03-01-2023

CS051A

INTRO TO COMPUTER SCIENCE WITH TOPICS IN AI

12: Intro to AI and Neural Networks

Alexandra Papoutsaki

she/her/hers

Lectures

Zilong Ye

he/him/his

Labs

Lecture 12: Intro to AI and Neural Networks

- Administrative
- Artificial Intelligence
- Neural Networks

Assignment 6 is out

- Make sure to sign up for 03/20 group presentations on AI + Ethics.
  - Talk to your classmates or use Slack to find 1-2 more students interested in the same topic (must be in the same section).

Course feedback

- Thank you for providing us with feedback!
- For the most part, we are in a good pace and the load outside lectures/labs seems reasonable.
- Assignments and mentor sessions are a highlight: Yay!
  - I communicated that to mentors who are really appreciative of your feedback.
  - Lectures provide the basic elements but only if you put them to practice you will be able to really “own” them.
- Make sure you review material after each lecture:
  - Code is always linked at the end of the slides so you can follow along.
  - Practice problems are linked both at the course website and the end of the slides.
  - While working on the assignment, ask yourself what concepts from the lectures are you using? E.g., in assignment 5, we used dictionaries, in assignment 6, we will use recursion etc.

Lecture 12: Intro to AI and Neural Networks

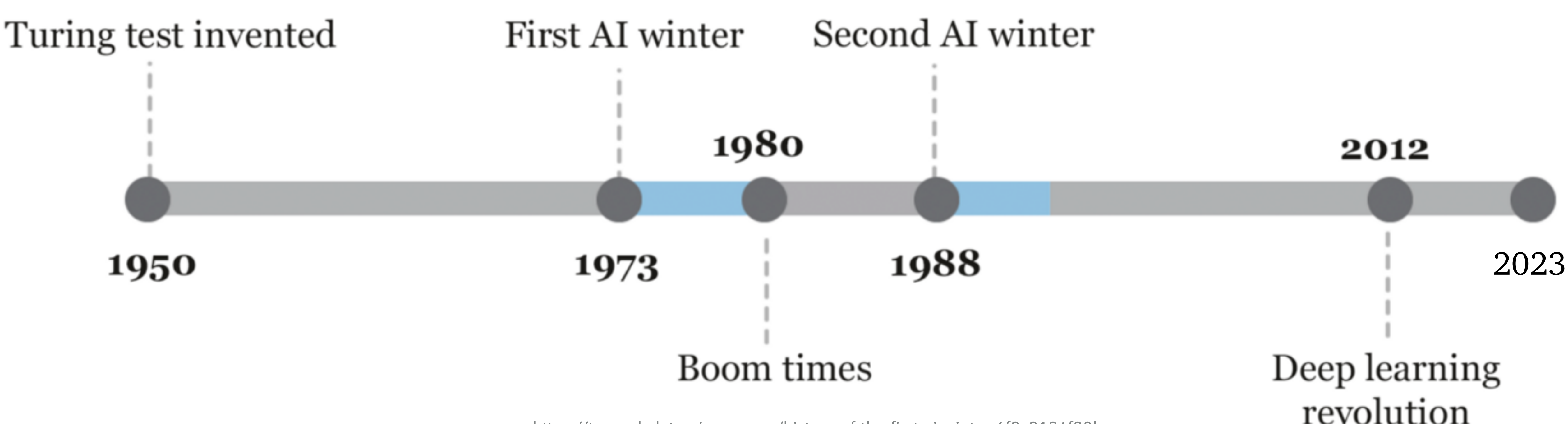
- Administrative
- Artificial Intelligence
- Neural Networks

The field of Artificial Intelligence

- A (huge!) subfield of Computer Science.
- Term coined in mid-1950s by John McCarthy and the field officially began at the Dartmouth Conference.
- But ideas around intelligent machines are old:
  - Storytelling devices in antiquity,
  - Mary Shelley's Frankenstein,
  - Karel Čapek's R.U.R. introduced the word robot.

AI winters followed by boom times

- Excitement around AI has been often followed by stalling in funding and falling short of expectations.
- Since ~2010, we are experiencing a deep learning revolution due to advances in computation supported by hardware and data.



https://towardsdatascience.com/history-of-the-first-ai-winter-6f8c2186f80b

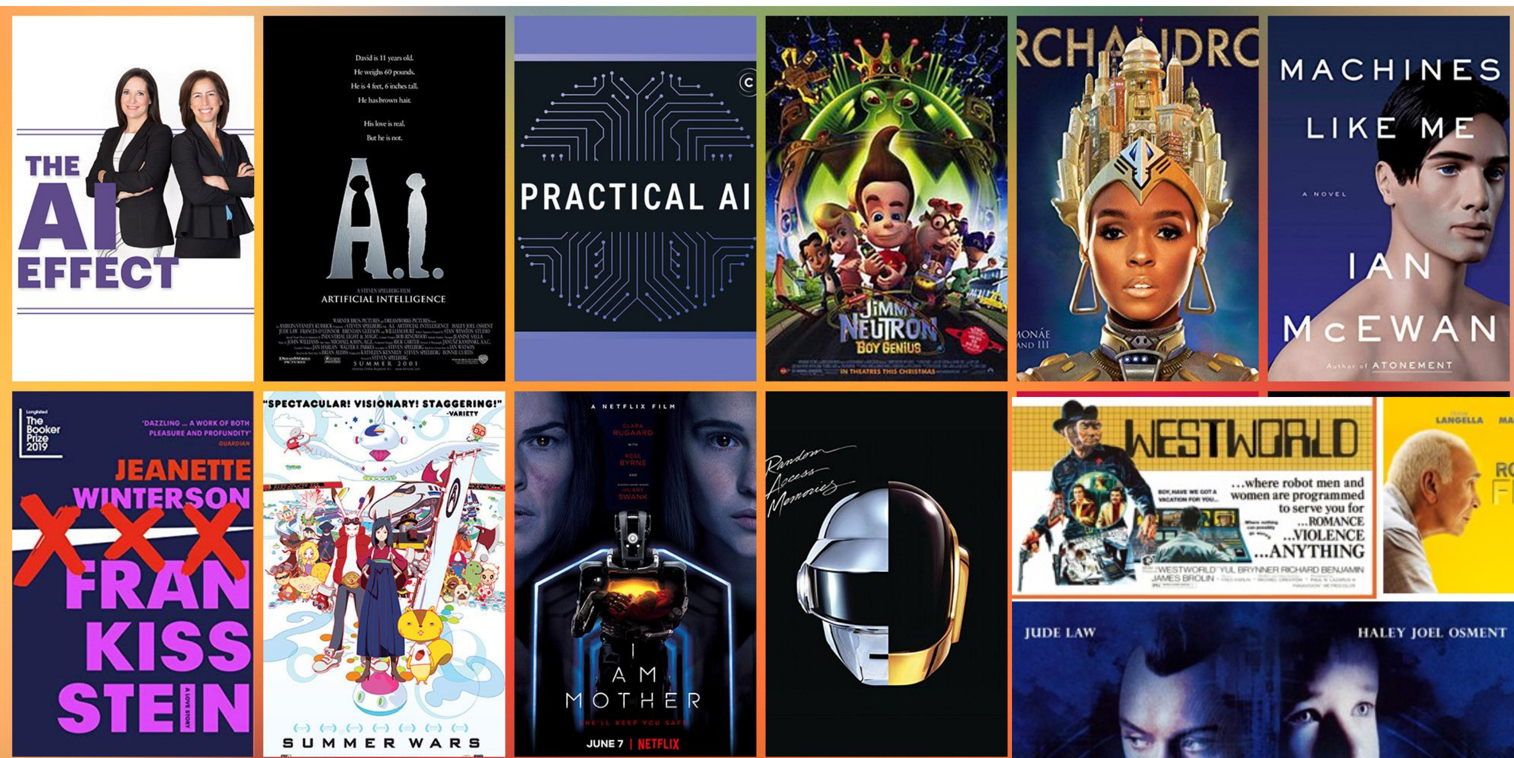
What is Artificial Intelligence?

- *“Can machines think?”*
  - Alan Turing
- *“Every aspect of learning or any other feature of intelligence can be so precisely described that a machine can be made to simulate it”*
  - John McCarthy, Marvin Minsky, Claude Shannon, and Norbert Wiener

Traits of intelligence

- reason, use strategy, solve puzzles, and make judgments under uncertainty,
- represent knowledge, including common sense knowledge,
- plan,
- learn,
- communicate in natural language,
- sense (e.g., see, hear, etc.), and
- act (e.g., move and manipulate objects, change own location to explore, etc.)

Artificial Intelligence in popular media



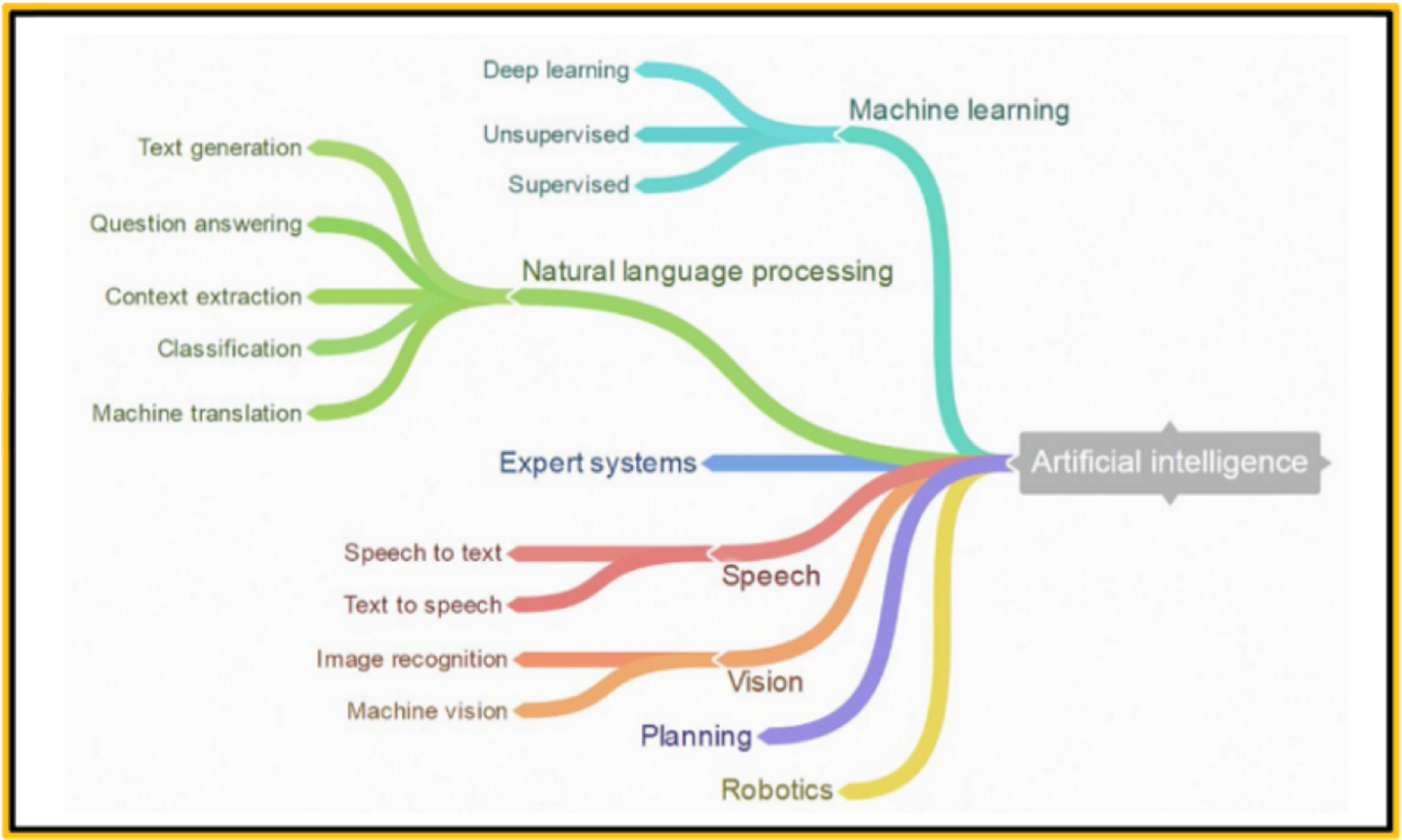
https://www.esquire.com/entertainment/movies/g29653413/best-ai-artificial-intelligence-movies-tv-shows-books-podcasts/

https://www.newworldai.com/top-22-best-artificial-intelligence-and-robotics-movies-of-all-time/

Weak AI, Strong AI, Super AI

- Artificial Narrow Intelligence (ANI) or Weak AI: a machine performs a narrow single task extremely well.
- Artificial General Intelligence (AGI) or Strong/Full AI: a machine performs any intellectual task that a human being can.
    - Can we avoid human-level error and cognitive hazards?
        - Prejudice, change blindness, sloppy reasoning.
    - A fantasy of slavery:
        - The worker who never tires or strikes
        - The “lover” who can’t say “no”
- Artificial Super Intelligence (ASI): a machine far surpasses the brightest and most gifted human minds.

Subfields of AI



https://storage.googleapis.com/cdn.thenewstack.io/media/2018/07/53726660-marconi.png

Challenges

- Perception
  - Perceive the environment via sensors.
- Computer Vision
  - Process visual information.
  - Object identification, face recognition, motion tracking.
- Natural language processing and generalization
  - Speech recognition, language understanding.
  - Language translation, speech generalization, summarization.

Challenges

- Knowledge representation
  - Encode known information.
    - Facts about environment, etc.
- Machine Learning
  - Learn from environment/data (more soon).
- Reasoning/problem solving.
- Robotics

Accomplishments

- Language
  - Capturing spoken language as text is getting pretty okay.
    - Is this the same as understanding spoken language?
  - AI can speak and sing (easier in some languages, interestingly).
    - Mobile: Siri, Ok Google, etc.
    - Home assistants: Alexa, Google Home, etc.
    - Chatbots: chatGPT, Bing vs Google

Accomplishments

- Self-driving cars
  - Driver assisting technologies (breaking, lane drift avoidance, etc) are pretty great already.
  - Good on highways.
  - Okay off-roading.
  - Urban driving very hard.

