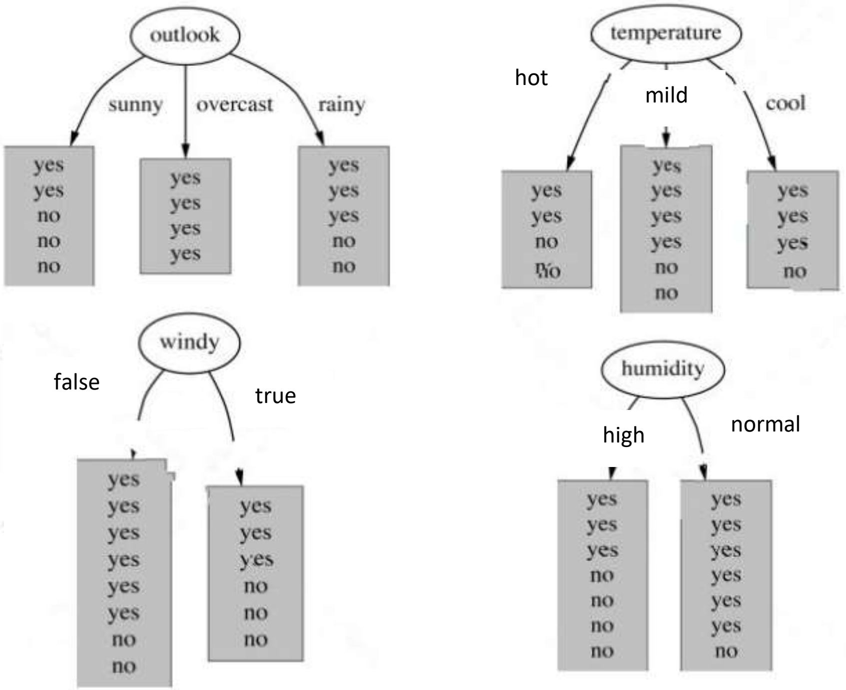
$= .940 - (7/14).985 - (7/14).592$
$= .151$

Gain (S, Wind)

