Understanding Artificial Intelligence

July 2024

Mitchell Spradlin — Amazon

The Class

- Mix of lectures, reading, and discussions
- Break halfway through class
- Raise hand to ask questions at any time
- Be respectful and inclusive

Background

Going around the room:

- What is your name?
- What school do you go to?
- What grade are you going in to?
- What are you most hoping to learn this week?

Overview

Day 1 The Landscape, Decision Trees

Day 2 Social Media Algorithms

Day 3 Fraud Detection

Day 4 Chatbots

Day 5 Deepfakes, the AI Control Problem

The AI Landscape

Intelligence The ability to perceive or infer information.

Artificial Intelligence (AI) Intelligence exhibited by machines.

The above definition is very broad. Most people mean a computer system that can make decisions that change depending on the circumstances.

History

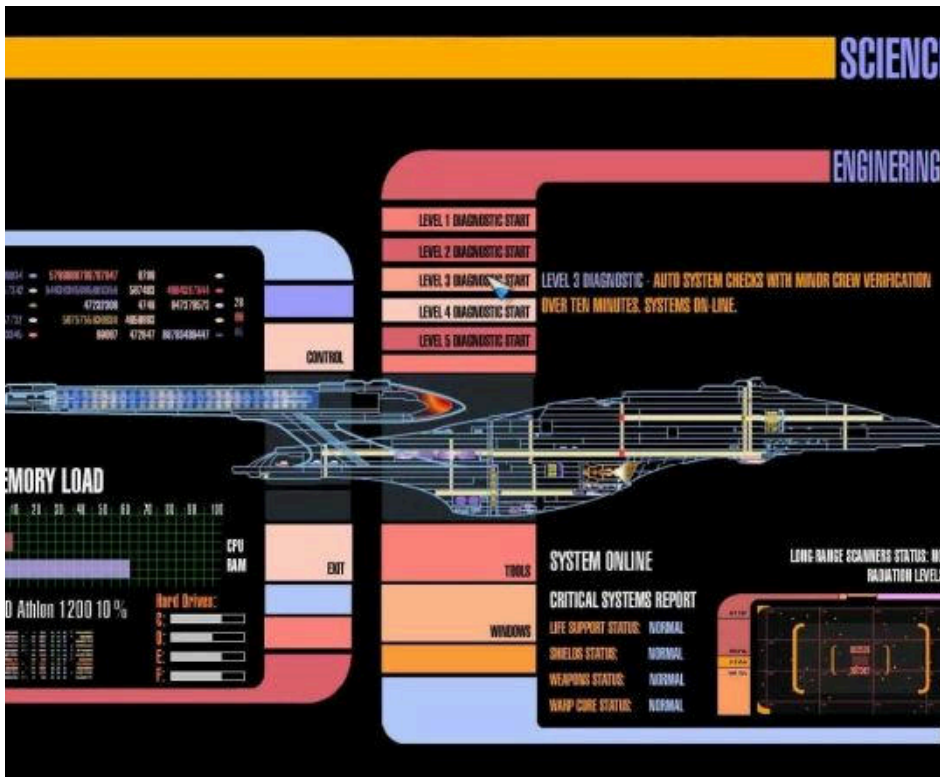
The idea of intelligent artificial beings has been around for a long time.

Figure 1: Talos, giant bronze protector of Crete, c. 300 BC

History

- AI assumes that the process of human thought can be mechanized
- Formal reasoning developed with a long history:
  - Aristotle — Logic
  - al-Khwārizmī — Algebra and algorithms
  - Gödel — Logic, incompleteness proof
  - Many, many, more all over the world

Deep Blue



In 1985, AI was largely the stuff of science fiction

Deep Blue

In 1985, Dr. Feng-Hsiung Hsu began developing a chess supercomputer while a doctoral student at Carnegie Mellon.

In won a computer chess championship in 1987.

Deep Blue

In 1988, he developed Deep Thought. He joined IBM Research after getting his doctorate in 1989.

Deep Thought played Garry Kasparov in 1989 and lost badly. It was then renamed Deep Blue.

Deep Blue

Between 1989 and 1996, Deep Blue was developed using highly-tuned algorithms, including decision trees, and custom chips were created to run the algorithms quickly.

Deep Blue

Deep Blue played Garry Kasparov in Feburary of 1996. It made history by being the first computer to win a game against the reigning world champion, but Kasparov won 4-2.