Accomplishments

- Emotions
  - It’s generally not too hard for humans to read emotions.
  - It’s super hard for computers
    - People have different faces,
    - People emote with their whole body,
    - People don’t always say what they mean.
  - Limited success so far.

Accomplishments

- Automated Reasoning
  - Really good on isolated problems
    - Chess, Go, Starcraft, …
  - Really hard in general

Accomplishments

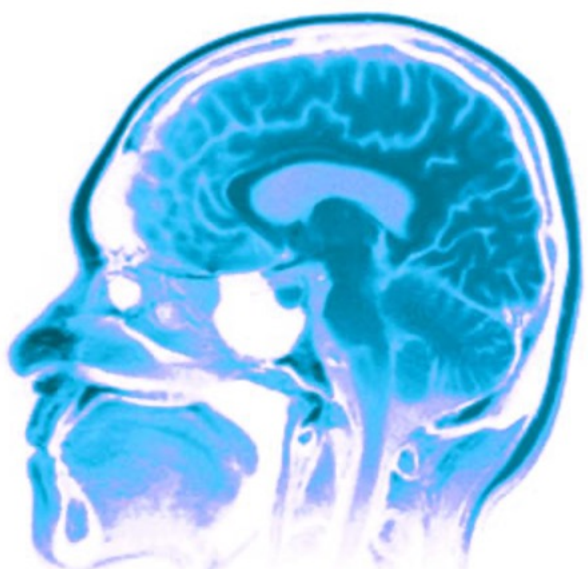
- Robots
  - Robots have had a variety of locomotion methods
  - Walking with legs, is challenging:
    - Differing terrains, stairs, running, ramps, etc.
  - Some impressive developments:
    - Boston dynamics: https://www.youtube.com/watch?v=tF4DML7FIWk and https://www.youtube.com/watch?v=-e1_QhJ1EhQ

Lecture 12: Intro to AI and Neural Networks

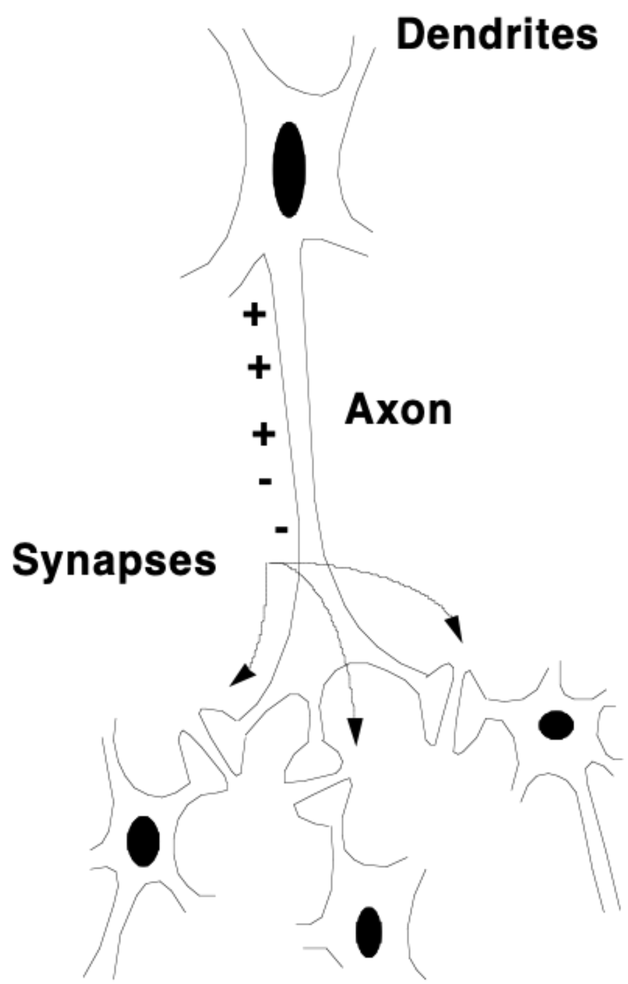
- Administrative
- Artificial Intelligence
- Neural Networks

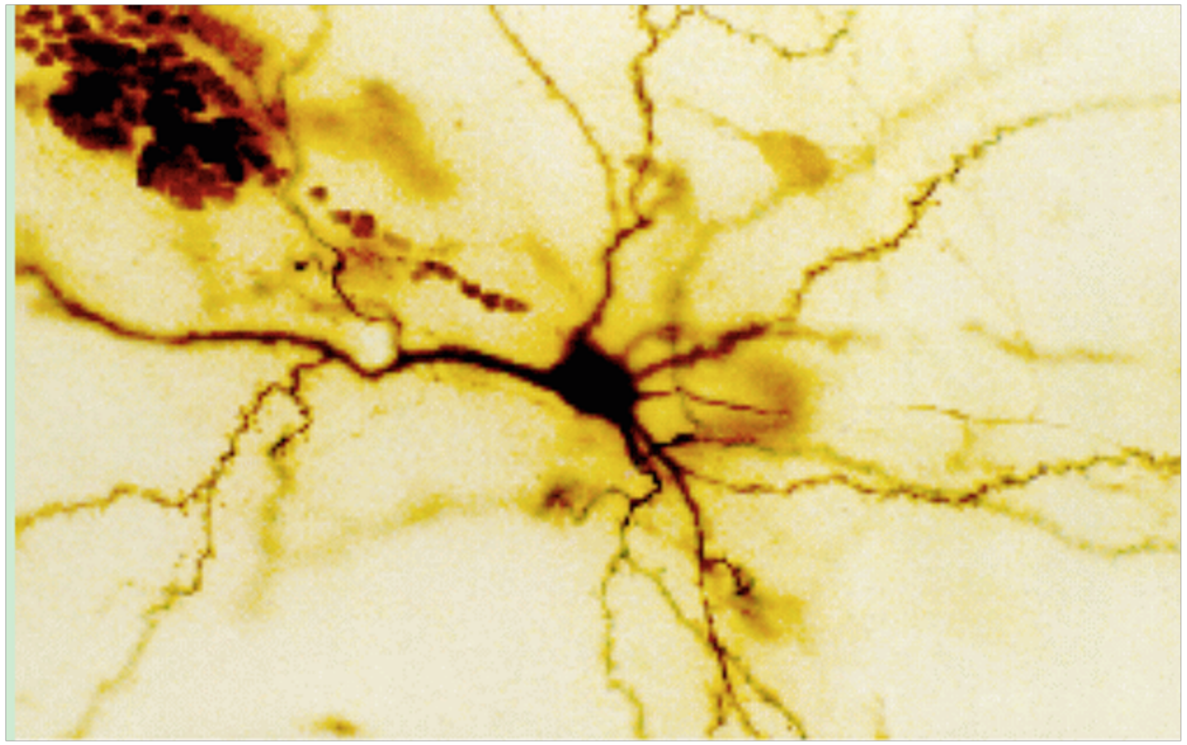
Neural Networks

- Try to mimic the structure and function of our nervous system.
  - *People like biologically motivated approaches*



Our nervous system





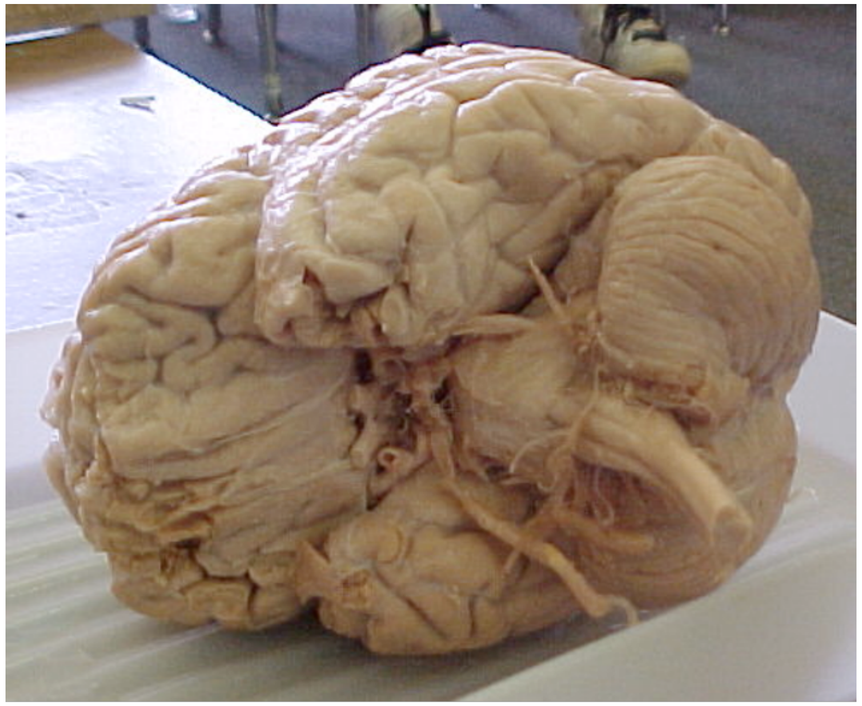
neuron

Our nervous system: *the computer science view*

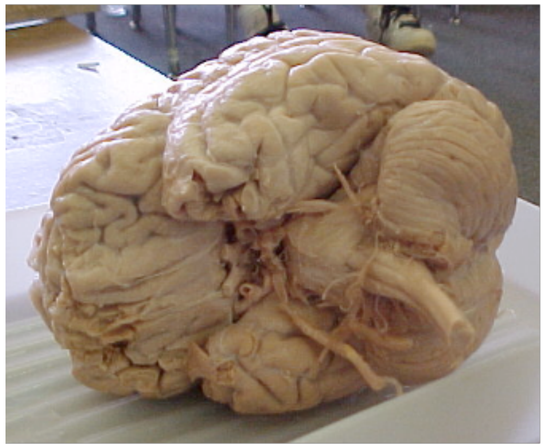
- The human brain is a large collection of interconnected neurons.
- A neuron is a brain cell:
  - They collect, process, and disseminate electrical signals.
  - They are connected via synapses.
  - They fire depending on the condition of the neighbors.

Our nervous system

- The human brain:
  - Contains ~10^{11} (100 billion) neurons.
  - Each neuron is connected to ~10^4 (10,000) other neurons.
  - Neurons can fire as fast as 10^{-3} seconds.
- How does this compare to computers?



Human vs. Machine



- 10^{11} neurons
- 10^{11} neurons
- 10^{14} synapses
- 10^{-3} "cycle" time

- 10^{10} transitions
- 10^{11} bits of RAM/memory
- 10^{13} bits on disk
- 10^{-9} cycle time

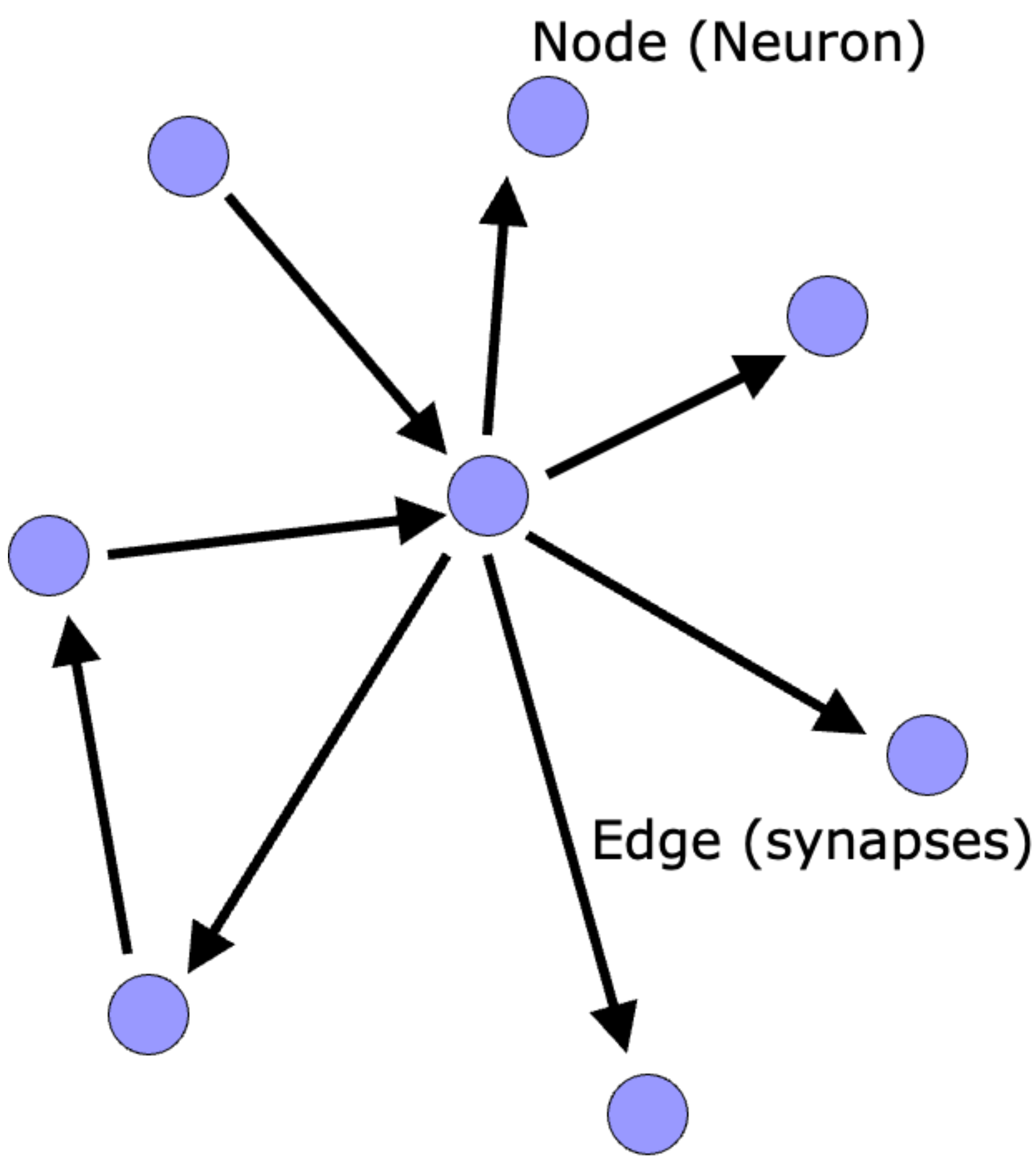
Brains are still pretty fast

Who is this?