$= .940 - (8/14).811 - (6/14)1.0$
$= .048$

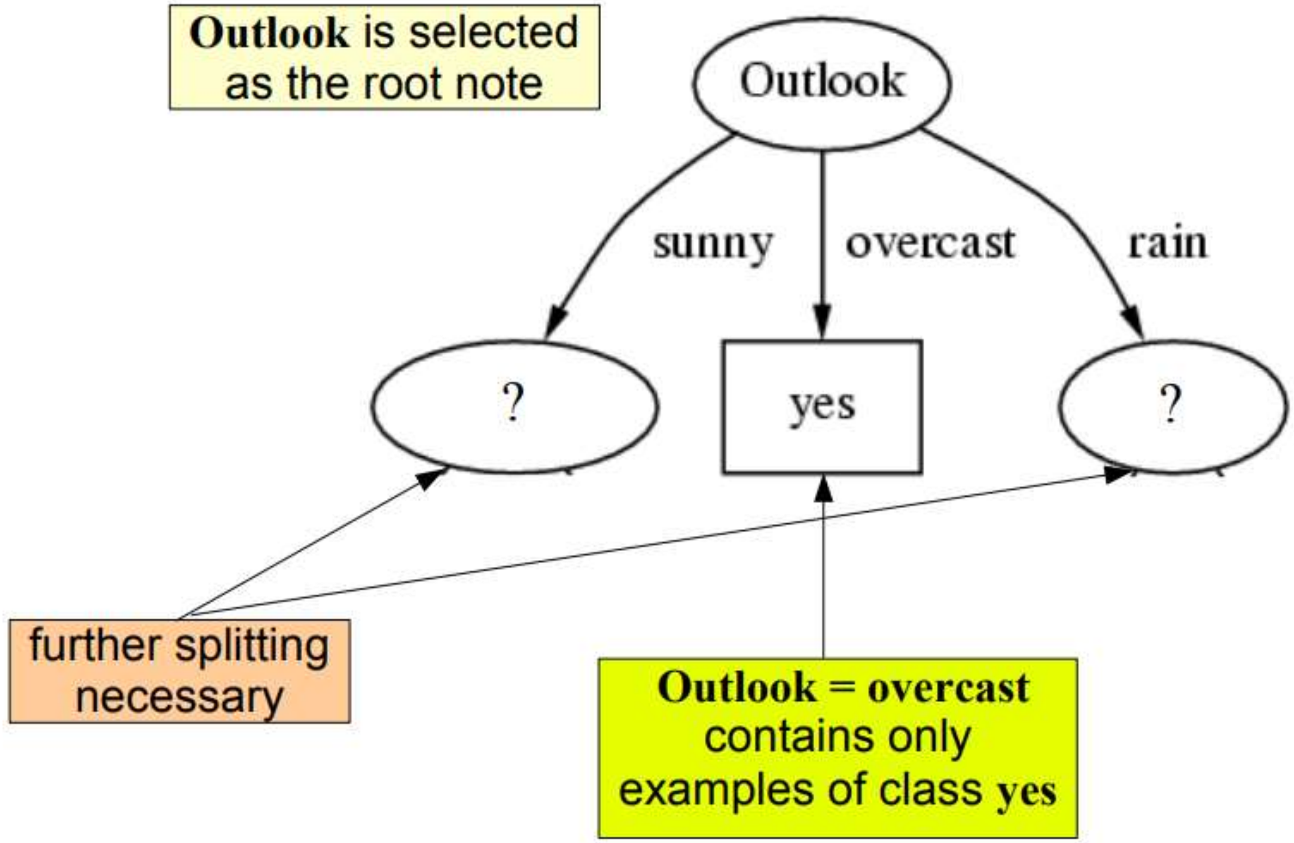
$Gain(S, Outlook)=\boxed{0.246}$

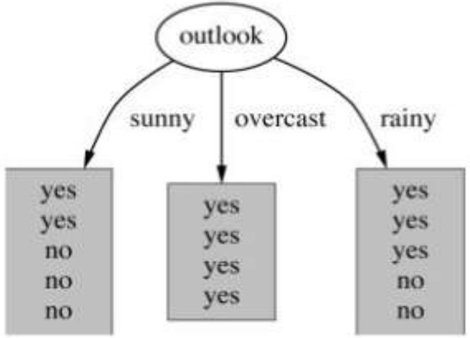
$Gain(S, Temperature)=0.029$



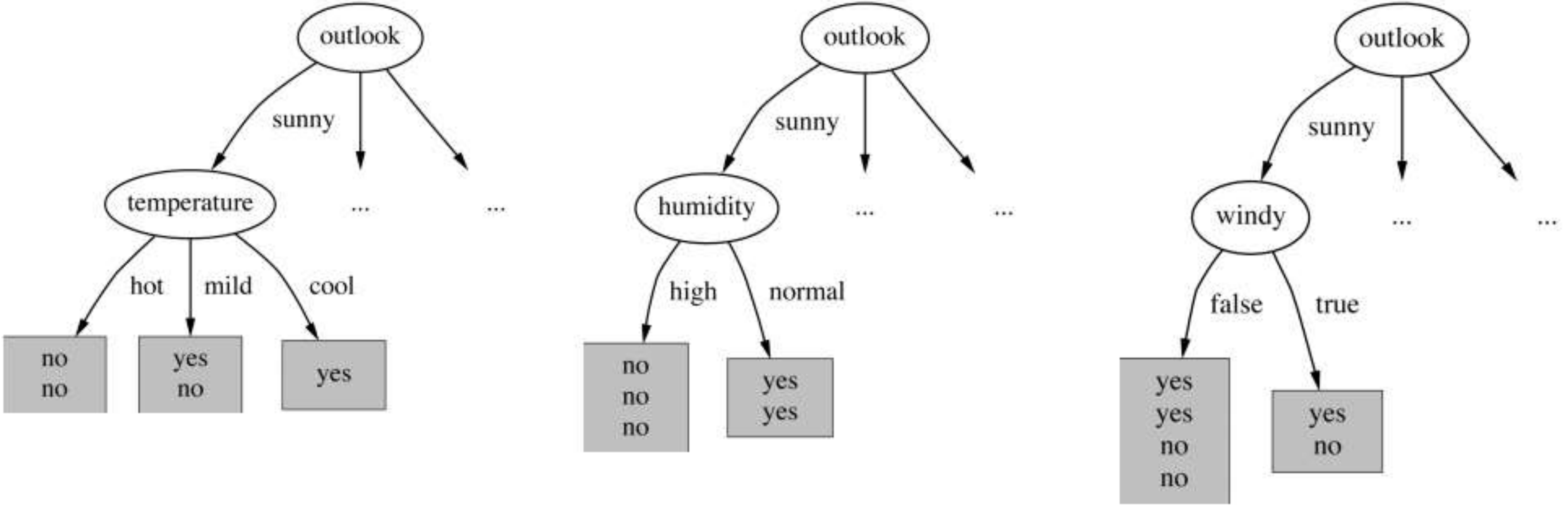
Select the attribute which provides largest 'impurity reduction'/Information Gain