Garry Kasparov

- Youngest world champion ever (22) in 1985
- Was ranked #1 for more than 25 years (!)
- Retired in 2007, pursued activism

Deep Blue

Deep Blue was upgraded, rematch in May 1997.

- First game move 44 bug, Kasparov wins
- Second game Blue wins, cheating accusation
- Third game esoteric opening, draw

Deep Blue

- Fourth game draw, bad time management by Kasparov
- Fifth game draw, Kasparov missed a win
- Sixth game, dubious opening by Kasparov, Blue wins

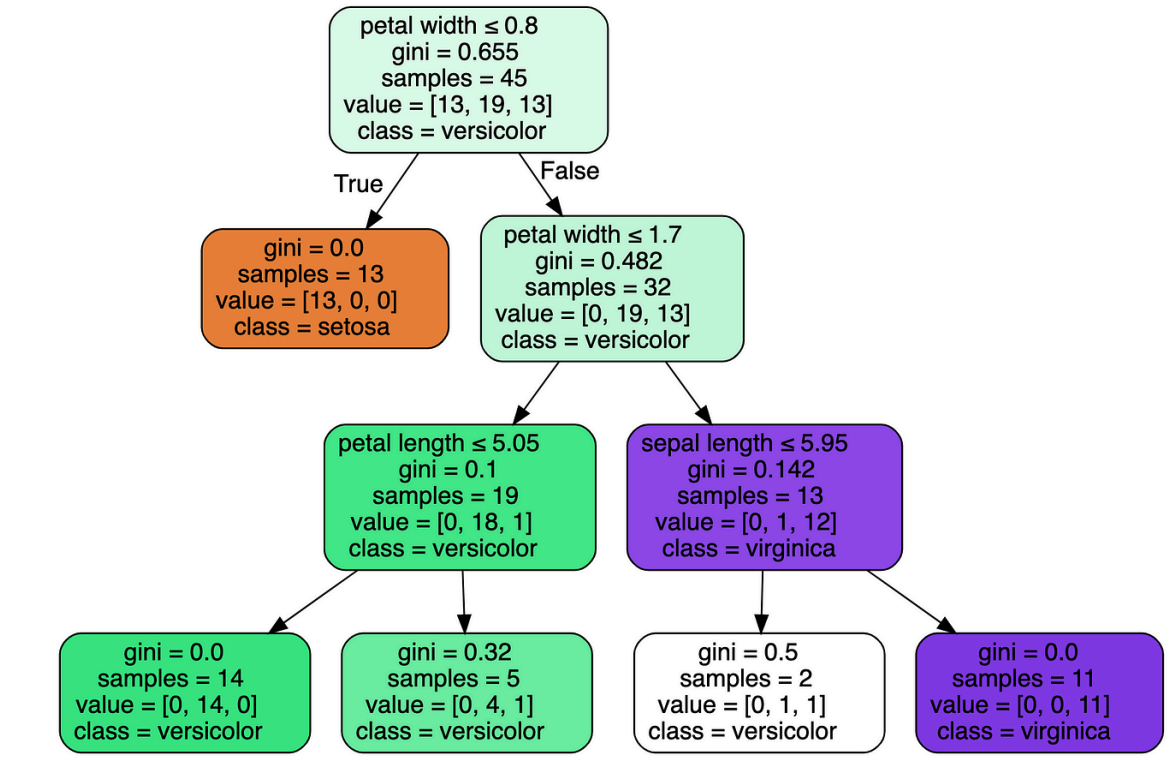
Group Exercise

Split into groups and discuss:

- What are three examples of AI you have encountered in your life?
- How can AI be used as a useful tool?
- Are there any risks associated with AI?

We will then have a class discussion.