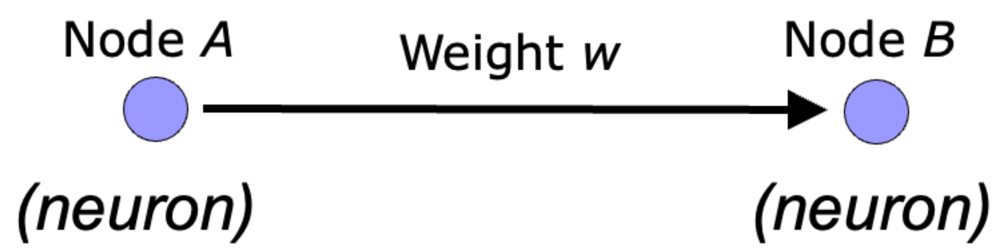
Brains are still pretty fast

- If you follow basketball, you will be able to identify Lebron James in under a second!
- Given a neuron firing time of 10^{-3} seconds, how many neurons could fire in about a second?
  - A few hundred maybe a thousand.
- What are possible explanations?
  - Either neurons are performing some very complicated computations or brain is taking advantage of the **massive** parallelization (remember, neurons are connected ~10,000 other neurons.

Artificial Neural Networks - Our approximation

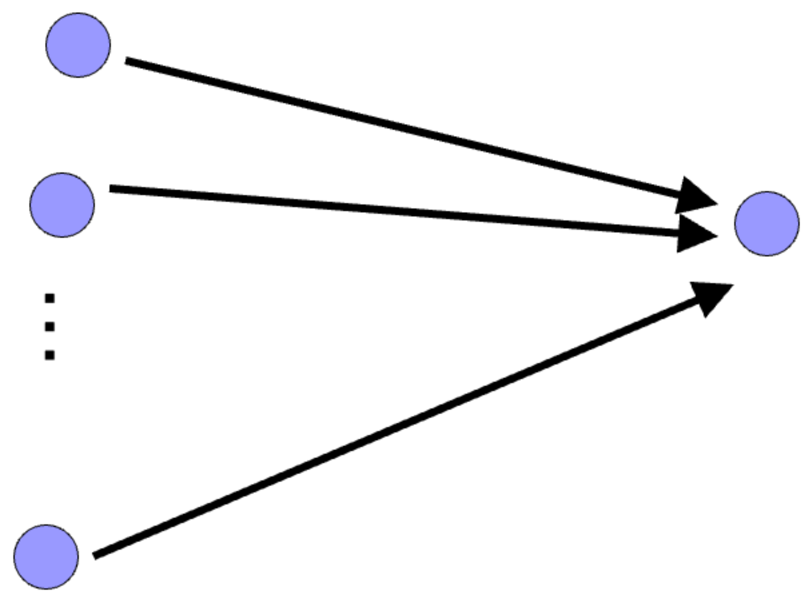


Strength of signal



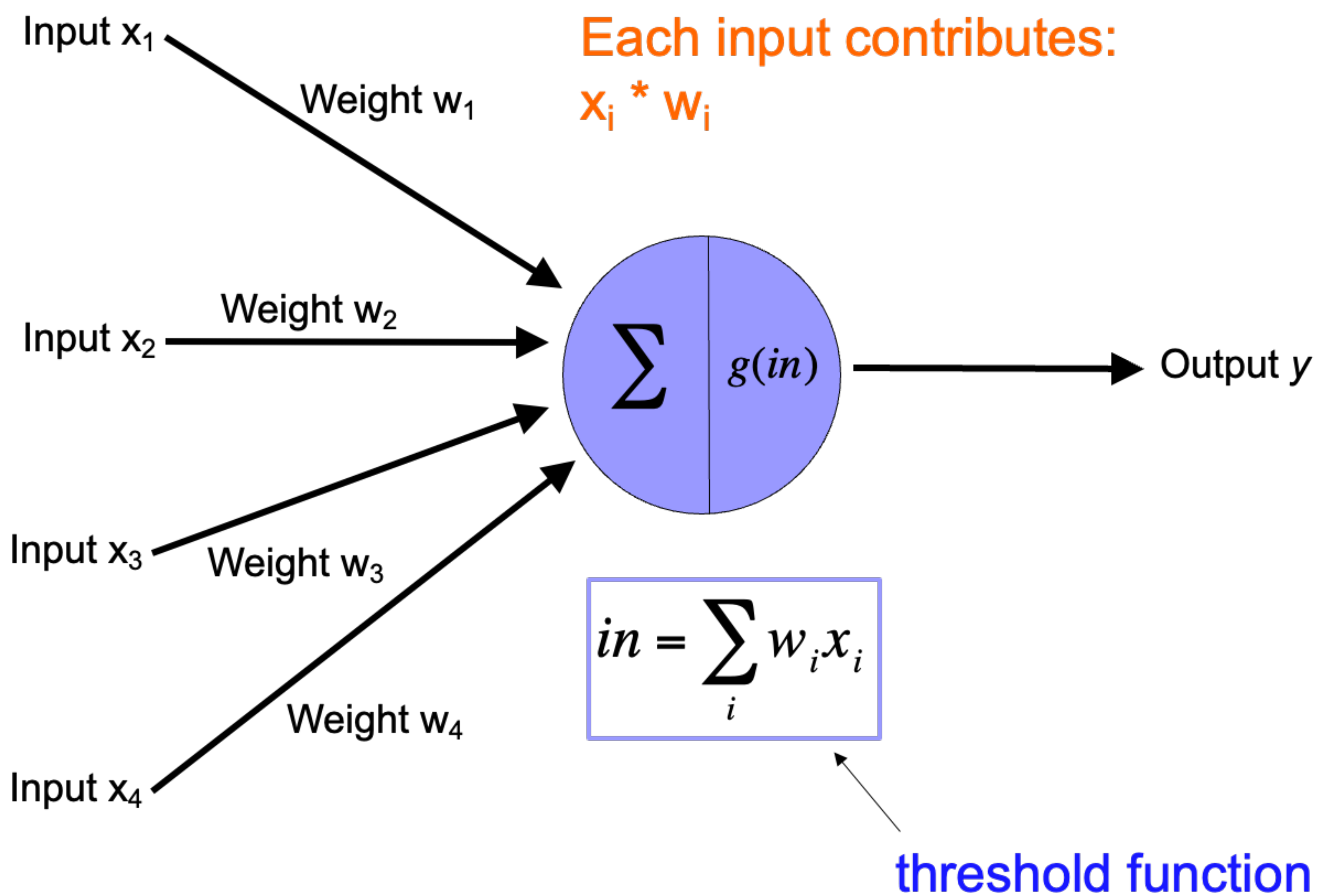
- w is the strength of signal sent between A and B.
- If A fires and w is positive, then A stimulates B.
- If A fires and w is negative, then A inhibits B.

Firing a neuron



- A given neuron has many, many connecting, input neurons.
- If a neuron is stimulated enough, then it also fires.
- How much stimulation is required is determined by its threshold.

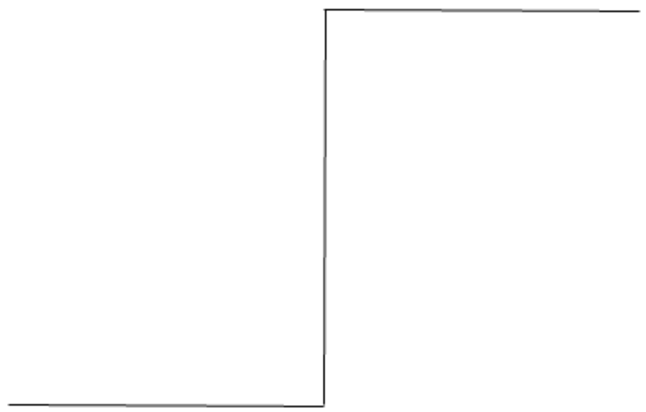
A single neuron/perceptron



Possible threshold functions

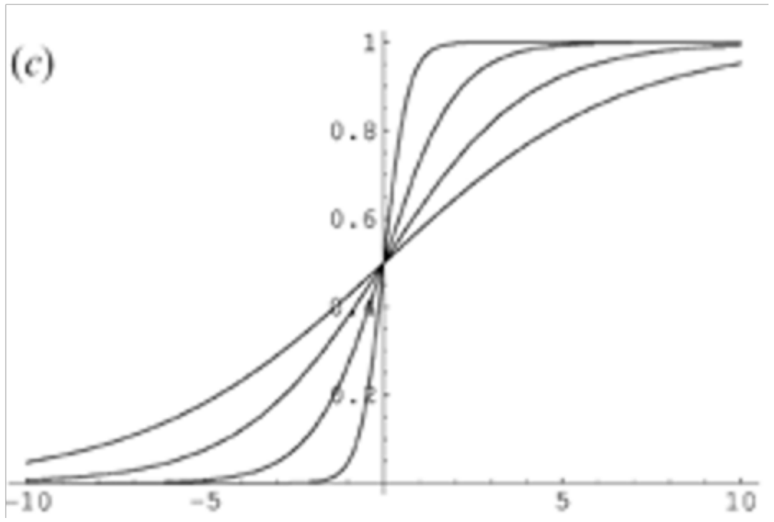
hard threshold

$$g(x) = \begin{cases} 1 & \textit{if } \mathrm{x} \geq \textit{threshold} \\ 0 & \textit{otherwise} \end{cases}$$

