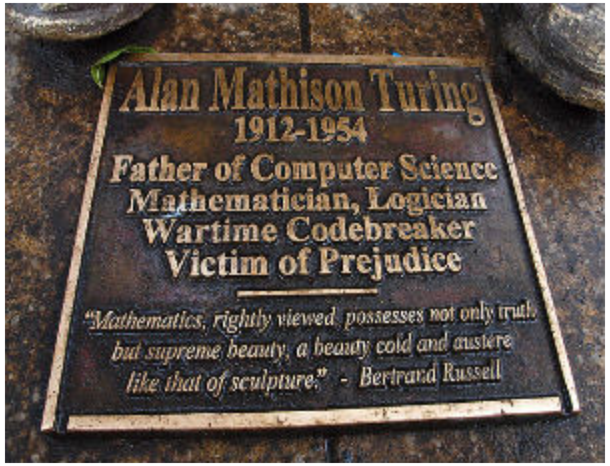
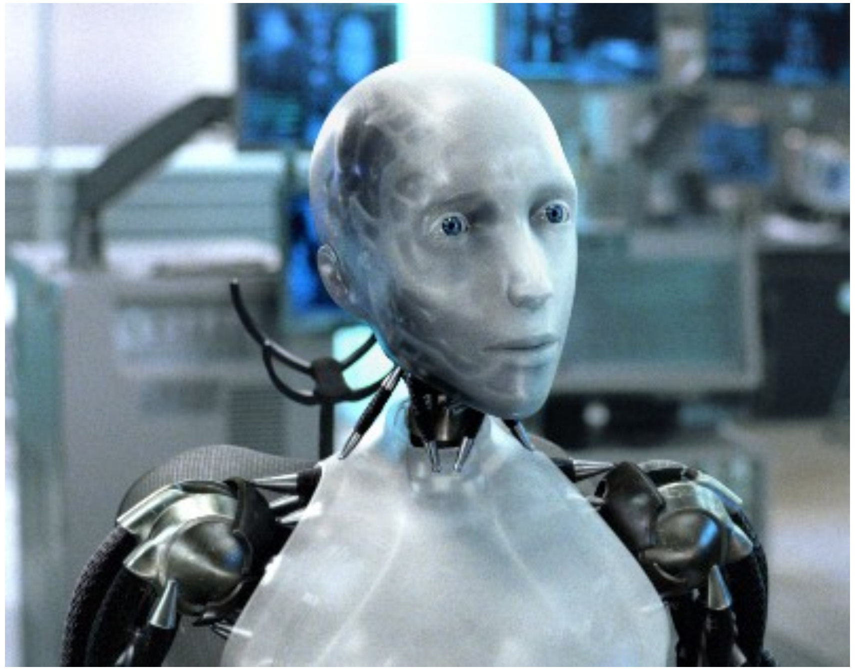
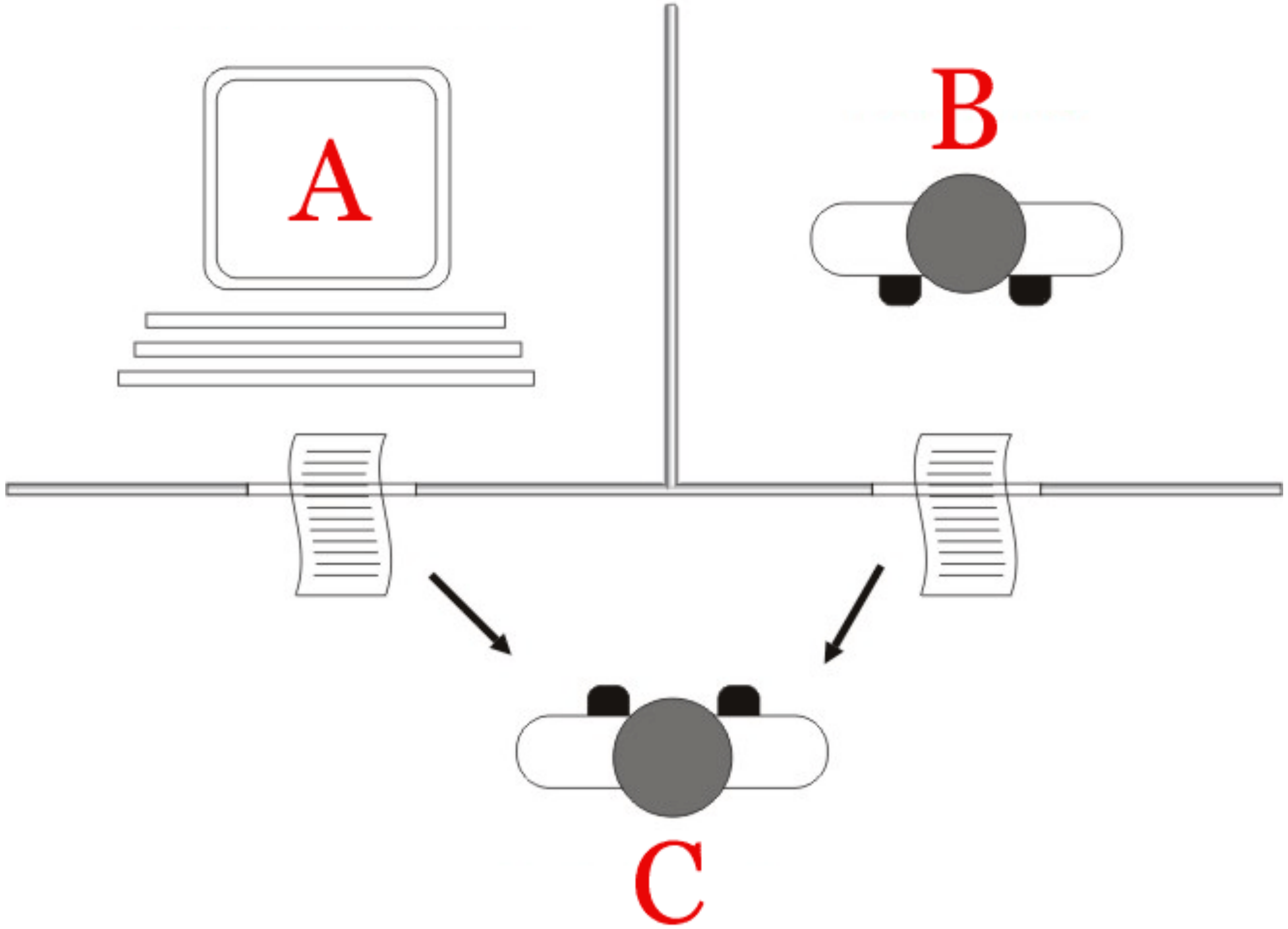
AI and Machine Learning





