

Lecture 13: What is AI?/Neural Network Basics

Key Questions

- Why is "AI" an intrinsically problematic term?
- What are some general approaches we can take to design a novel AI system?
- In your own words, describe the components of (or parameters for) a neurode (neural network "neuron") and how it comes to an answer for a given input; draw a diagram if it helps.
- Given the following truth table, can you come up with a neurode that implements the underlying boolean function?

b1	b2	b1 b2
T	T	F
T	F	T
F	T	T
F	F	T

Notes

- "Automated decision making" captures most of what we want out of "AI"
 - We can debate whether these decisions constitute intelligence
- **Opinion:** If strong AI is possible, it might not be ethical to create it. What do you think?
- **Opinion:** Superhuman AI might not be possible, since humans' generalization power comes with inaccuracies and blind spots. What do you think?
- Neural networks specifically are based on the *metaphor* of animal brains
 - The topology of neural networks is usually a stack of layers (emphatically not true of animal brains)
- Activation of a neurode = $f(\sum_i w_i x_i)$, e.g. $\mathbf{w} \cdot \mathbf{x} \geq t$
 - Any function is possible, usually these are differentiable; sigmoid, tanh, ReLU, threshold, ...

Your Questions