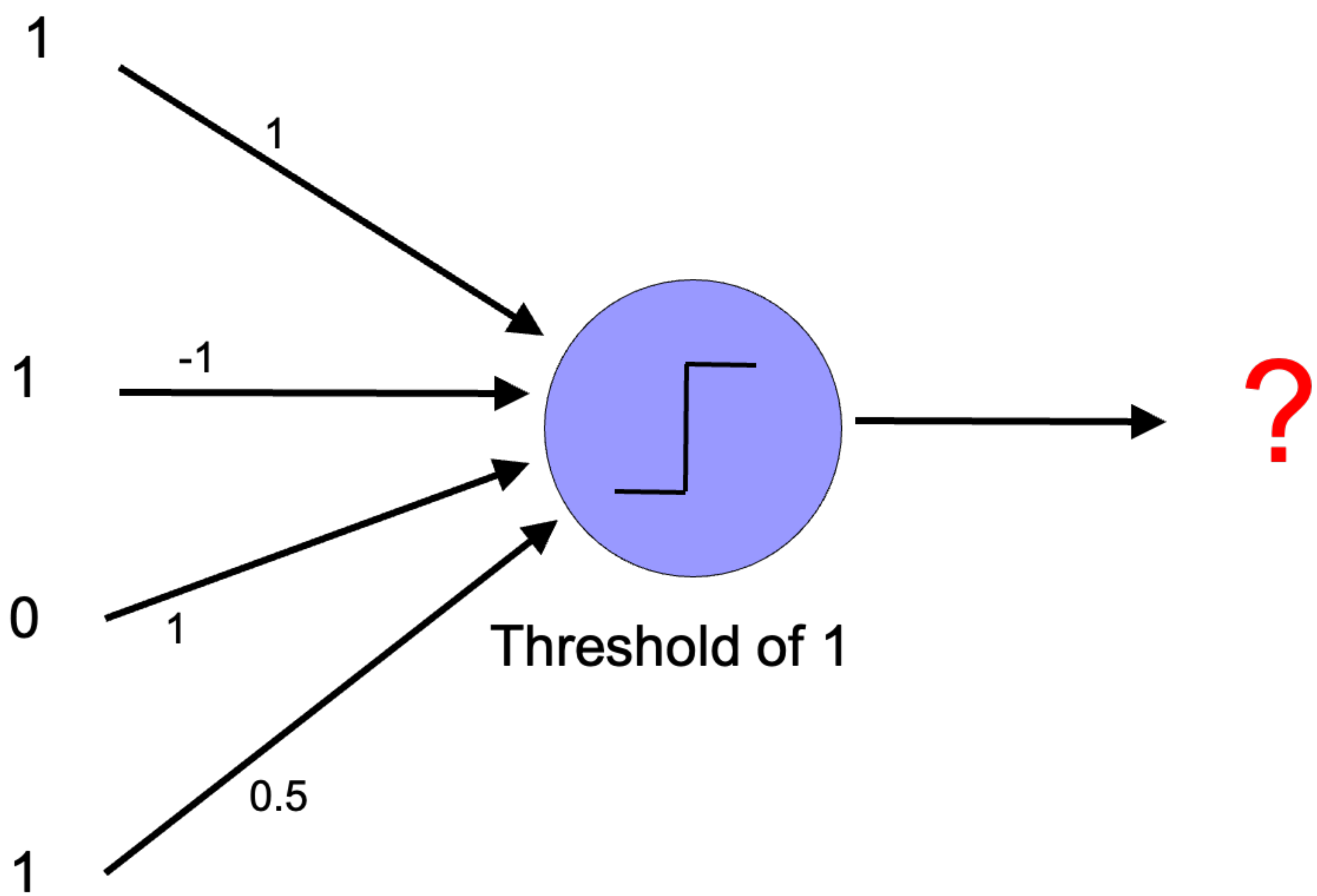


sigmoid

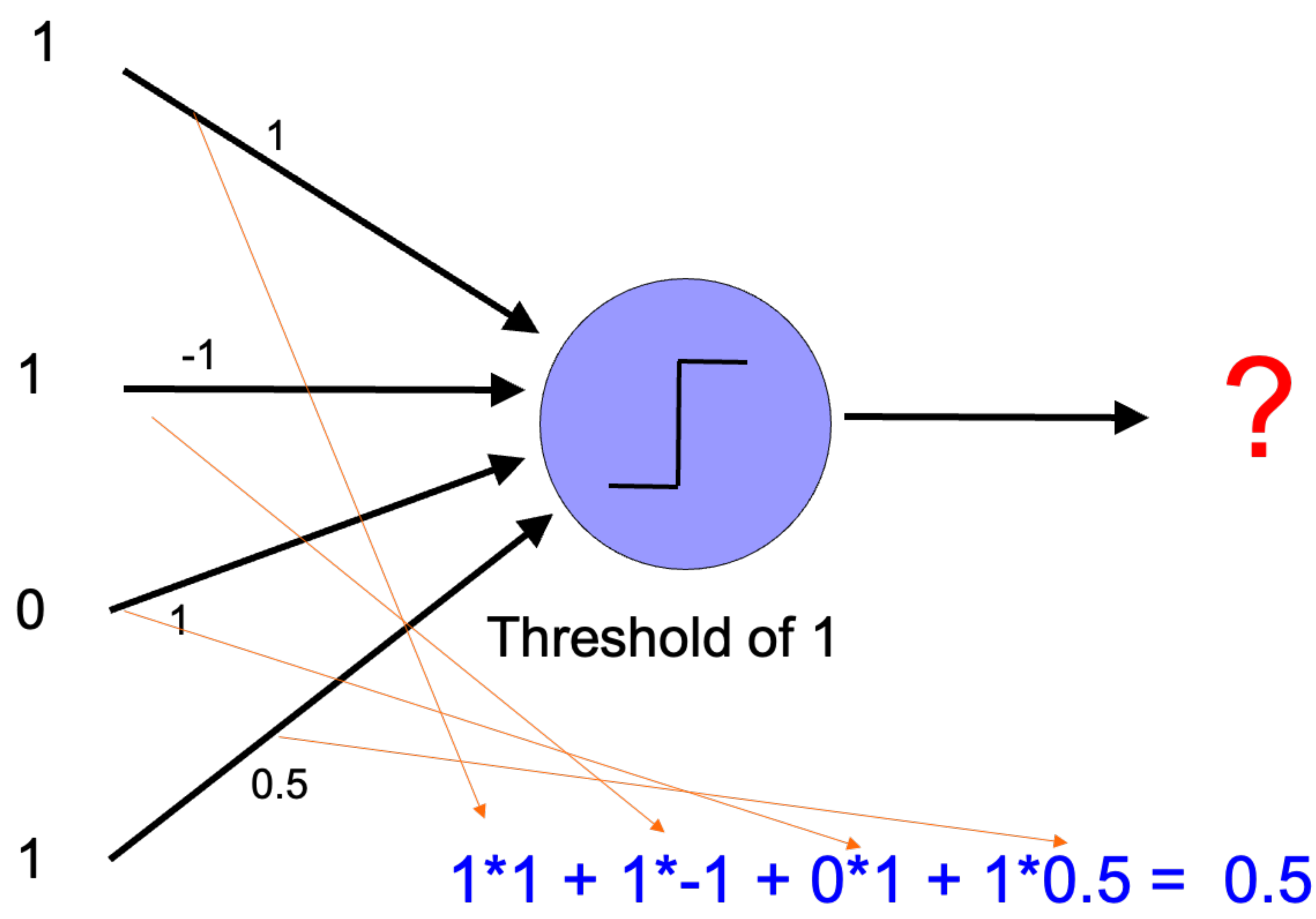
$$g(x) = \frac{1}{1 + e^{-ax}}$$



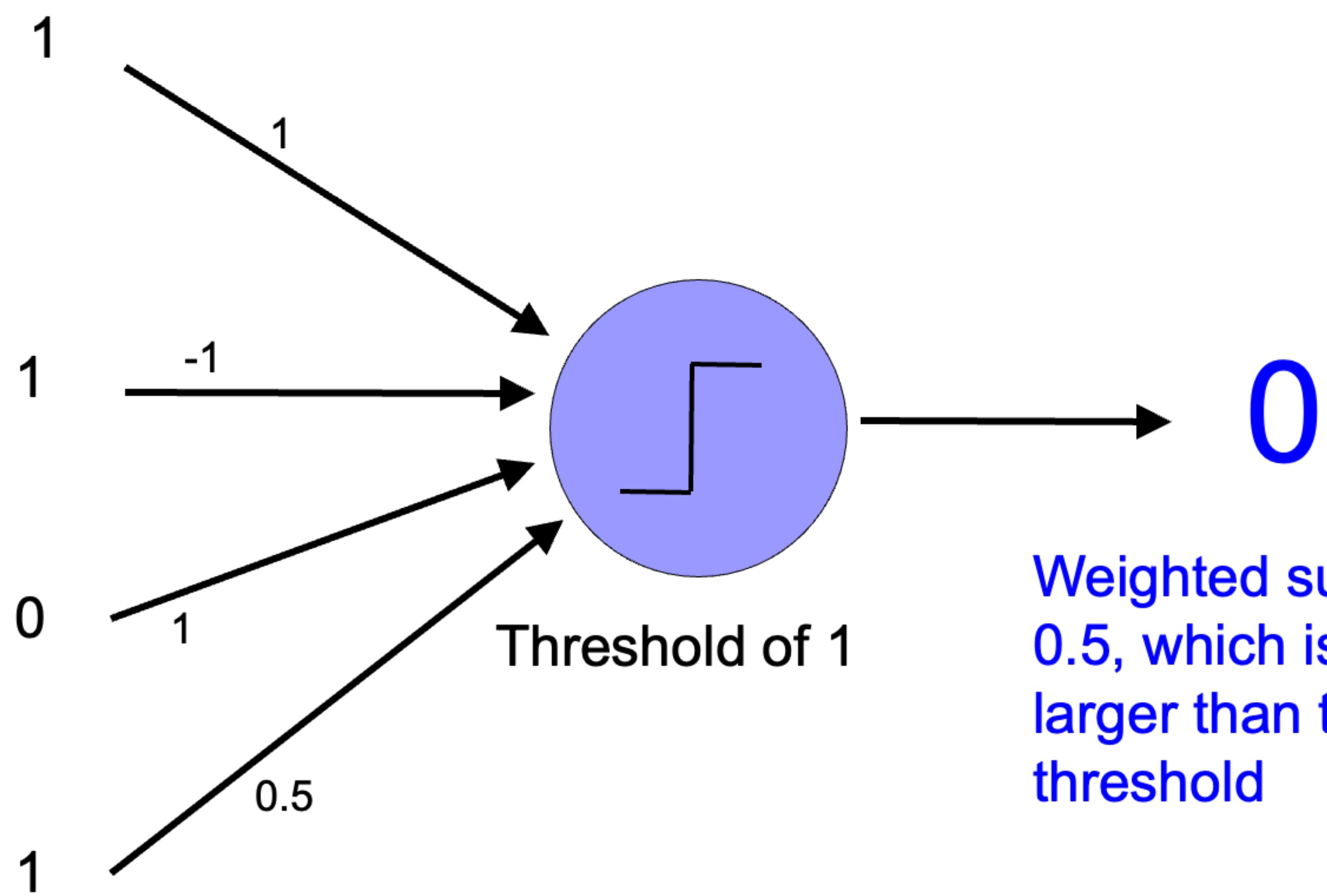
A single neuron/perceptron



A single neuron/perceptron

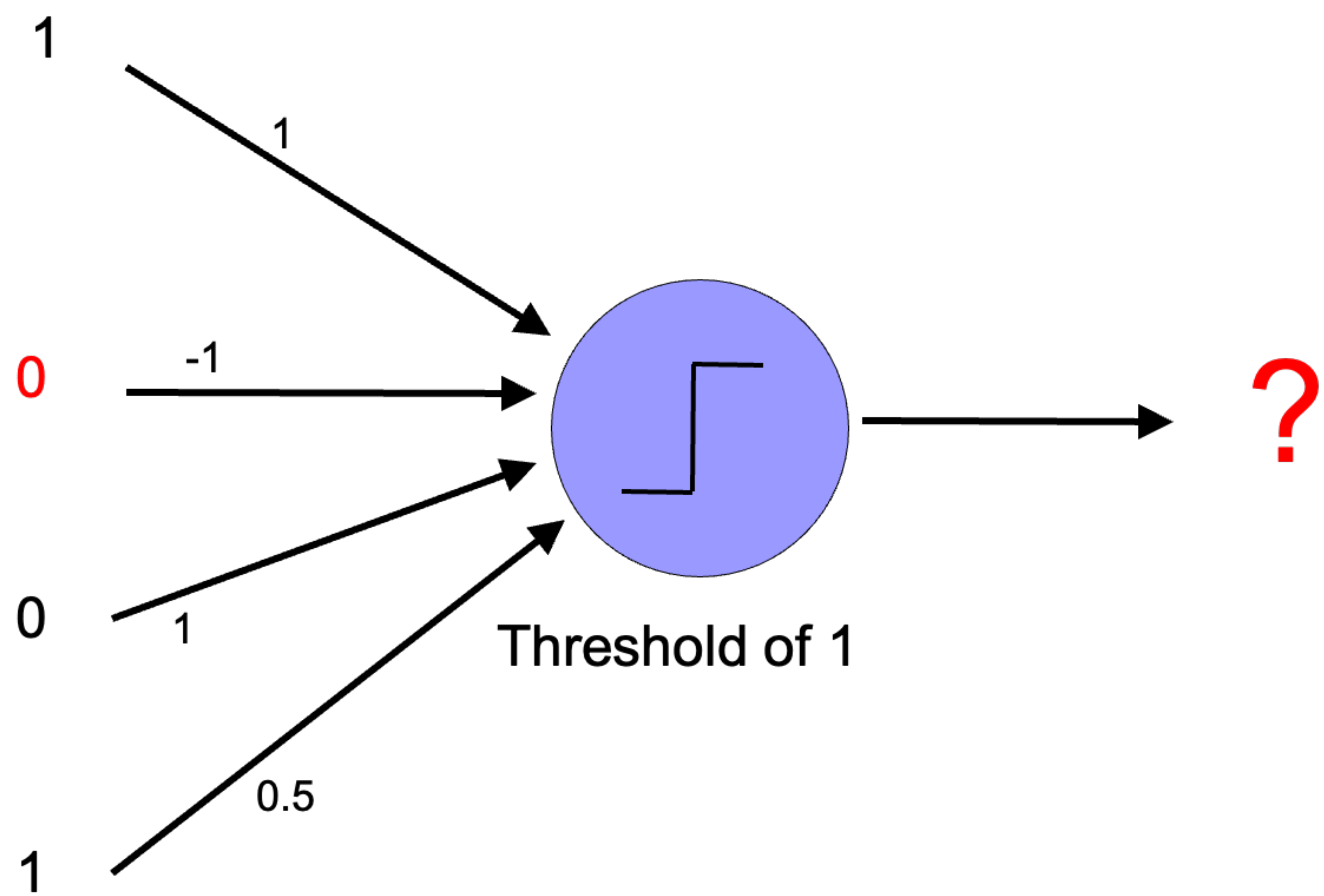


A single neuron/perceptron

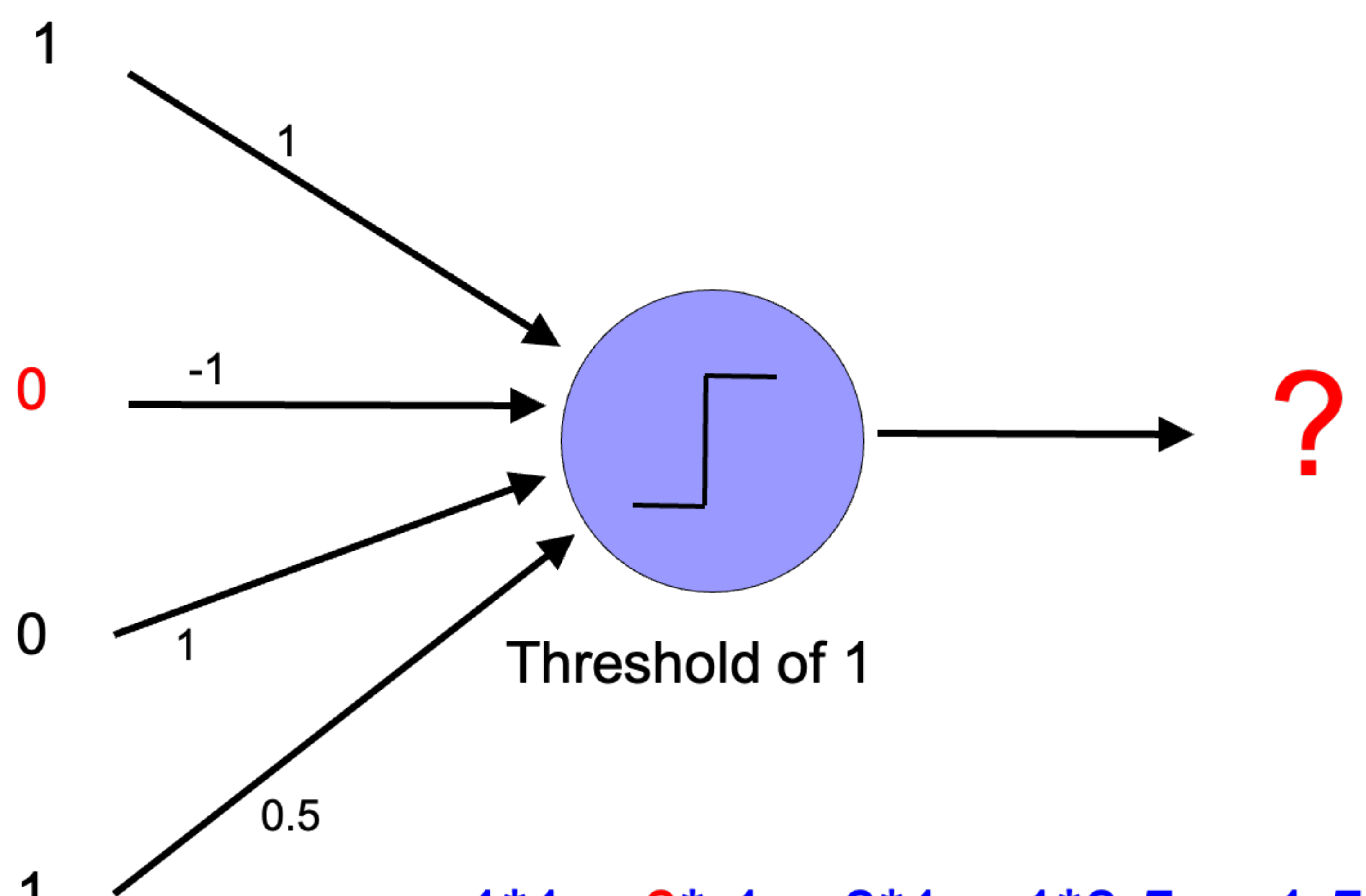


Weighted sum is 0.5, which is not larger than the threshold

A single neuron/perceptron

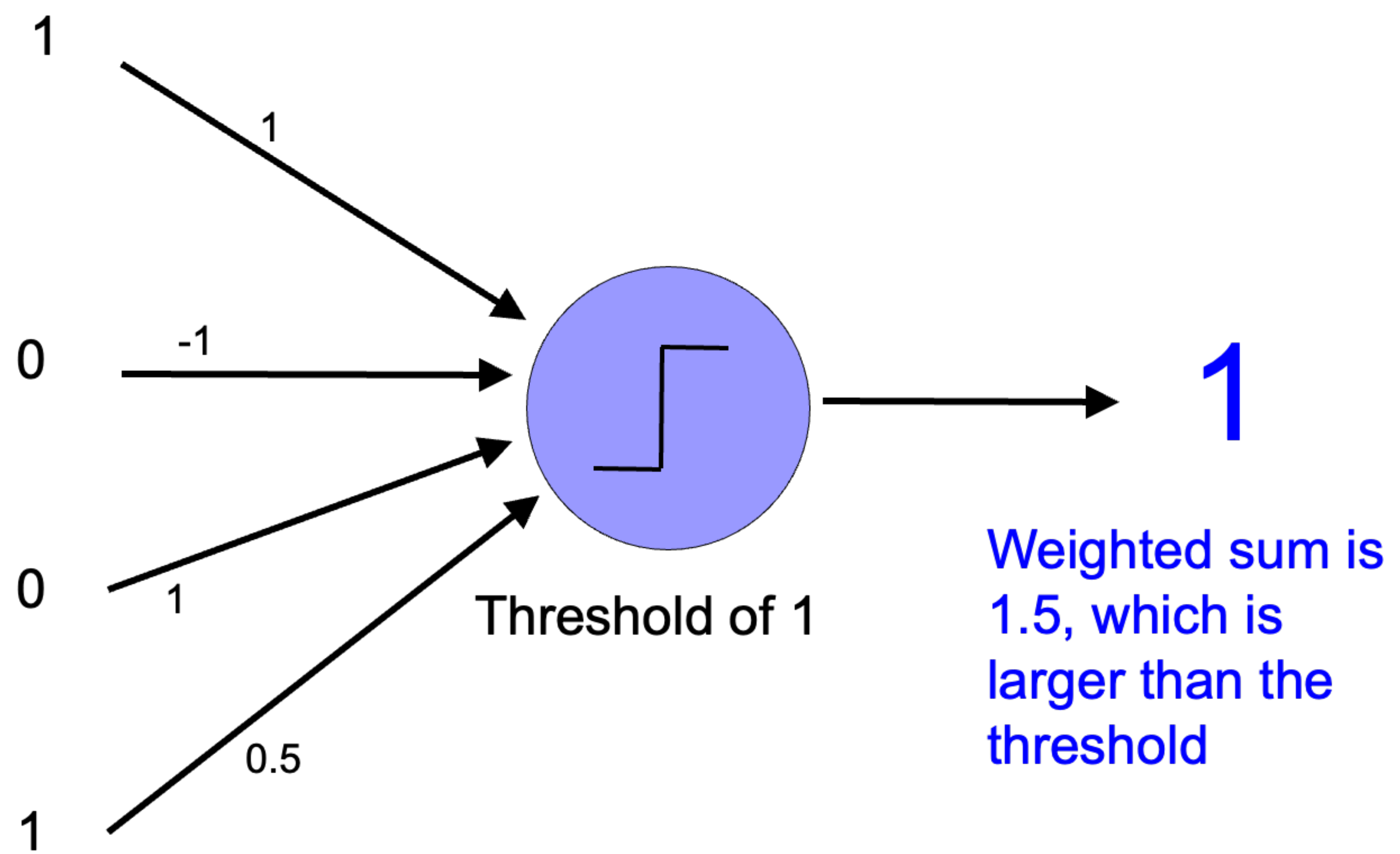


A single neuron/perceptron

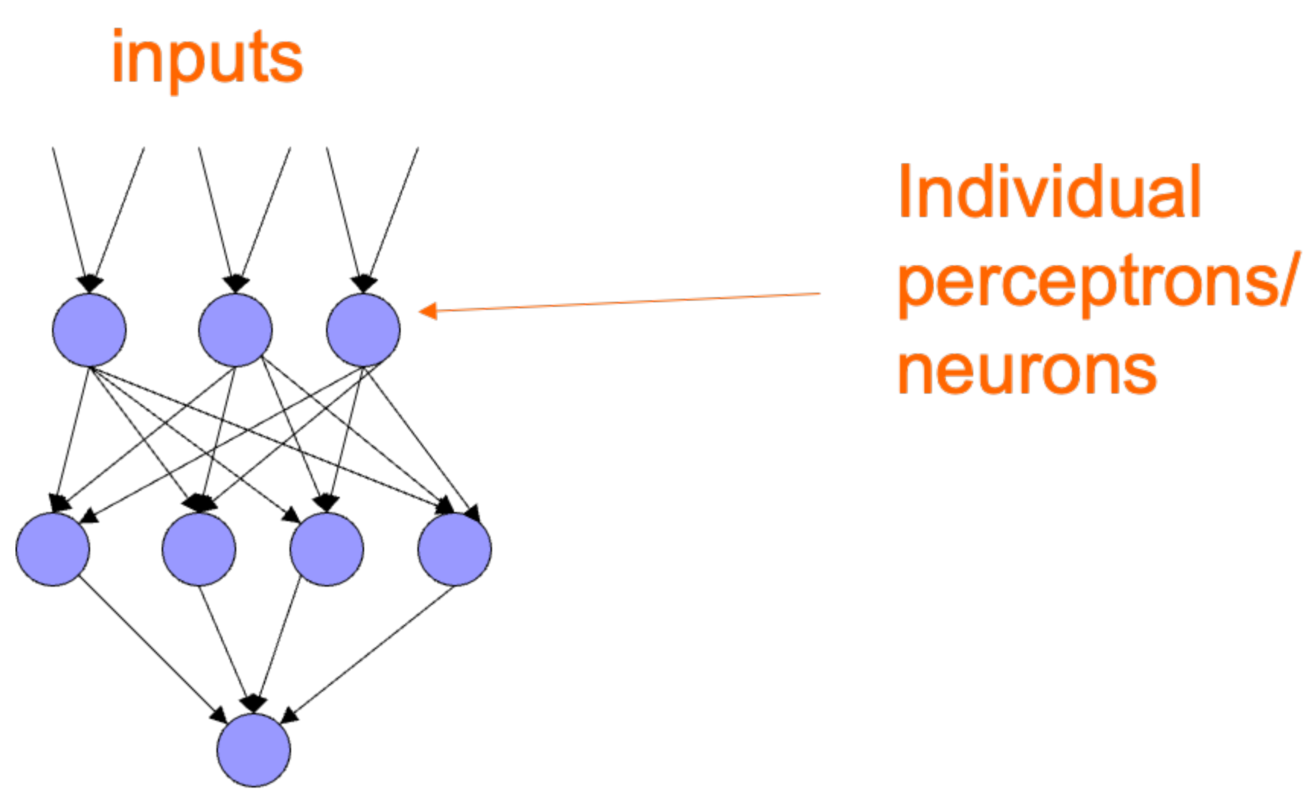


$1*1 + 0*-1 + 0*1 + 1*0.5 = 1.5$

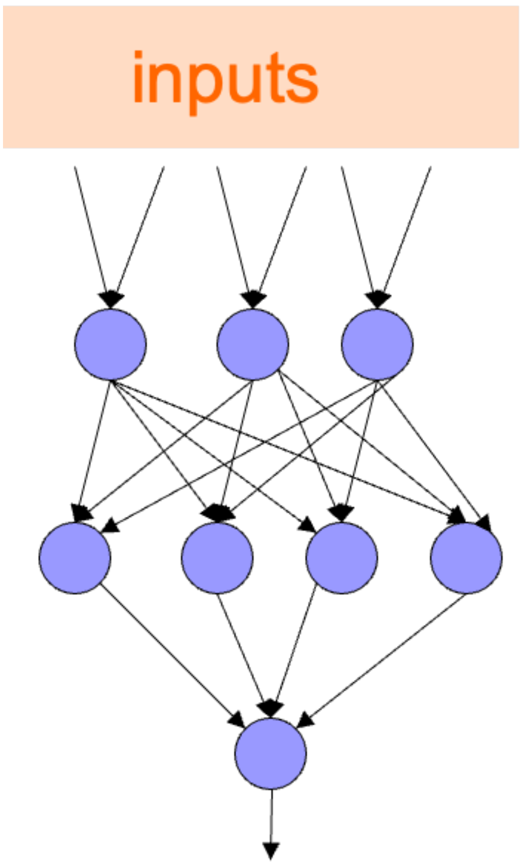
A single neuron/perceptron



Neural network