The Turing Test



Turing attack

How can we show a machine is Intelligent? Let A = machine. Let C = Intelligent. Let B = someone that "we" claim is intelligent. How can we show A = C? Hmm. It's subjective? Well most (normal) say B = C. So if can we show that A = B, then we can show that A = C!

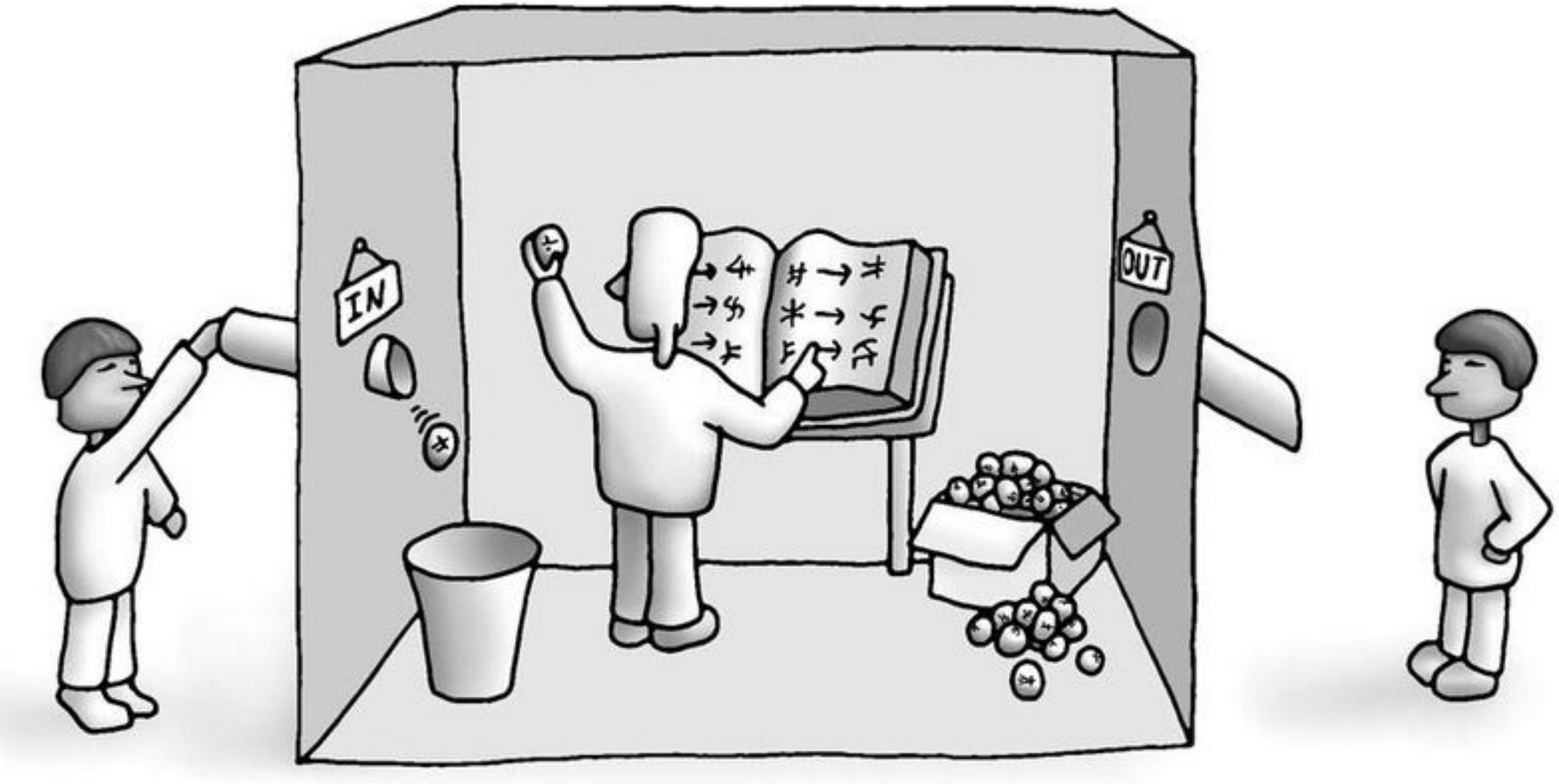
Turing attack

So subjective criteria when normal can be reasoned with transitivity!

E.g., let C = Greatest basketball player ever. Most "normal" say Jordan. Let Jordan = B. Now is Lebron = the greatest ever? Let Lebron = A.

Well we have to show that A = C, or A = B. Let's assume A = B. Do we wind up in a contradiction? If so A!=B. If not A = B and A = C. Is Lebron the greatest?