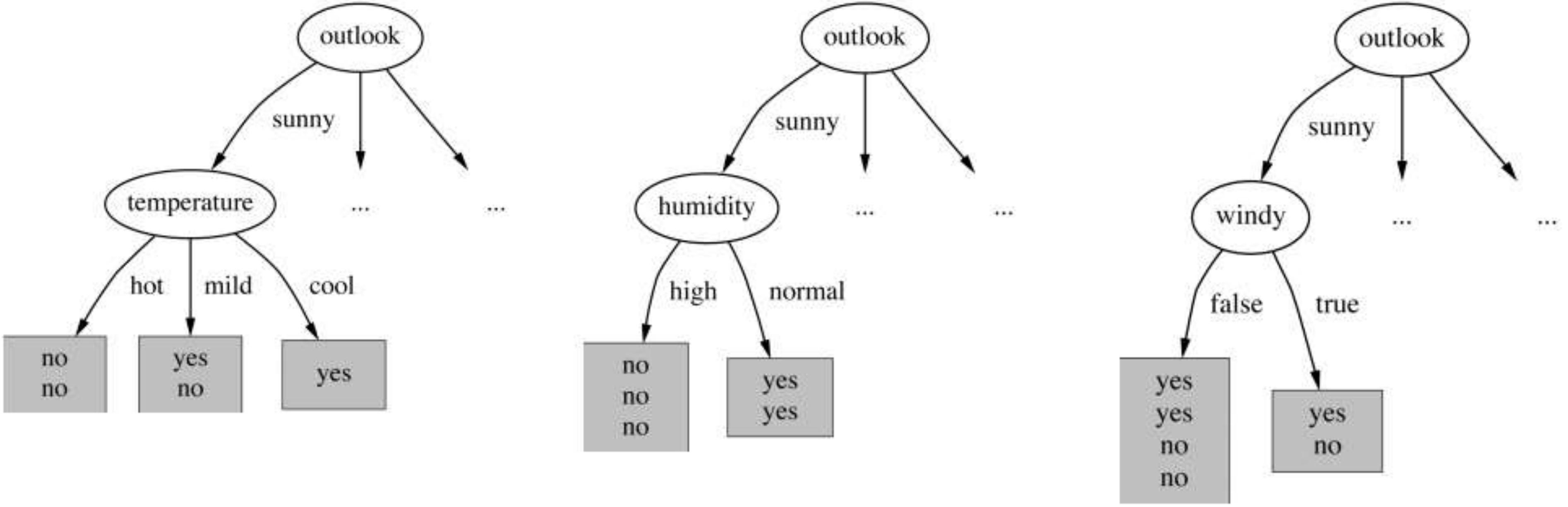
Step-6: Assign root node





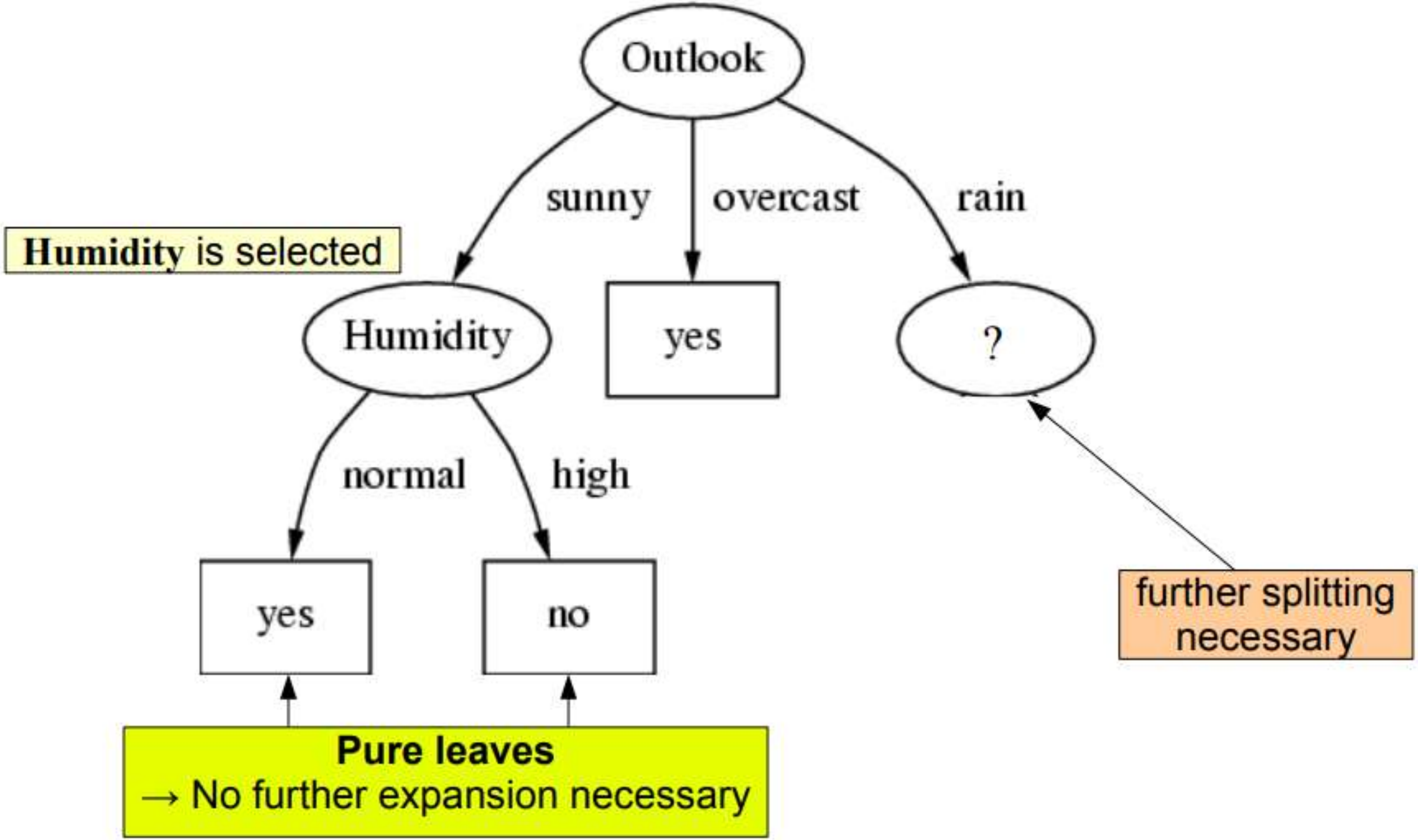
Recurse and repeat Steps 1-6





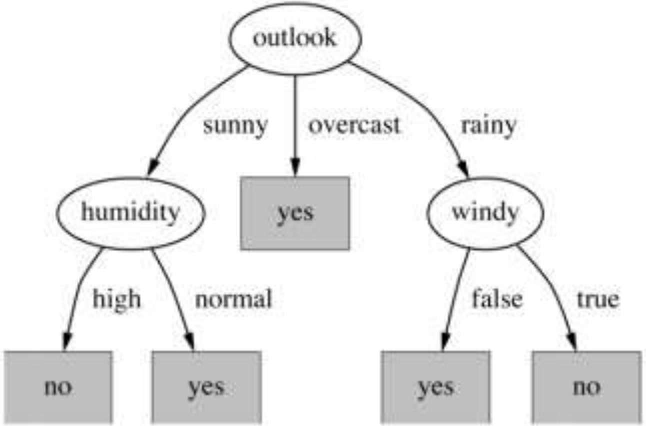
$$\begin{aligned} \text{Gain}(\mathit{Temperature}) &= 0.571 \text{ bits} \\ \text{Gain}(\mathit{Humidity}) &= 0.971 \text{ bits} \\ \text{Gain}(\mathit{Windy}) &= 0.020 \text{ bits} \end{aligned}$$

Humidity is selected


Outlook
sunny
overcast
rain
Humidity is selected
Humidity
yes
?
normal
high
yes
no
further splitting necessary
Pure leaves
→ No further expansion necessary