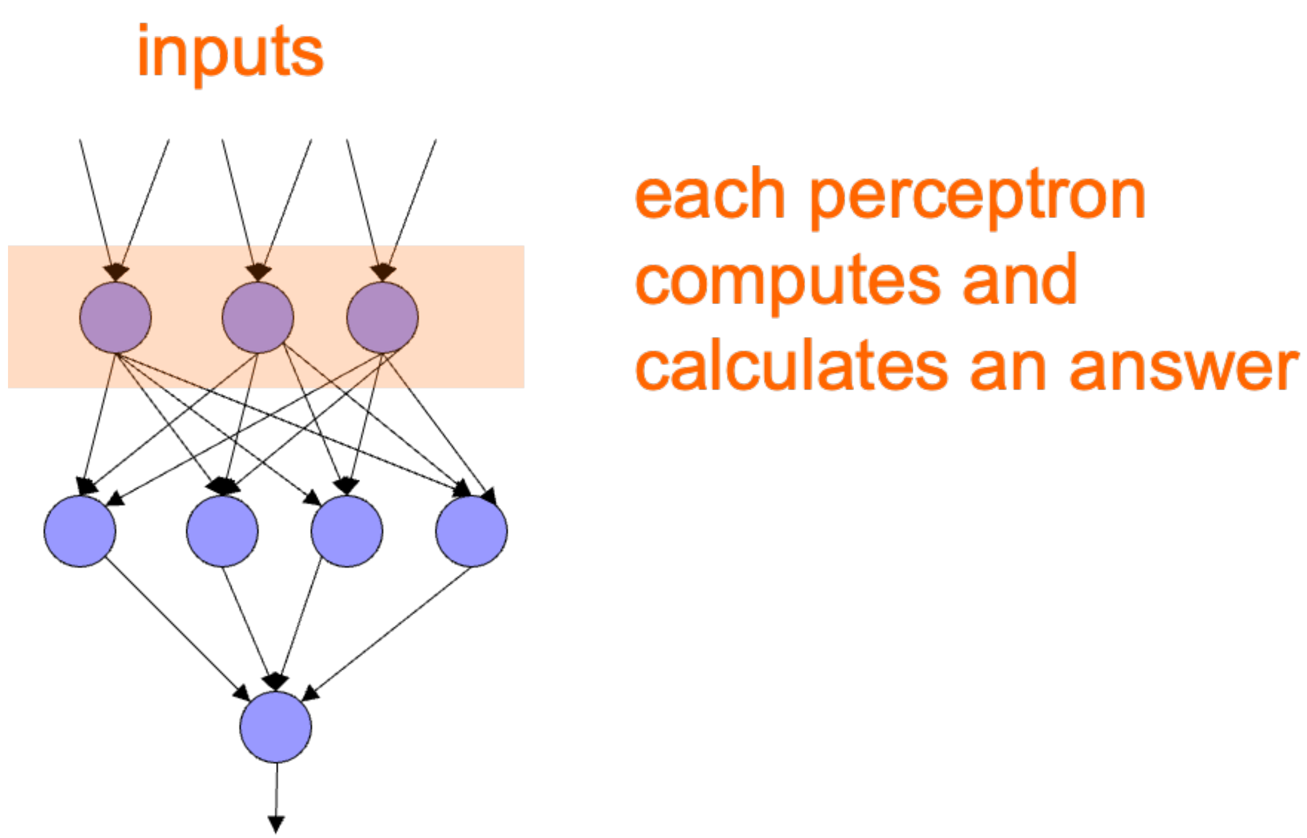


Neural network

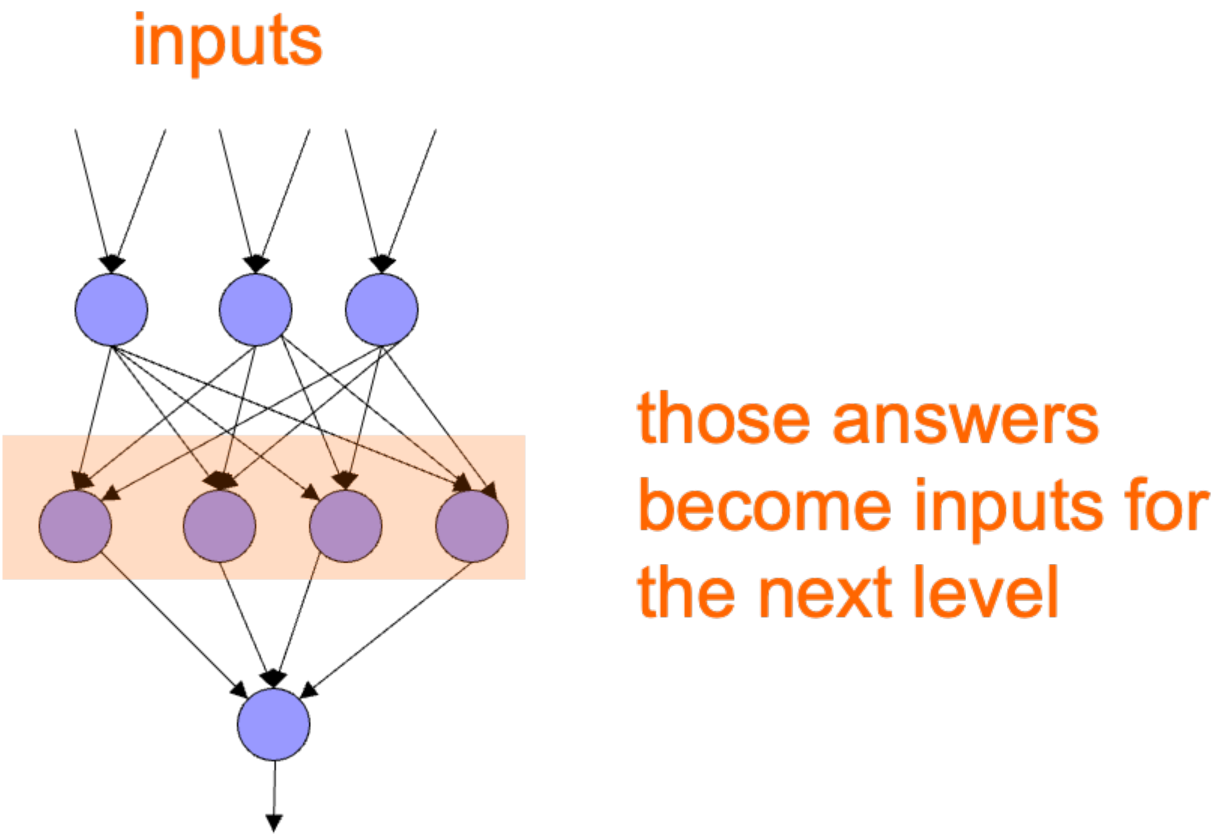


some inputs are provided/entered

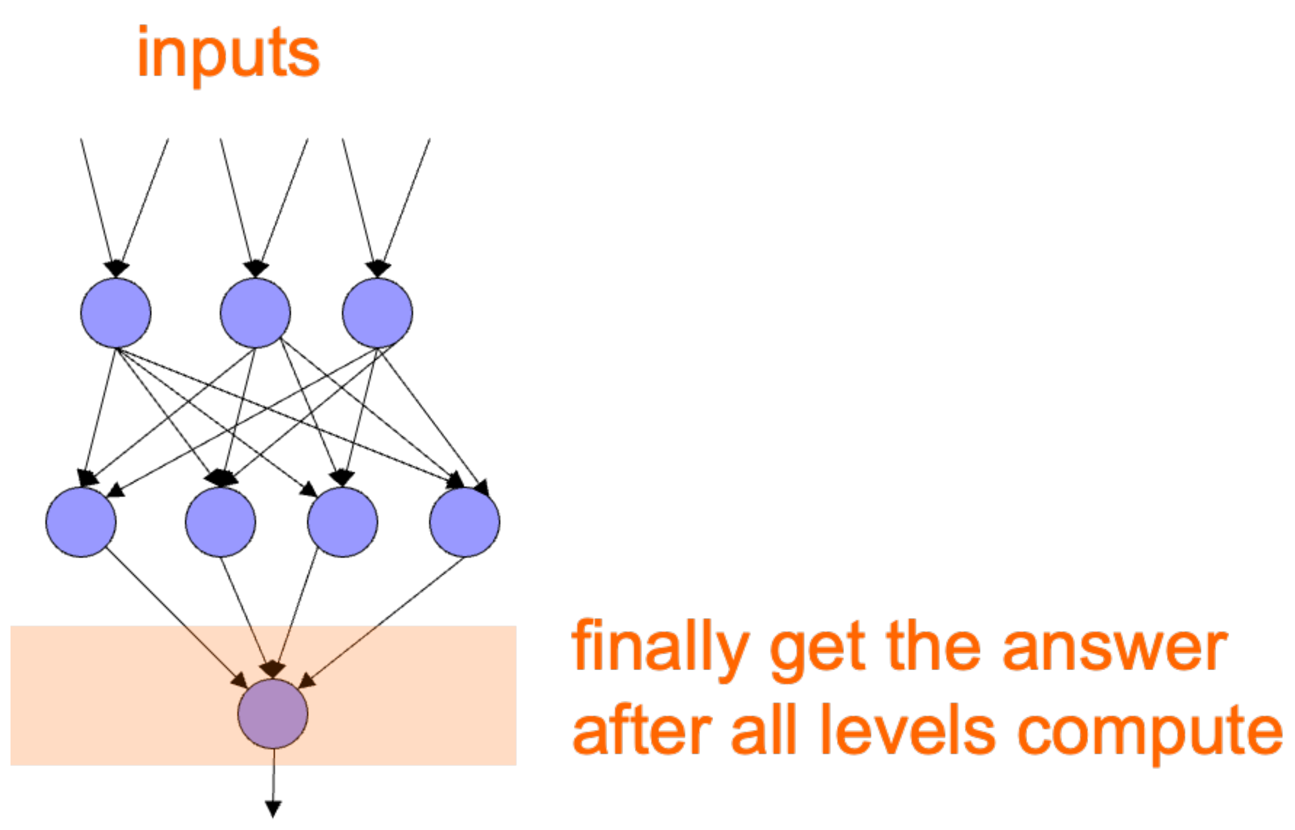
Neural network



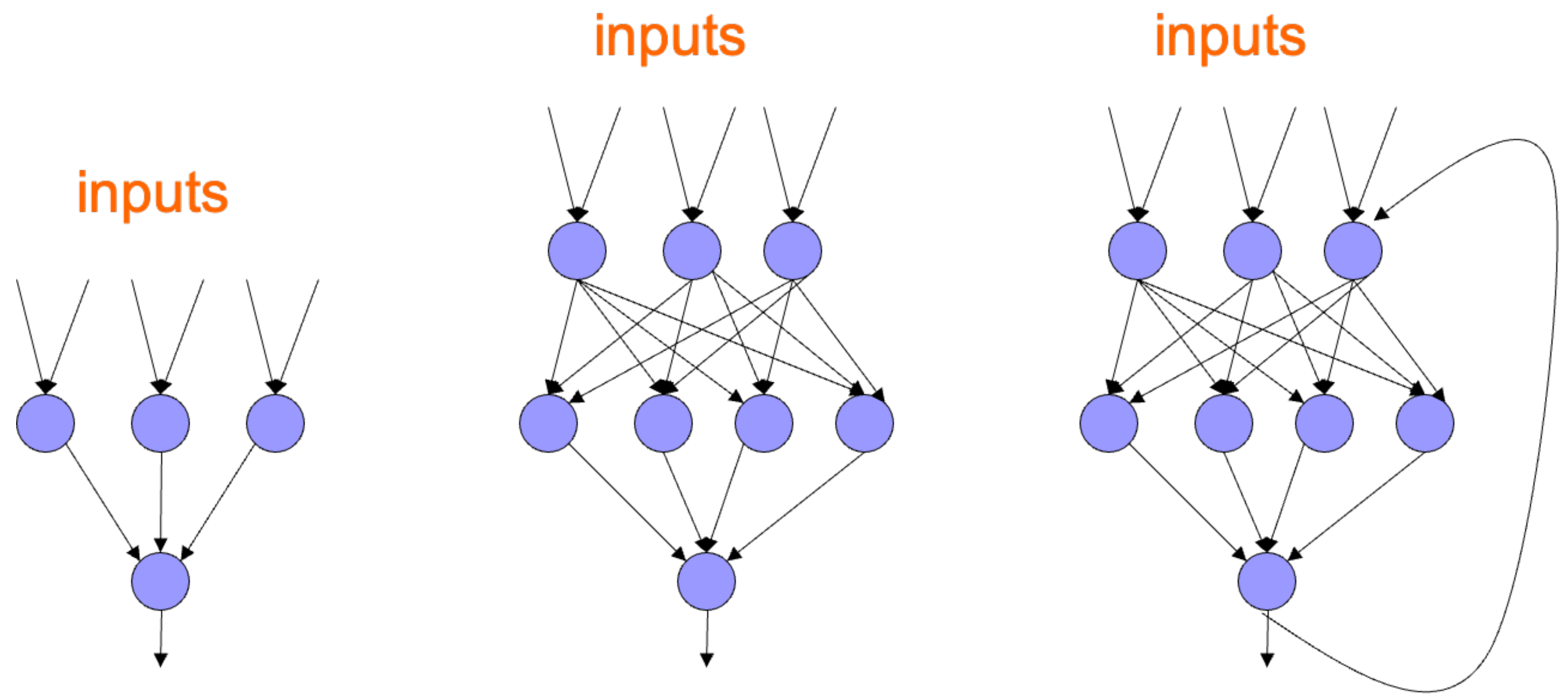
Neural network



Neural network



Different kinds of neural networks



History of neural networks

- McCulloch and Pitts (1943) – introduced model of artificial neurons and suggested they could learn.
- Hebb (1949) – Simple updating rule for learning.
- Rosenblatt (1962) - the perceptron model.
- Minsky and Papert (1969) – wrote Perceptrons.
- Bryson and Ho (1969, but largely ignored until 1980s--Rosenblatt) – invented back-propagation learning for multilayer networks.