Chinese Room Argument



Chinese Room Argument

Searle defines:
Strong AI = Understanding
Weak AI = pattern recognition and mappings

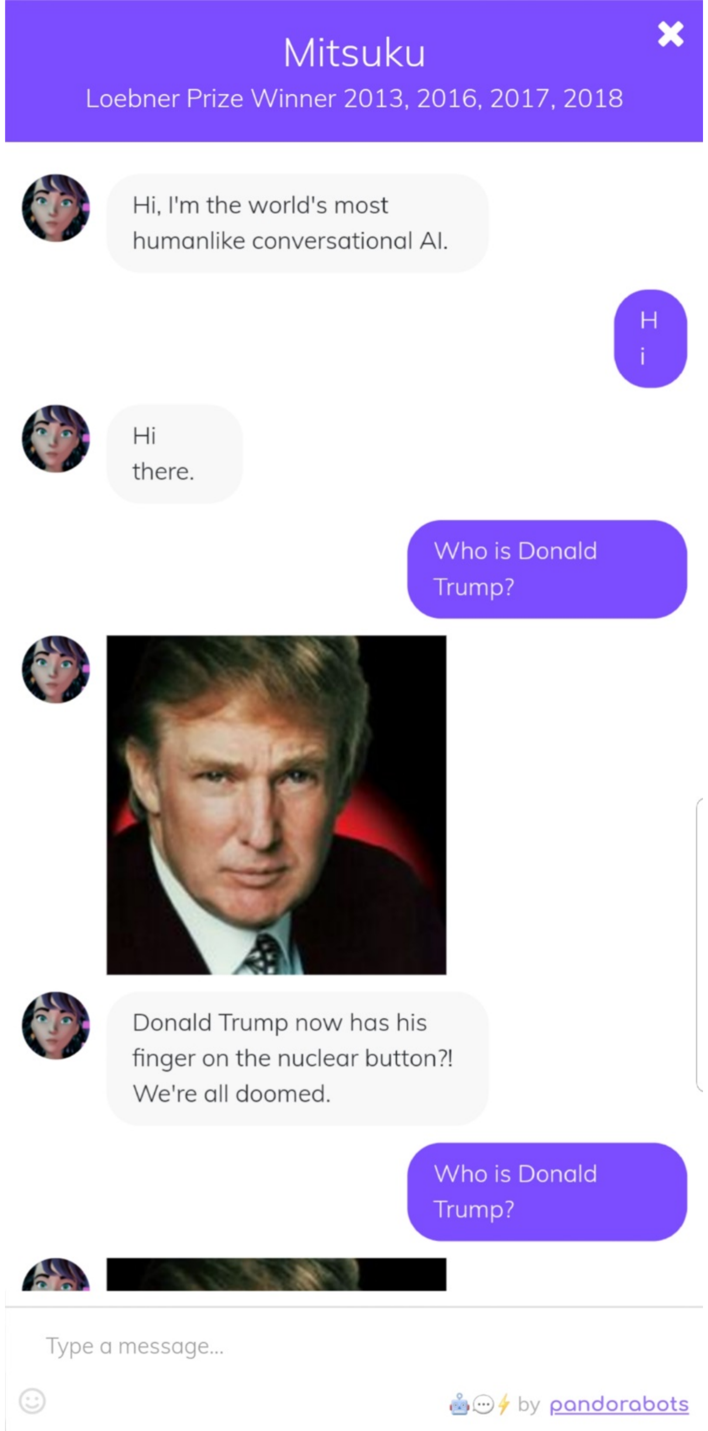
What do you think?

Loebner prize Mitsuka

The Loebner Prize is an annual competition in artificial intelligence that awards prizes to the computer programs considered by the judges to be the most human-like. The format of the competition is that of a standard Turing test. No one has ever won the silver or gold medal!

Mitsuko

Loebner bronze winner in 2013, 2016, 2017, 2018



Unsupervised vs. Supervised Learning

Supervised learning is aided by training data and human correction. Here's some training data. Learn the patterns. Make your best guess at what the patterns are. We'll feed you test data to figure out if you've understood it. If you stray of course we'll correct you and retrain. Examples include Decision Trees and Neural Networks.

Unsupervised learning is uncorrected and runs on data. It can't classify things "yet". But is very good at clustering and anomaly detection.

Decision Tree Learning

Task:

– Given: collection of examples (x, f(x))

– Return: a function *h* (*hypothesis*) that approximates *f*

– *h is a decision tree*

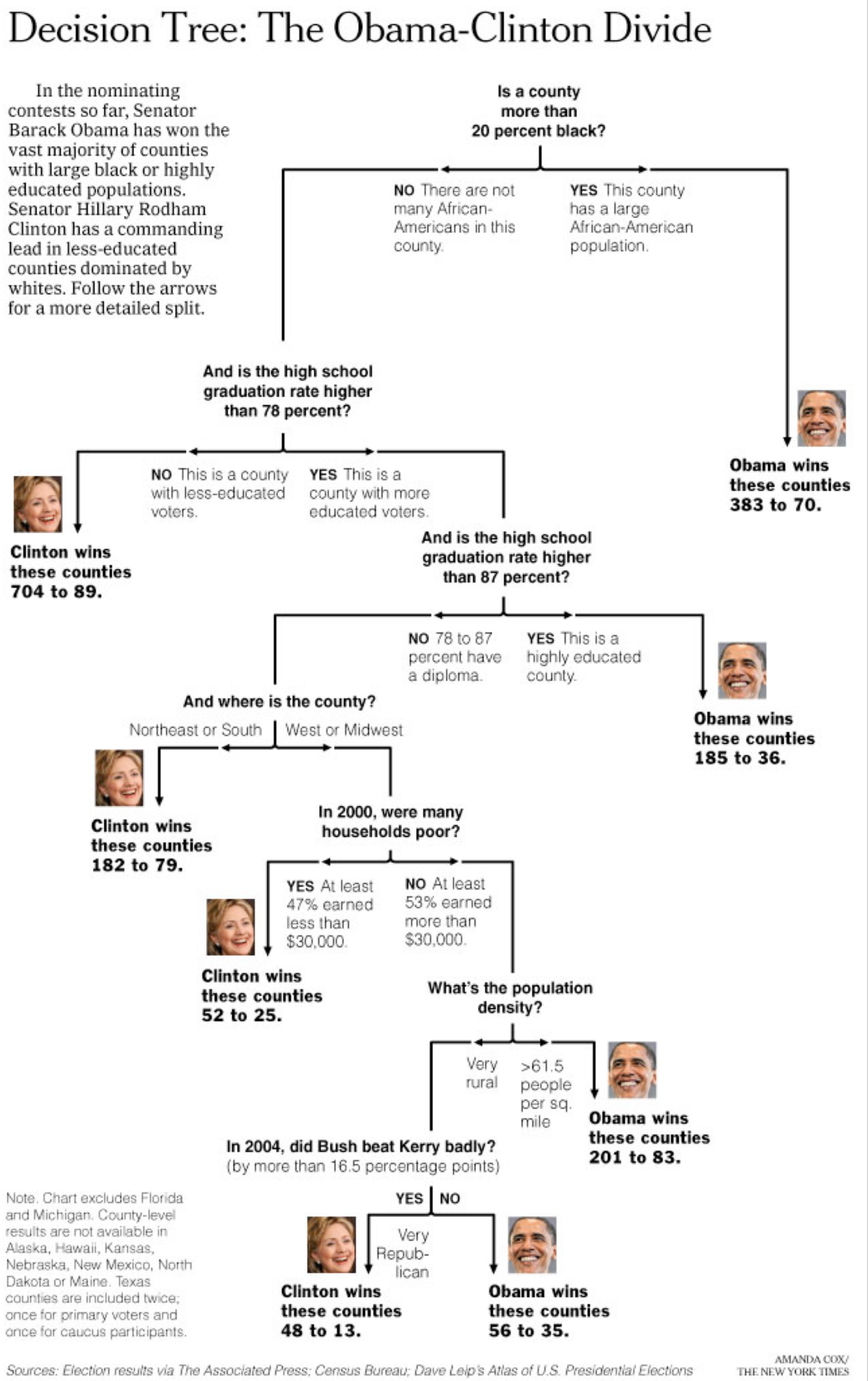
- **Input:** an object or situation described by a set of attributes (or features)
- **Output:** a "decision" – the predicts output value for the input.