Decision Trees



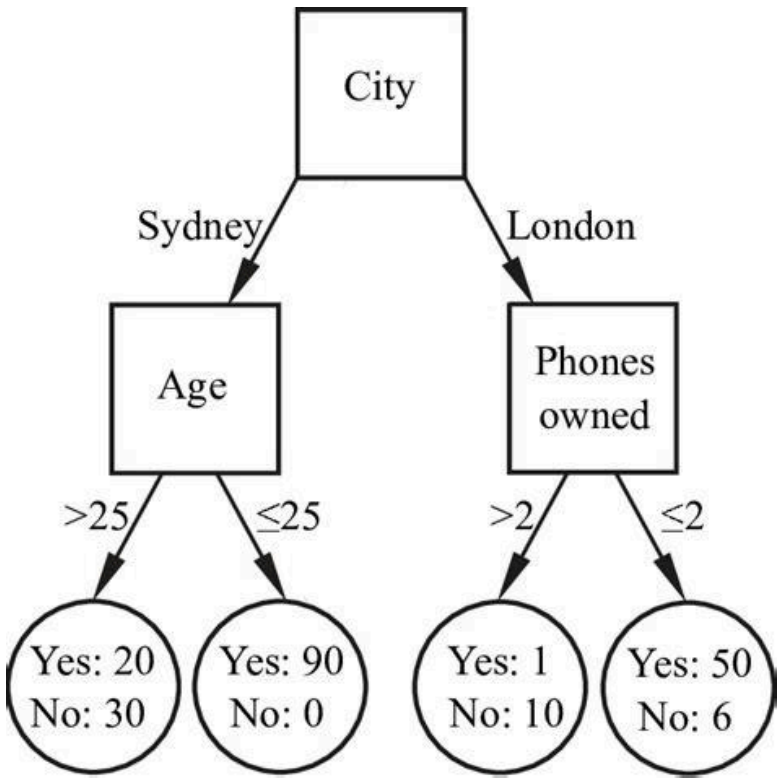
Uses

- Commonly used in **expert systems**
- Is a **decision support** model
- Can encode policies and best practices

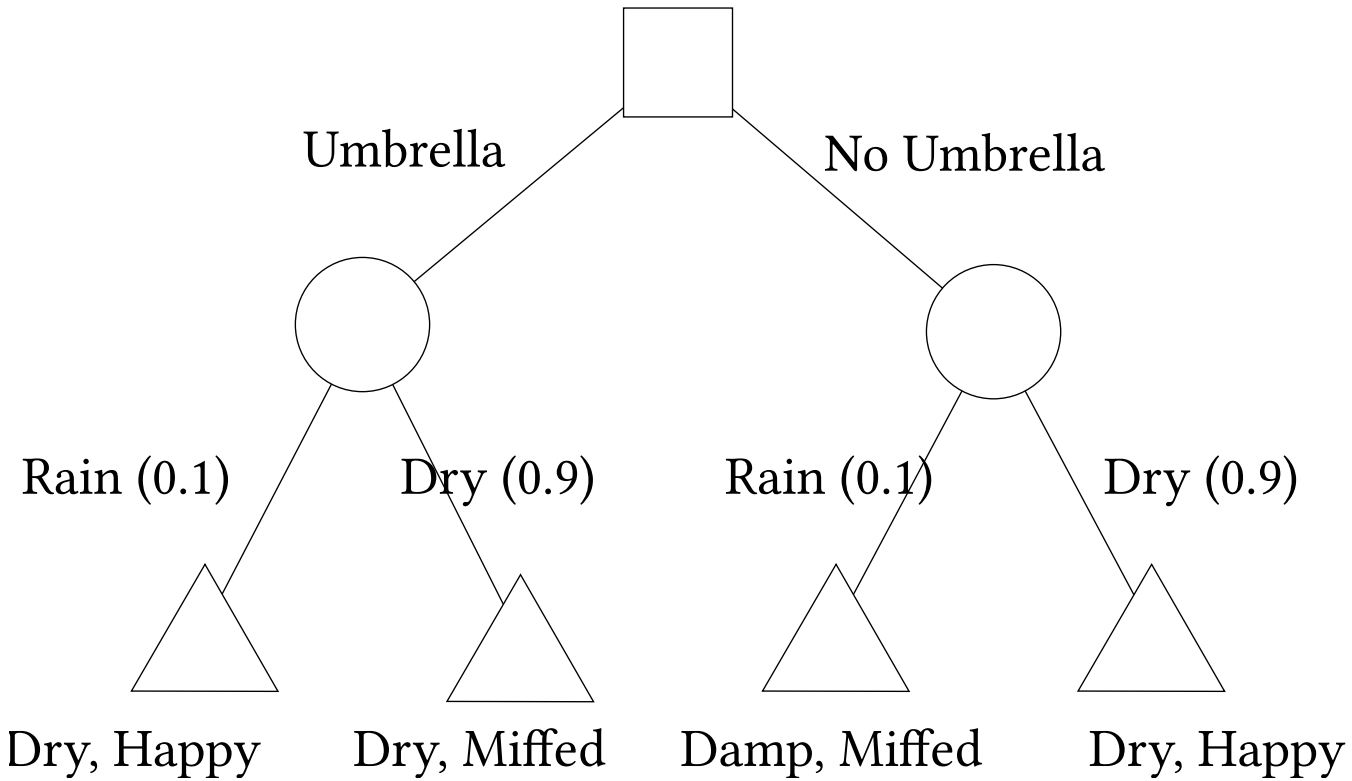
Example of decision trees in action:

https://akinator.com

Approach



- Start at root
- Squares mean a decision
- Continue until you get to the end



- Circles mean chance, good for modeling outcomes

Exercise: Make a Decision Tree by Hand

Create a decision tree to decide if a given animal is a fish based on the table below.

	Can survive without coming to surface?	**Has flippers?**	**Fish?**
1	Yes	Yes	Yes
2	Yes	Yes	Yes
3	Yes	No	No
4	No	Yes	No
5	No	Yes	No

Exercise: Make a Decision Tree by Hand

- What was your solution?
- How can you tell if it works as you expect?

Exercise: Make a Decision Tree for Tic-Tac-Toe

Make a decision tree to play tic-tac-toe. Write down your rules such that another person could follow them.

When we're done, we will play out a game on the board using the tree.

	No surfacing?	Flippers?	Head?	Tentacles?	Fish?
1	Yes	Yes	Yes	No	Yes
2	Yes	Yes	Yes	Yes	Yes
3	Yes	No	No	No	No
4	No	Yes	Yes	No	No
5	No	Yes	No	Yes	No
6	No	No	No	No	No
7	No	Yes	Yes	Yes	No
8	Yes	Yes	No	No	No

Algorithm: ID3

1. Calculate the *entropy* of every attribute a of data set S.
2. Split the set S into subsets using the attribute for which the resulting entropy after splitting is minimized.
3. Make a decision tree node containing that attribute.
4. Recurse on subsets using the remaining attributes.

Entropy is a measure of how disordered data is.

Entropy

In information theory, the "informational value" of some data is related to how surprising the event is. If something highly likely occurs, then the data contains little information. If something unlikely occurs, then there the event is very informative.

- Knowledge of a losing lottery number: Uninformative
- Knowledge of the winning number: Very informative

$$\text{Data point } x = \left(x_{a_1}, x_{a_2}, ..., x_{a_n}\right)$$

$$\text{Data set X} = x_1, ..., x_n$$

$$\text{Probability of event } E_i = p(E_i)$$

$$= \frac{\text{Count of } x \text{ where } x_{a_i} = E_i}{\text{Total count of } X} = |E_i \frac{|}{|} X|$$

$$\text{Information} = I(E_i) = \log_2\left(\frac{1}{p(E_i)}\right)$$

$$= -\log_2(p(E_i)))$$

$$\text{Entropy of } X = \mathrm{H}(X) = \sum_{i=1}^{n} p(E_i) * I(E_i)$$

$$\text{Information Gain of attribute } A = IG(X, A)$$

$$= \mathrm{H}(X) - \mathrm{H}(X|A)$$

$$= \mathrm{H}(X) - \sum_{E_a \in A} |E_a \frac{|}{-} X| * \mathrm{H}(x \in X \mid x_a = E_a)$$

Next we'll work an example.

Entropy: Example

	No surfacing?	**Flippers?**	**Fish?**
1	Yes	Yes	Yes
2	Yes	Yes	Yes
3	Yes	No	No
4	No	Yes	No
5	No	Yes	No

We'll compare information gain for splitting on `No surfacing?` and `Flippers?`.

Splitting on `No surfacing?`:

	No surfacing?	**Flippers?**	**Fish?**
1	Yes	Yes	Yes
2	Yes	Yes	Yes
3	Yes	No	No
4	No	Yes	No
5	No	Yes	No

Overall entropy is: $\mathrm{H}(X_{\text{Fish?}}) = \sum_{i=1}^{n} p(E_i) * I(E_i)$

	No surfacing?	**Flippers?**	**Fish?**
1	Yes	Yes	Yes
2	Yes	Yes	Yes
3	Yes	No	No
4	No	Yes	No
5	No	Yes	No

$$
\begin{aligned}
&= p(\text{Yes}) * I(\text{Yes}) \\
&\qquad + p(\text{No}) * I(\text{No}) \\
&= \frac{2}{5}\left(-\log_2\left(\frac{2}{5}\right)\right) \\
&\qquad + \frac{3}{5}\left(-\log_2\left(\frac{3}{5}\right)\right) \\
&\approx 0.97 \text{ bits of entropy}
\end{aligned}
$$