Training the perceptron

- First wave in neural networks in the 1960's.
- Single neuron.
- Trainable: its threshold and input weights can be modified.
- If the neuron doesn't give the desired output, then it has made a mistake.
- Input weights and threshold can be changed according to a learning algorithm.

Examples - Logical operators

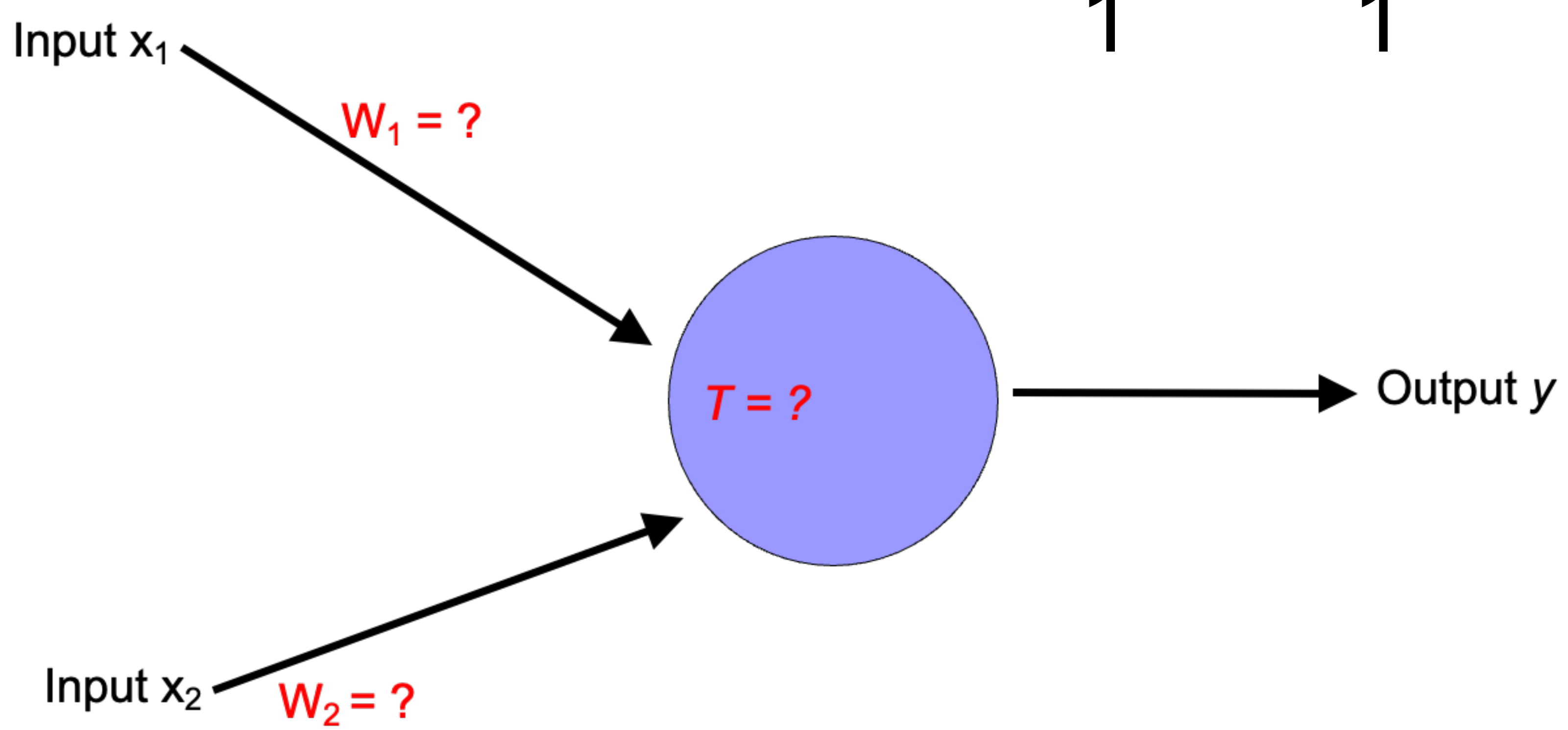
- AND – if all inputs are 1, return 1, otherwise return 0
- OR – if at least one input is 1, return 1, otherwise return 0
- NOT – return the opposite of the input
- XOR – if exactly one input is 1, then return 1, otherwise return 0