- The input attributes and the outputs can be discrete or continuous.

- We will focus on decision trees for Boolean classification:
- each example is classified as positive or negative.
-

Can we learn how counties vote?

New York Times
April 16, 2008



Decision Trees:
a sequence of tests.
Representation very natural for humans.
Style of many "How to" manuals and trouble-shooting procedures.

Decision Tree

- **What is a decision tree?**

- **A tree with two types of nodes:**

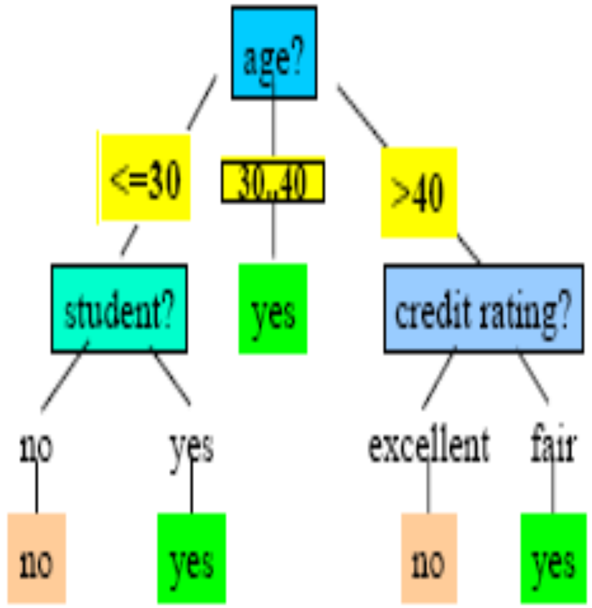
  - **Decision nodes**
  - **Leaf nodes**

- **Decision node:** **Specifies a choice or test of some attribute with 2 or more alternatives;**

- **→ every decision node is part of a path to a leaf node**

- **Leaf node:** **Indicates classification of an example**

- Decision Tree example (is a customer going to buy a computer or not):



Inductive Learning Example

Food (3)	Chat (2)	Fast (2)	Price (3)	Bar (2)	BigTip
great	yes	yes	normal	no	yes
great	no	yes	normal	no	yes
mediocre	yes	no	high	no	no
great	yes	yes	normal	yes	yes

Etc.

Instance Space X: Set of all possible objects described by attributes (often called features).

Target Function f: Mapping from Attributes to Target Feature (often called label) (f is unknown)

Hypothesis Space H: Set of all classification rules h_i we allow.

Training Data D: Set of instances labeled with Target Feature

What is the best Variable (Feature) to use as an indicator of a BigTip?

Entropy & Information Gain

$$InfoGain(feature_d) = Entropy(D) - Entropy(feature_d)$$

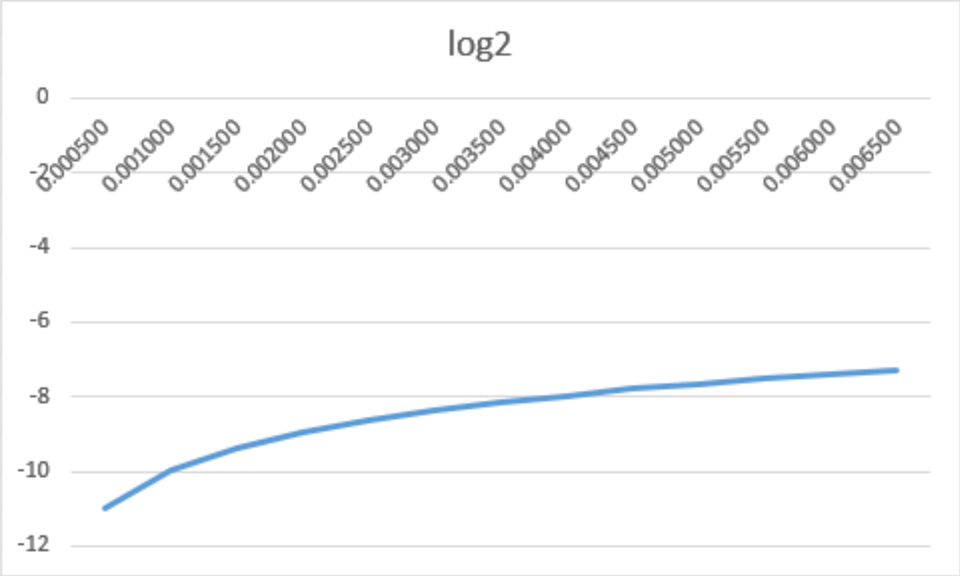
$$InforGain(feature_d, D) = Entropy(D) - \sum_{t \in feature} (\frac{|feature_d = t|}{|D|} * H(feature_d = t))$$