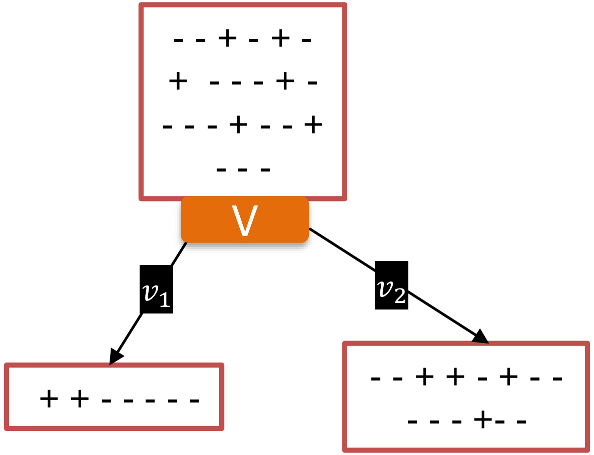
Final Decision Tree

Day	Temperature	Outlook	Humidity	Windy	Play Golf?
07-05	hot	sunny	high	false	no
07-06	hot	sunny	high	true	no
07-07	hot	overcast	high	false	yes
07-09	cool	rain	normal	false	yes
07-10	cool	overcast	normal	true	yes
07-12	mild	sunny	high	false	no
07-14	cool	sunny	normal	false	yes
07-15	mild	rain	normal	false	yes
07-20	mild	sunny	normal	true	yes
07-21	mild	overcast	high	true	yes
07-22	hot	overcast	normal	false	yes
07-23	mild	rain	high	true	no
07-26	cool	rain	normal	true	no
07-30	mild	rain	high	false	yes



Properties of an impurity measure

- Class labels: Binary {+1,-1}
- q



An ***impurity measure*** is a function $i(V)$ such that

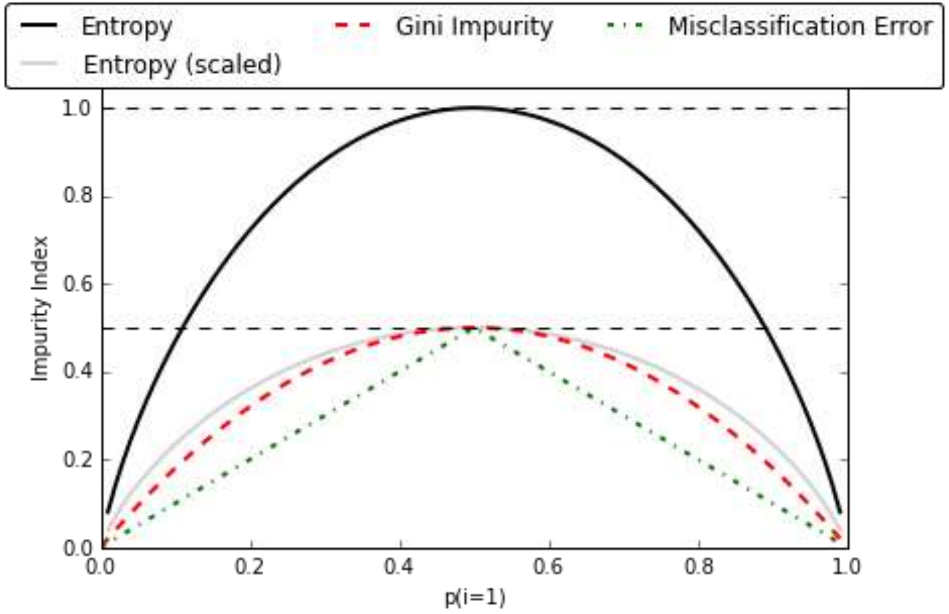
- $i(V) \geq 0$, with $i(V) = 0$ iff V consists of a single class
- a larger value of $i(V)$ indicates that the distribution defined by $(q, (1-q))$ is closer to the uniform distribution

Impurity function: candidates

Entropy: $i(V) = -\big(q \log q + (1-q)\log(1-q)\big)$

Gini index: $i(V) = 2q(1-q)$

Misclassification rate: $i(V) = \min(q, 1-q)$



References and Reading

- Cool demo: http://www.r2d3.us/visual-intro-to-machine-learning-part-1/
- Entropy
  - https://towardsdatascience.com/entropy-how-decision-trees-make-decisions-2946b9c18c8
  - https://plus.maths.org/content/information-surprise
  - In decision trees: https://bricaud.github.io/personal-blog/entropy-in-decision-trees/

- Textbook References
  - [TM] Machine Learning by Tom Mitchell (3.1 – 3.5, 3.7 – 3.8)
  - [PRML] Pattern Recognition and Machine Learning by Chris Bishop (1.2 (intro), 1.6)
  - [DHS] Duda and Hart (8.1 – 8.4)

- Code
  - https://scikit-learn.org/stable/modules/tree.html
  - https://scikit-learn.org/stable/auto_examples/tree/plot_unveil_tree_structure.html