AND

x_1	x_2	x_1 **and** x_2
0	0	0
0	1	0
1	0	0
1	1	1

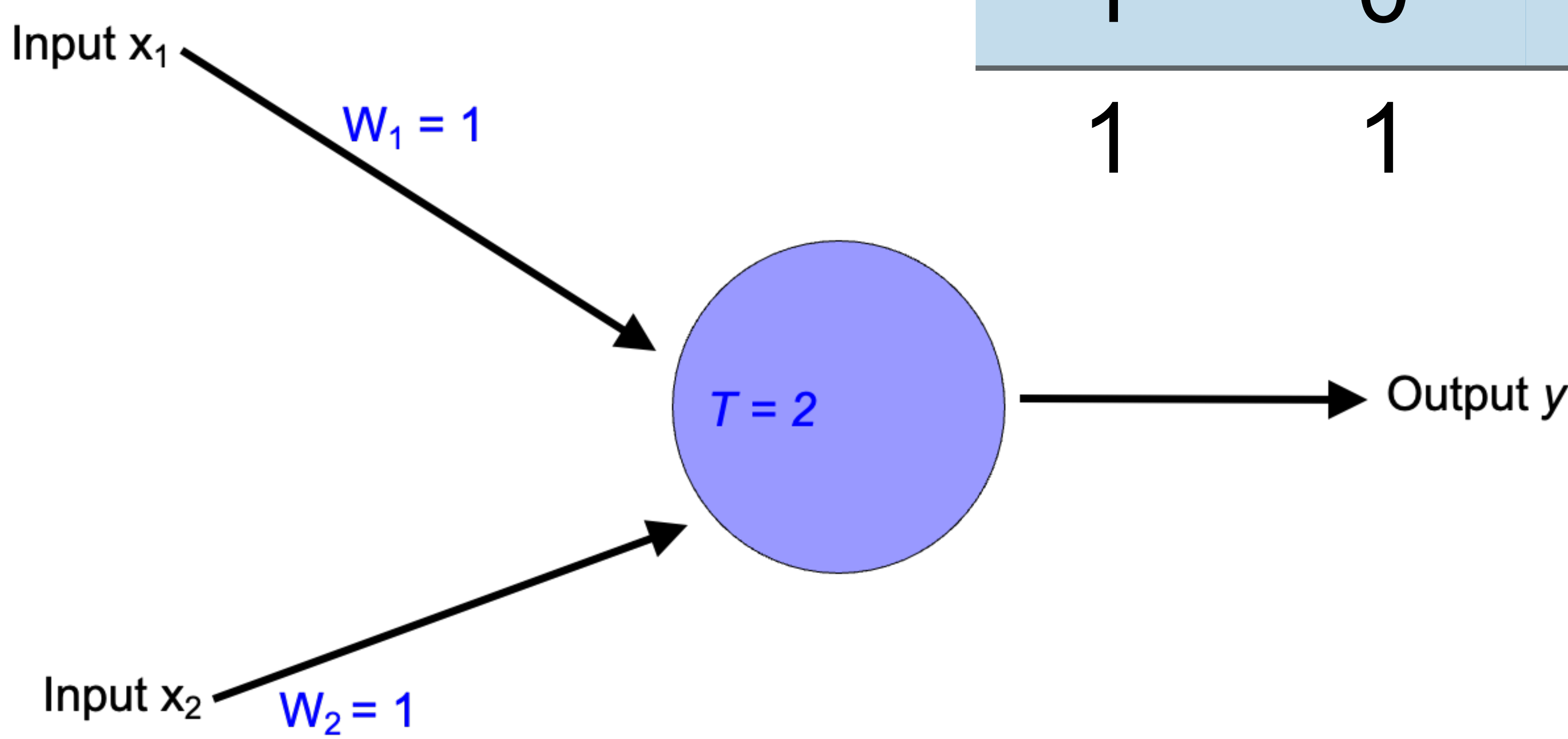
AND

x_1	x_2	x_1 **and** x_2
0	0	0
0	1	0
1	0	0
1	1	1



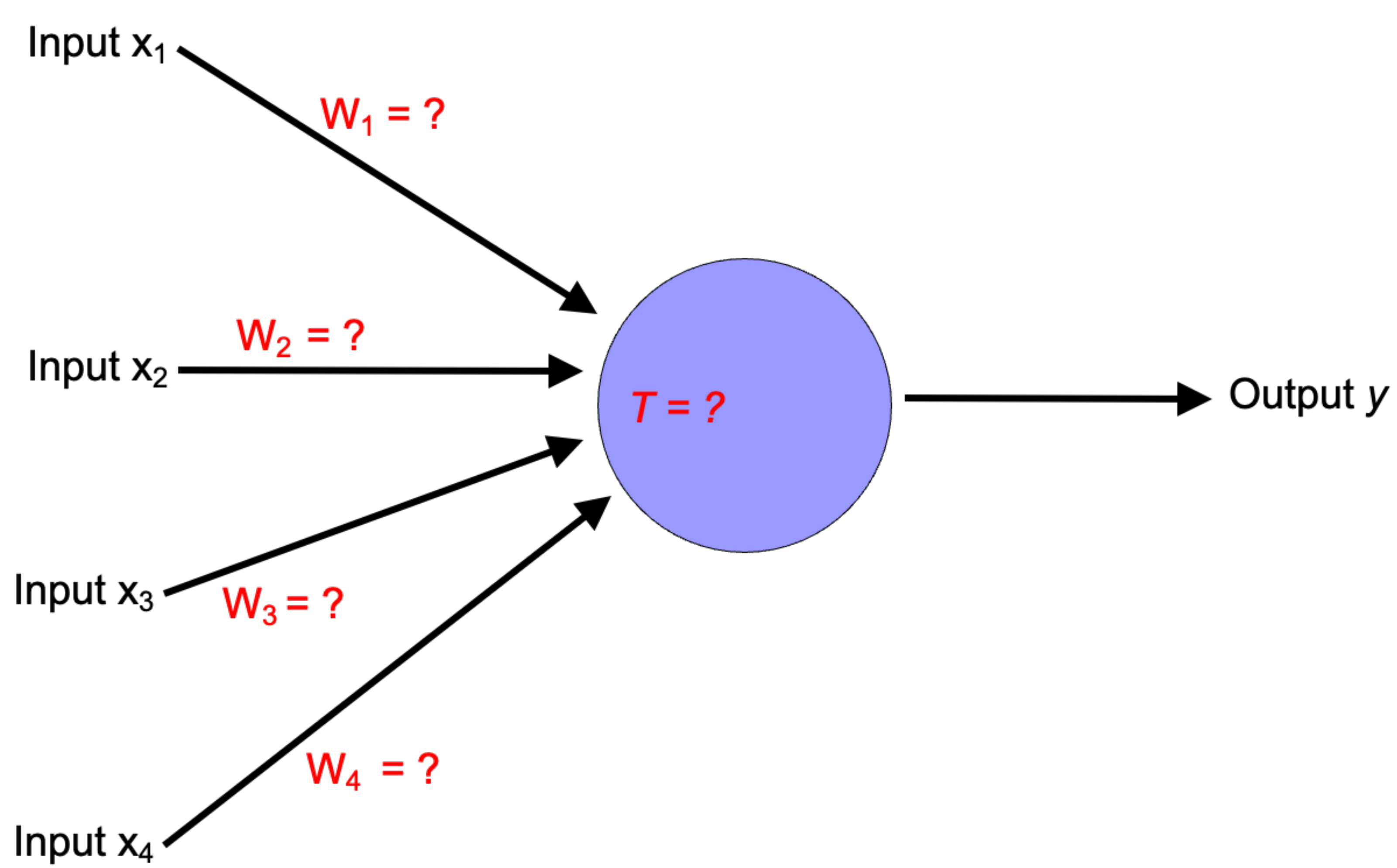
AND

x_1	x_2	x_1 **and** x_2
0	0	0
0	1	0
1	0	0
1	1	1

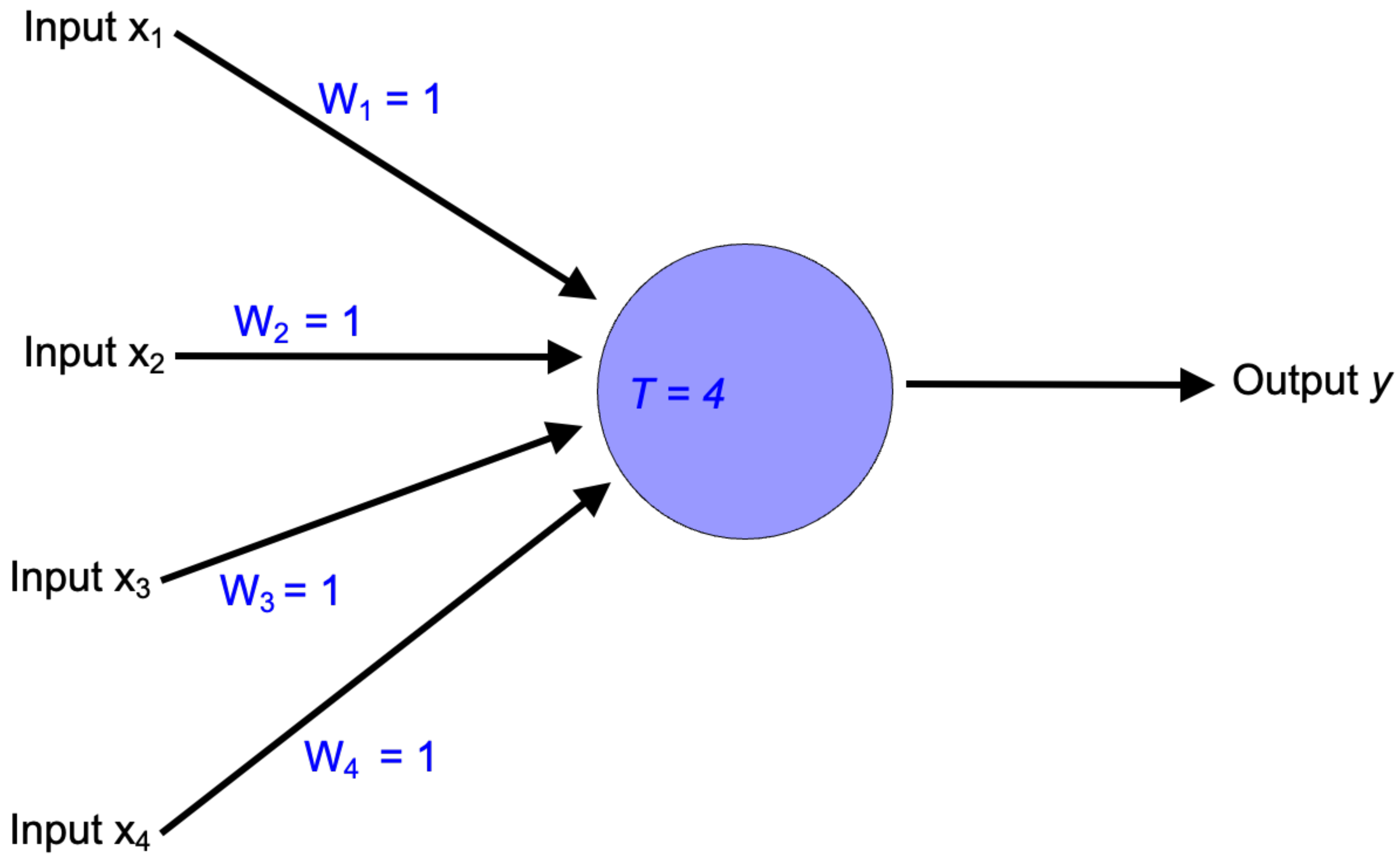


Inputs are either 0 or 1

AND



AND