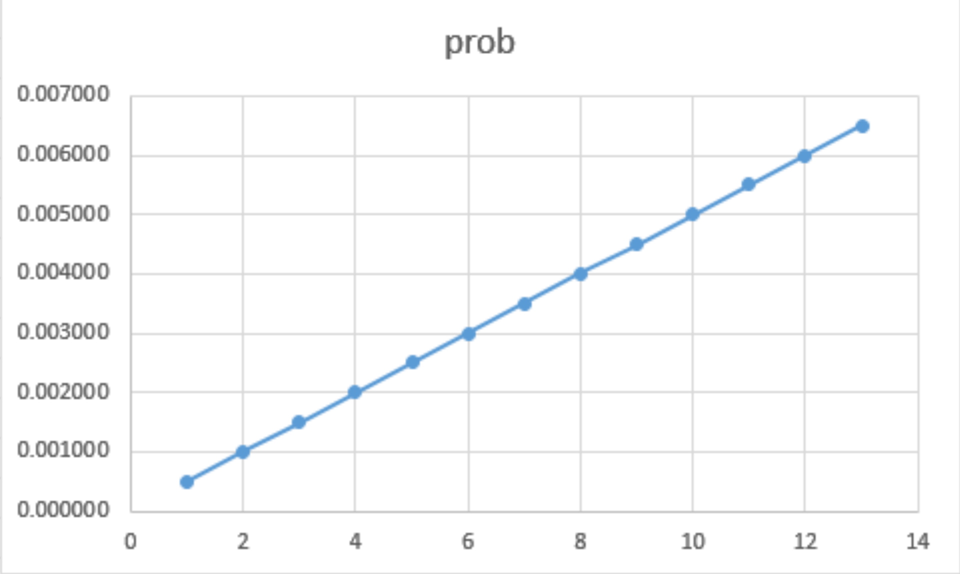
$$=$$

$$Entropy(D) - \sum_{t \in feature} (\frac{|feature_d = t|}{|D|} * (- \sum_{k \in target} (P(target = k, feature_d = t) * log_2(P(target = k, feature_d = t))))$$

Obviously ☺
Don’t Panic. Entropy is just a way to measure disorder = uncertainty in data and uncertainty in “data mappings”.

total possible				
2000	occurences	total	prob	log2
	1	2000	0.000500	-10.9658
	2	2000	0.001000	-9.96578
	3	2000	0.001500	-9.38082
	4	2000	0.002000	-8.96578
	5	2000	0.002500	-8.64386
	6	2000	0.003000	-8.38082
	7	2000	0.003500	-8.15843
	8	2000	0.004000	-7.96578
	9	2000	0.004500	-7.79586
	10	2000	0.005000	-7.64386
	11	2000	0.005500	-7.50635
	12	2000	0.006000	-7.38082
	13	2000	0.006500	-7.26534





Excel example

Food	Chat	Fast	Price	Bar	Big tip
great	yes	yes	normal	no	yes
great	no	yes	normal	no	yes
mediocore	yes	no	high	no	no
great	yes	yes	normal	yes	no

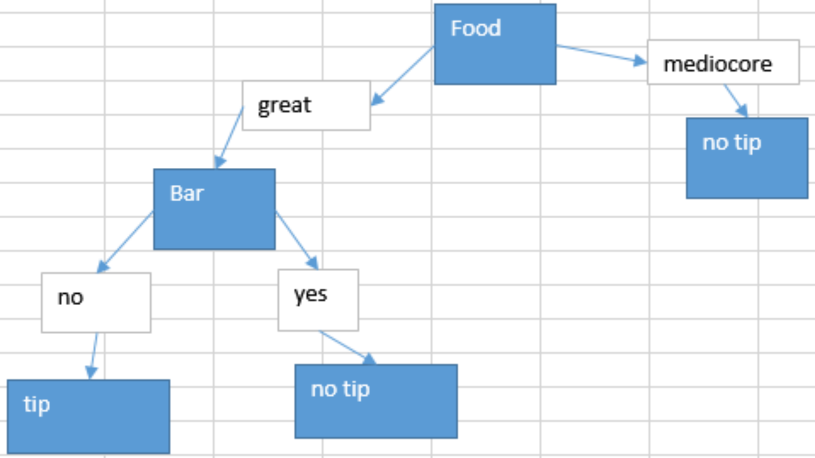
	Food	Chat	Fast	Price	Bar	Big tip
p(a)	0.7500	0.7500	0.7500	0.7500	0.7500	0.5000
p(b)	0.2500	0.2500	0.2500	0.2500	0.2500	0.5000
log2(a)	-0.4150	-0.4150	-0.4150	-0.4150	-0.4150	-1.0000
log2(b)	-2.0000	-2.0000	-2.0000	-2.0000	-2.0000	-1.0000
P(a)*log2(a)	-0.3113	-0.3113	-0.3113	-0.3113	-0.3113	-0.5000
P(b)*log2(b)	-0.5000	-0.5000	-0.5000	-0.5000	-0.5000	-0.5000
Entropy()	**0.8113**	**0.8113**	**0.8113**	**0.8113**	**0.8113**	**1.0000**
mult	0.1875	0.1875	0.1875	0.1875	0.1875	0.2500
entropy a (mappings)	0.9183	0.9183	0.9183	0.9183	0.9183	
entropy b (mappings)	**0.0000**	**0.0000**	**0.0000**	**0.0000**	**0.0000**	
Info Gain	0.3113	0.3113	0.3113	0.3113	0.3113	

entropy of 1 = chaos!

could pick any

P(yes)	0.5
p(no)	0.5
log2(yes)	-1
log2(no)	-1
P(yes)*log2(yes)	-0.5
P(no)*log2(no)	-0.5
Entropy(big tip)	**1**