Entropy of `No surfacing?` = Yes: $\mathrm{H}(X_{\text{Fish?}}|\text{No surfacing?} = \text{Yes})$

	No surfacing?	**Flippers?**	**Fish?**
1	Yes	Yes	Yes
2	Yes	Yes	Yes
3	Yes	No	No
4	No	Yes	No
5	No	Yes	No

$$
\begin{aligned}
&= p(\text{Yes}) * I(\text{Yes}) \\
&\qquad + p(\text{No}) * I(\text{No}) \\
&= \frac{2}{3}\left(-\log_2\left(\frac{2}{3}\right)\right) \\
&\qquad + \frac{1}{3}\left(-\log_2\left(\frac{1}{3}\right)\right) \\
&\approx 0.92 \text{ bits of entropy}
\end{aligned}
$$

Entropy of `No surfacing?` = No: $\mathrm{H}(X_{\text{Fish?}}|\text{No surfacing?} = \text{No})$

	No surfacing?	**Flippers?**	**Fish?**
1	Yes	Yes	Yes
2	Yes	Yes	Yes
3	Yes	No	No
4	No	Yes	No
5	No	Yes	No

$$= \sum_{i=1}^{n} p(E_i) * I(E_i)$$

$$= p(\text{No}) * I(\text{No})$$

$$= 1(-\log_2(1))$$

$$= 0 \text{ bits of entropy}$$

$$
\begin{aligned}
&IG\left(X_{\text{Fish?}}, A_{\text{No surfacing?}}\right) \\
&= \mathrm{H}(X_{\text{Fish?}}) - \mathrm{H}\left(X_{\text{Fish?}} | A_{\text{No surfacing?}}\right) \\
&= \mathrm{H}(X_{\text{Fish?}}) - \\
&\qquad \sum_{E_a \in A_{\text{No surfacing?}}} |E_a \frac{|}{|} X_{\text{Fish?}}|\ \mathrm{H}(X_{\text{Fish?}} \mid \text{No surfacing?} = E_a \\
&= \mathrm{H}(X_{\text{Fish?}}) - |E_{\text{Yes}} \frac{|}{|} X_{\text{Fish?}}|\ \mathrm{H}(\text{Yes}, \text{Yes}, \text{No}) + |E_{\text{No}} \frac{|}{|} X_{\text{Fish?}}|\ \mathrm{H} \\
&\approx 0.97 - \tfrac{3}{5} * 0.92 + \tfrac{2}{5} * 0 = 0.418 \text{ bits}
\end{aligned}
$$

Splitting on `Flipper?`:

	No surfacing?	**Flippers?**	**Fish?**
1	Yes	Yes	Yes
2	Yes	Yes	Yes
3	Yes	No	No
4	No	Yes	No
5	No	Yes	No

Entropy of Yes is 1. Entropy of No is 0. Information gain is $\approx 0.97 - \frac{4}{5} * 1 - \frac{1}{5} * 0 = 0.17$ bits.

Splitting

- Splitting on `No surfacing?` has information gain of 0.42
- Splitting on `Flipper?` has information gain of 0.17

Thus, `ID3` would split on `No surfacing?` first.

Since everything in the No branch is the same classification, those rows are removed from the data set.

After that, there's just `Flippers?` left to split on using the remaining data.

	Flippers?	**Fish?**
1	Yes	Yes
2	Yes	Yes
3	No	No

- This is the last attribute, so no calculations are necessary.
- Because rows get removed, the starting entropy can change attribute-to-attribute.

Other Algorithms

As with most machine learning algorithms, there are alternatives to choose from:

- C4.5 — Successor to ID3, can handle numeric values
- C5.0 — Successor to C4.5, more efficient but proprietary
- CART — Family of algorithms
- MARS — Family of algorithms

Each has different performance and complexity.

Strengths and Limitations

- Simple to interpret
- Valuable as a modeling process
- Generally good performance on large data
- Small changes in the data can render a tree inaccurate
- Less accurate compared to other techniques
- Gets very complicated if many factors are involved
- Prone to **overfitting**—matching the training data so closely it doesn't give good predictions