Because it's 50/50 it could go either way = randomness = entropy



on great

reduce set to great only values including Big tip

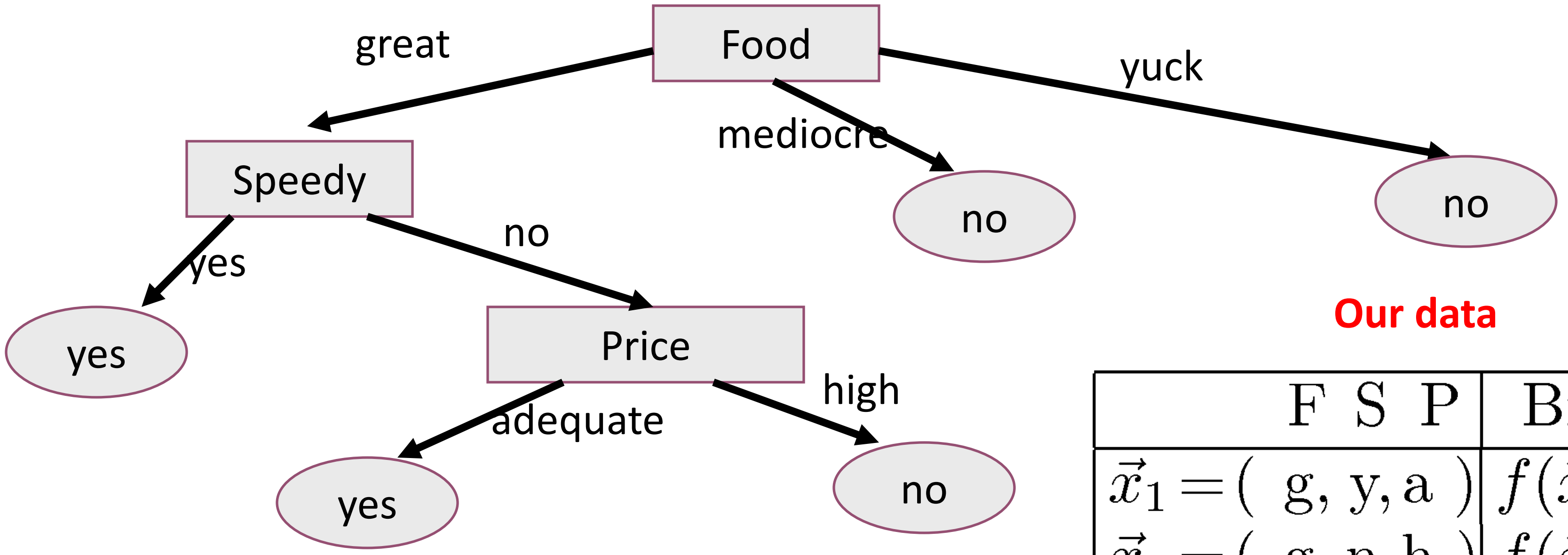
Food	Chat	Fast	Price	Bar	Big tip
great	yes	yes	normal	no	yes
great	no	yes	normal	no	yes
great	yes	yes	normal	yes	no

	Food	Chat	Fast	Price	Bar	Big tip
p(a)		0.6667	1.0000	1.0000	0.6667	0.6667
p(b)		0.3333	0.0000	0.0000	0.3333	0.3333
log2(a)		-0.5850	0.0000	0.0000	-0.5850	-0.5850
log2(b)		-1.5850	0.0000	0.0000	-1.5850	-1.5850
P(a)*log2(a)		-0.3900	0.0000	0.0000	-0.3900	-0.3900
P(b)*log2(b)		-0.5283	0.0000	0.0000	-0.5283	-0.5283
Entropy()		**0.9183**	**0.0000**	**0.0000**	**0.9183**	**0.9183**
mult		0.2222	0.0000	0.0000	0.2222	0.2222
entropy a (mappings)		1.0000	0.9183	0.9183	0.0000	
entropy b (mappings)		**0.0000**	**0.0000**	**0.0000**	**0.0000**	
Info Gain		0.2516	0.0000	0.0000	0.9183	

sub tree recalculation

0 due to new H(Big tip)

Decision Tree Example: "BigTip"



Our data

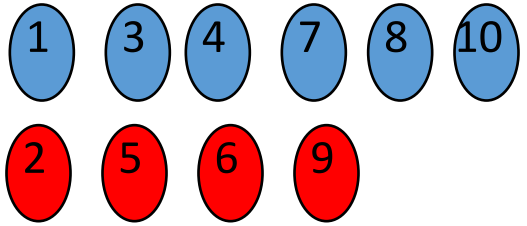
F S P	BigTip
$\vec{x}_1 = (\text{ g, y, a })$	$f(\vec{x}_1) = 1$
$\vec{x}_2 = (\text{ g, n, h })$	$f(\vec{x}_2) = 0$
$\vec{x}_3 = (\text{ g, y, h })$	$f(\vec{x}_3) = 1$
$\vec{x}_4 = (\text{ g, n, a })$	$f(\vec{x}_4) = 1$
$\vec{x}_5 = (\text{ m, y, a })$	$f(\vec{x}_5) = 0$
$\vec{x}_6 = (\text{ y, y, a })$	$f(\vec{x}_6) = 0$
$\vec{x}_7 = (\text{ g, y, a })$	$f(\vec{x}_7) = 1$
$\vec{x}_8 = (\text{ g, y, h })$	$f(\vec{x}_8) = 1$
$\vec{x}_9 = (\text{ m, y, a })$	$f(\vec{x}_9) = 0$
$\vec{x}_{10} = (\text{ g, y, a })$	$f(\vec{x}_{10}) = 1$

Is the decision tree we learned consistent?

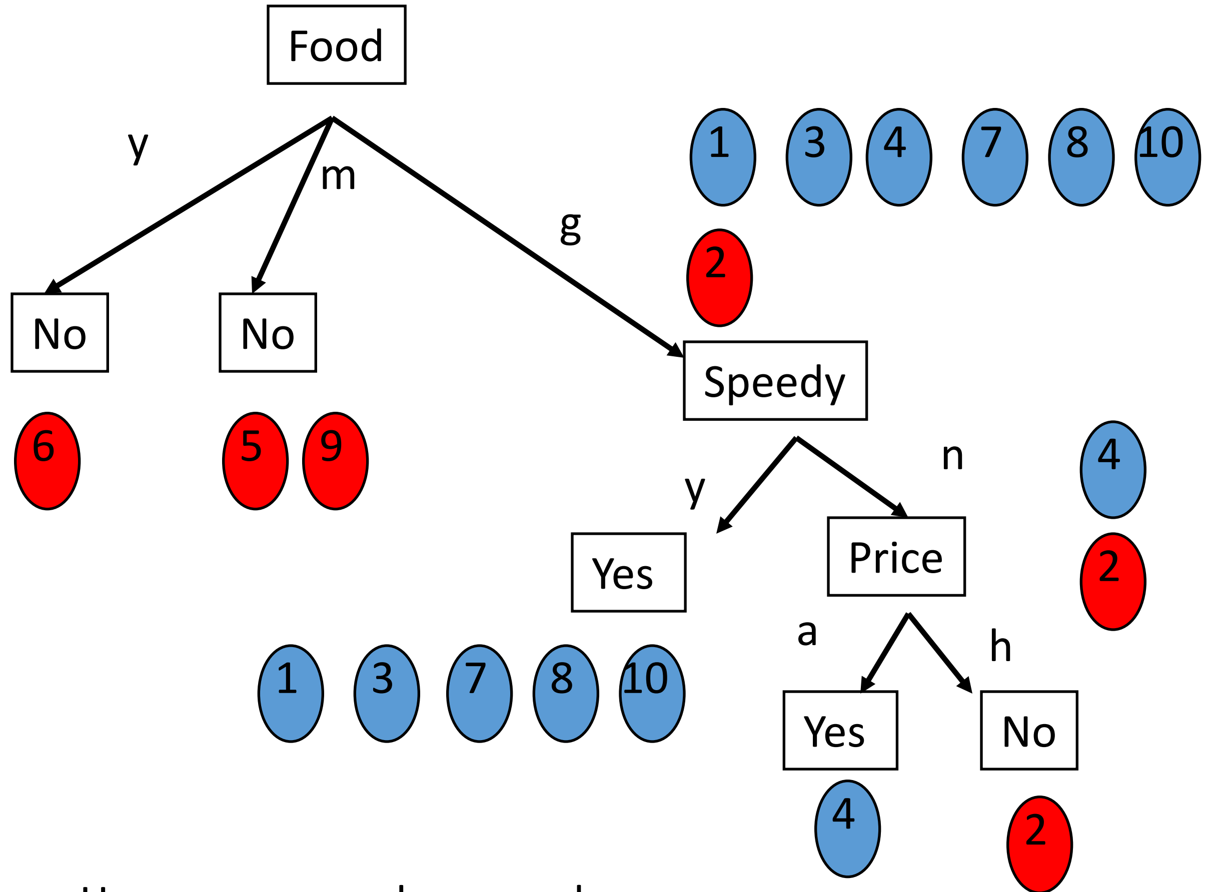
Yes, it agrees with all the examples!

Data: Not all 2x2x3 = 12 tuples
Also, some repeats! These are literally "observations."

Top-Down Induction of Decision Tree: Big Tip Example



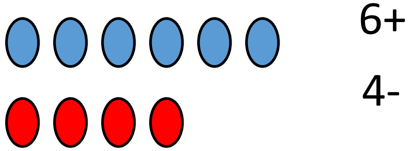
Node "done" when uniform label or "no further uncertainty."



How many + and - examples per subclass, starting with y?

Let's consider next the attribute Speedy

10 examples:

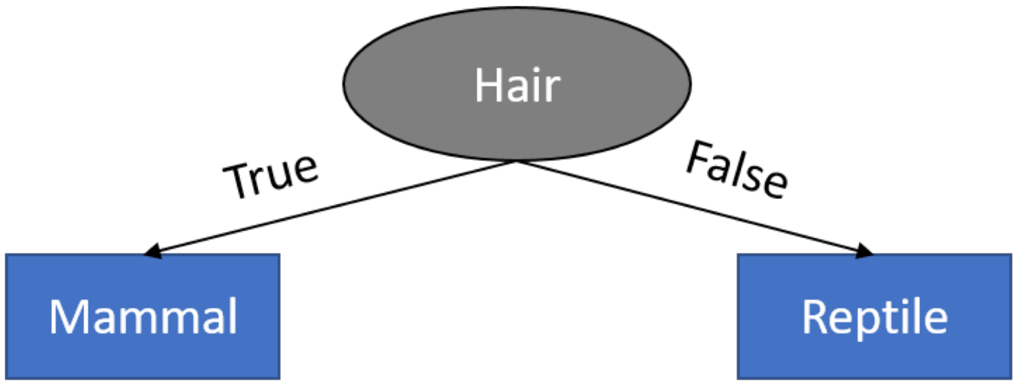


F S P	BigTip
$\vec{x}_1 = (\text{ g, y, a })$	$f(\vec{x}_1) = 1$
$\vec{x}_2 = (\text{ g, n, h })$	$f(\vec{x}_2) = 0$
$\vec{x}_3 = (\text{ g, y, h })$	$f(\vec{x}_3) = 1$
$\vec{x}_4 = (\text{ g, n, a })$	$f(\vec{x}_4) = 1$
$\vec{x}_5 = (\text{ m, y, a })$	$f(\vec{x}_5) = 0$
$\vec{x}_6 = (\text{ y, y, a })$	$f(\vec{x}_6) = 0$
$\vec{x}_7 = (\text{ g, y, a })$	$f(\vec{x}_7) = 1$
$\vec{x}_8 = (\text{ g, y, h })$	$f(\vec{x}_8) = 1$
$\vec{x}_9 = (\text{ m, y, a })$	$f(\vec{x}_9) = 0$
$\vec{x}_{10} = (\text{ g, y, a })$	$f(\vec{x}_{10}) = 1$

An example

	toothed	hair	breathes	legs	species
0	True	True	True	True	Mammal
1	True	True	True	True	Mammal
2	True	False	True	False	Reptile
3	False	True	True	True	Mammal
4	True	True	True	True	Mammal
5	True	True	True	True	Mammal
6	True	False	False	False	Reptile
7	True	False	True	False	Reptile
8	True	True	True	True	Mammal
9	False	False	True	True	Reptile

An example



	toothed	hair	breathes	legs	species
0	True	True	True	True	Mammal
1	True	True	True	True	Mammal
3	False	True	True	True	Mammal
4	True	True	True	True	Mammal
5	True	True	True	True	Mammal
8	True	True	True	True	Mammal

	toothed	hair	breathes	legs	species
2	True	False	True	False	Reptile
6	True	False	False	False	Reptile
7	True	False	True	False	Reptile
9	False	False	True	True	Reptile

An example

	toothed	breathes	legs	species
0	True	True	True	Mammal
1	True	True	True	Mammal
2	True	True	False	Reptile
3	False	True	True	Mammal
4	True	True	True	Mammal
5	True	True	True	Mammal
6	True	False	False	Reptile
7	True	True	False	Reptile
8	True	True	True	Mammal
9	False	True	True	Reptile

toothed == True

	toothed	breathes	legs	species
0	True	True	True	Mammal
1	True	True	True	Mammal
2	True	True	False	Reptile
4	True	True	True	Mammal
5	True	True	True	Mammal
6	True	False	False	Reptile
7	True	True	False	Reptile
8	True	True	True	Mammal

toothed == False

	toothed	breathes	legs	species
3	False	True	True	Mammal
9	False	True	True	Reptile

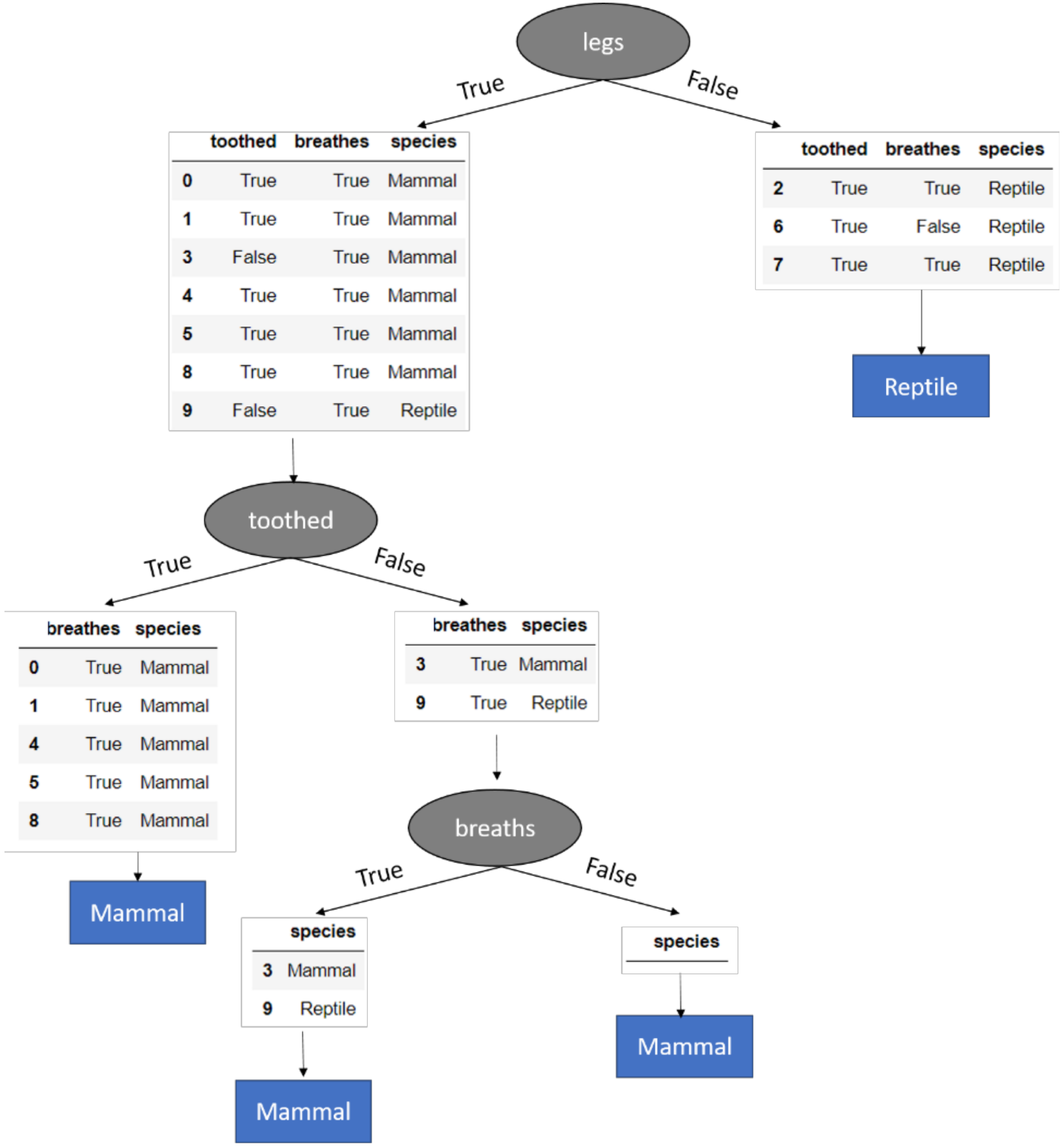
After computing the IG of feature *toothed* do this for features *breathes* and *legs*

1. Calculate the entropy for toothed == True

2. Calculate the entropy for toothed == False

3. Sum up the entropies of 1. and 2.

4. Subtract this sum from the whole datasets entropy → InfoGain


legs
True
False
toothed breathes species
0 True True Mammal
1 True True Mammal
3 False True Mammal
4 True True Mammal
5 True True Mammal
8 True True Mammal
9 False True Reptile
toothed breathes species
2 True True Reptile
6 True False Reptile
7 True True Reptile
Reptile
toothed
True
False
breathes species
0 True Mammal
1 True Mammal
4 True Mammal
5 True Mammal
8 True Mammal
Mammal
breathes species
3 True Mammal
9 True Reptile
breaths
True
False
species
3 Mammal
9 Reptile
Mammal
species
Mammal