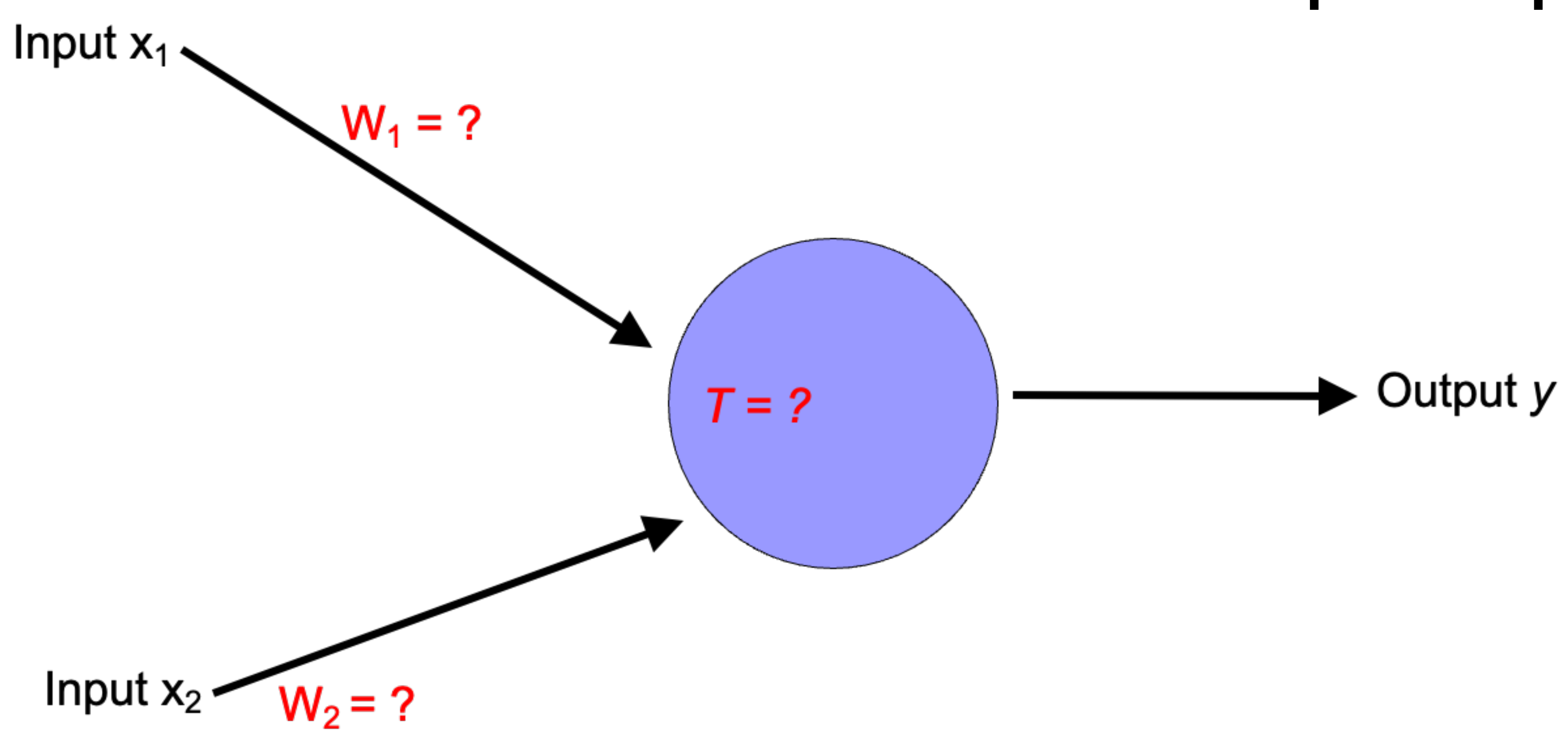
Inputs are either 0 or 1

OR

x_1	x_2	x_1 **or** x_2
0	0	0
0	1	1
1	0	1
1	1	1

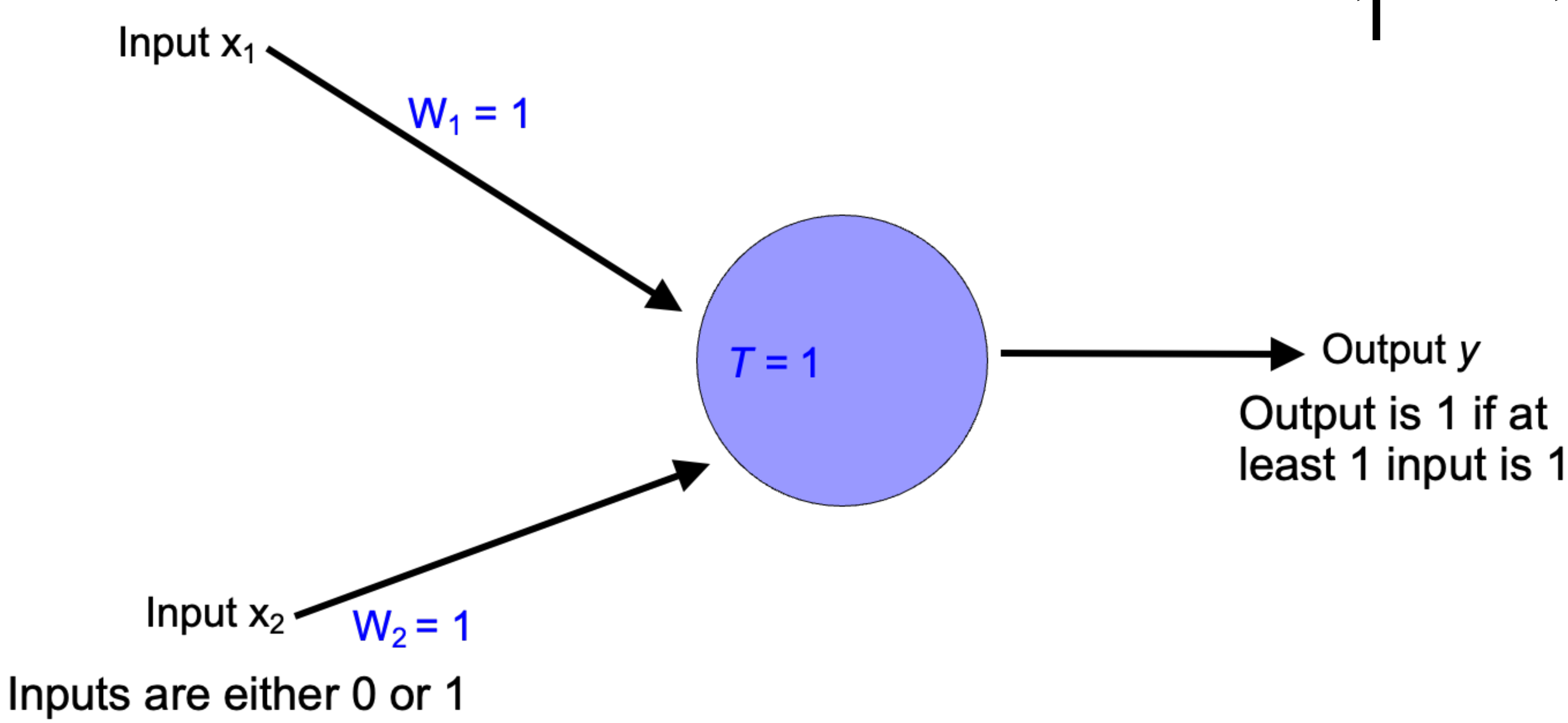
OR

x_1	x_2	x_1 **or** x_2
0	0	0
0	1	1
1	0	1
1	1	1

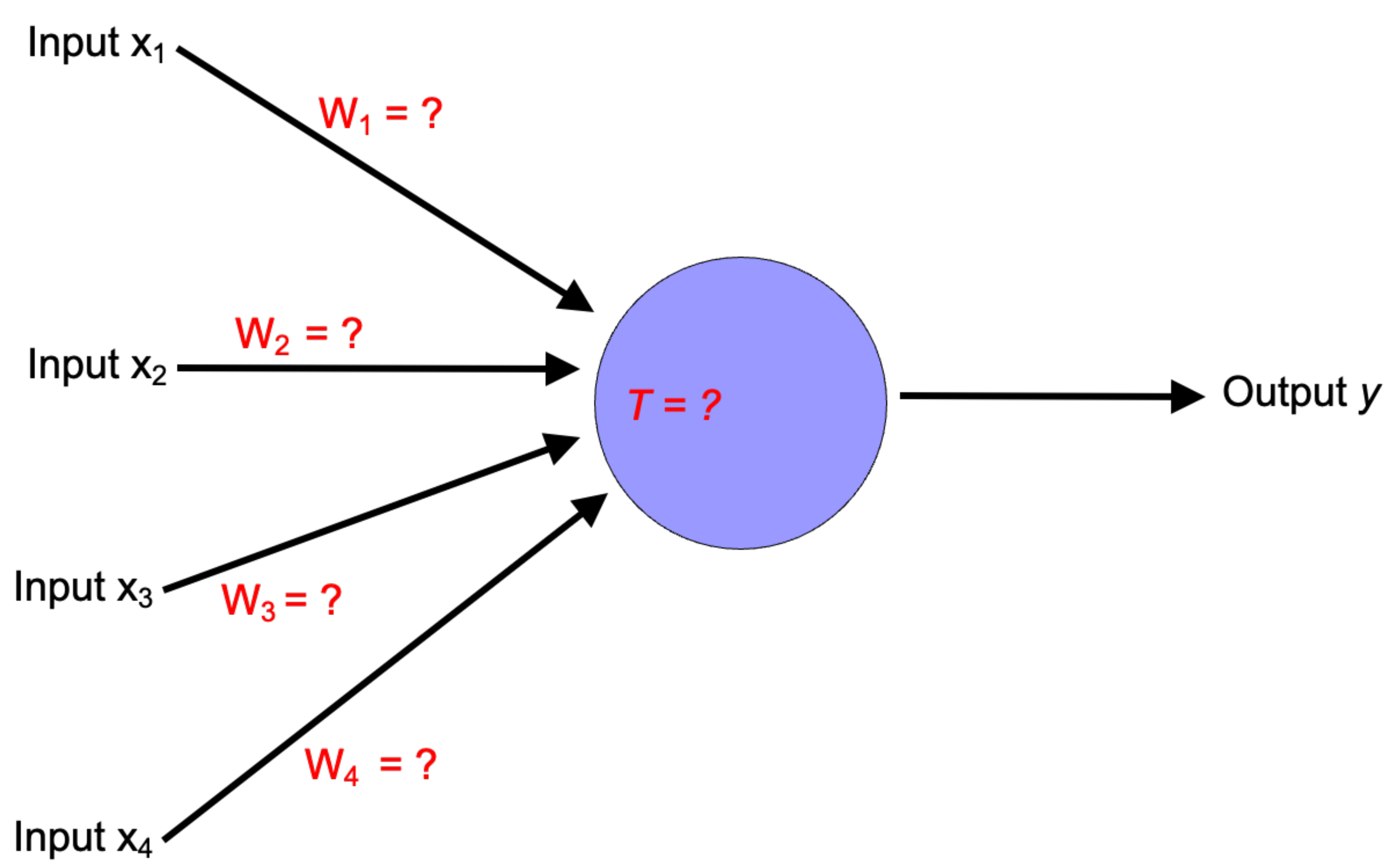


OR

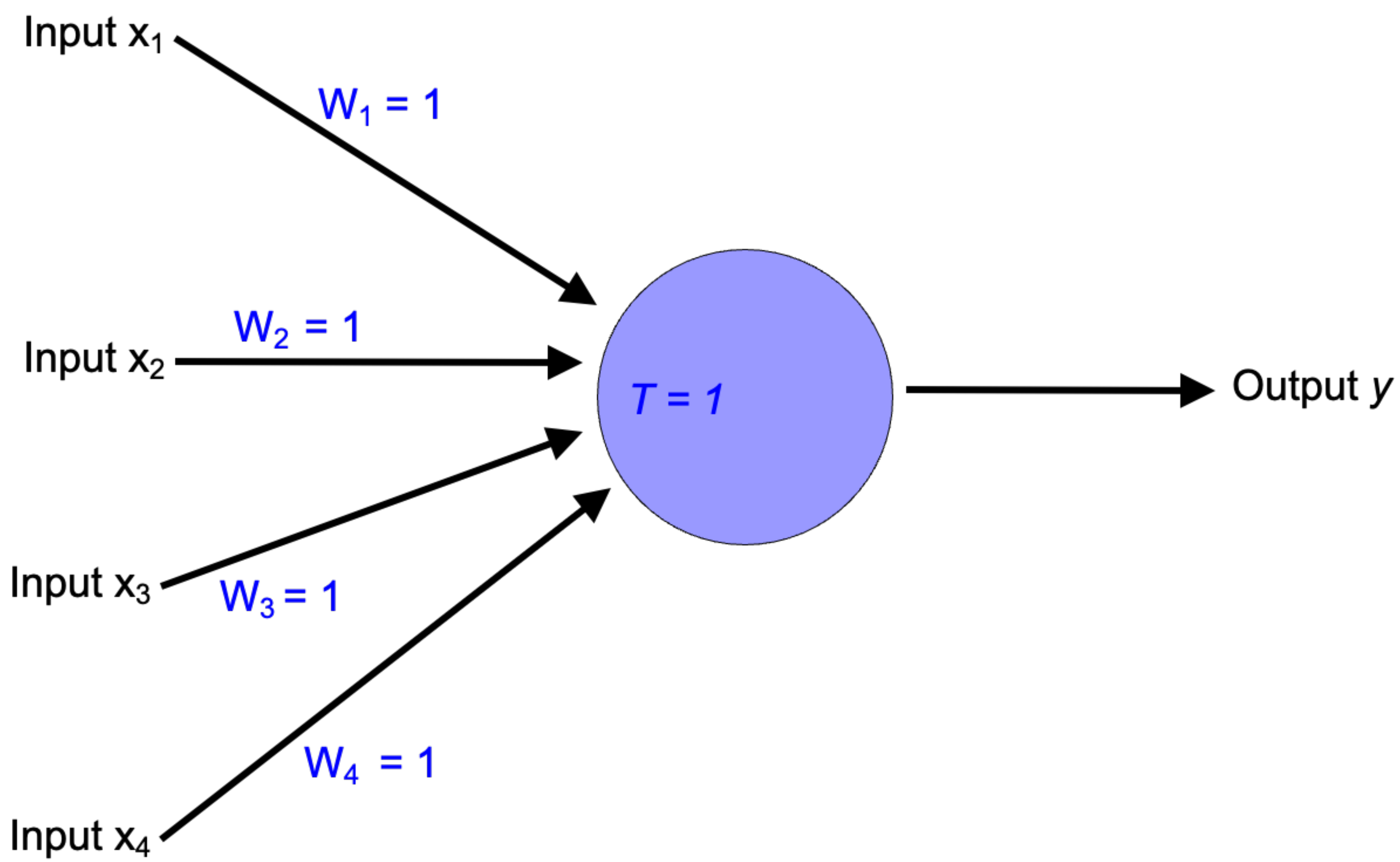
x_1	x_2	x_1 **or** x_2
0	0	0
0	1	1
1	0	1
1	1	1



OR



OR



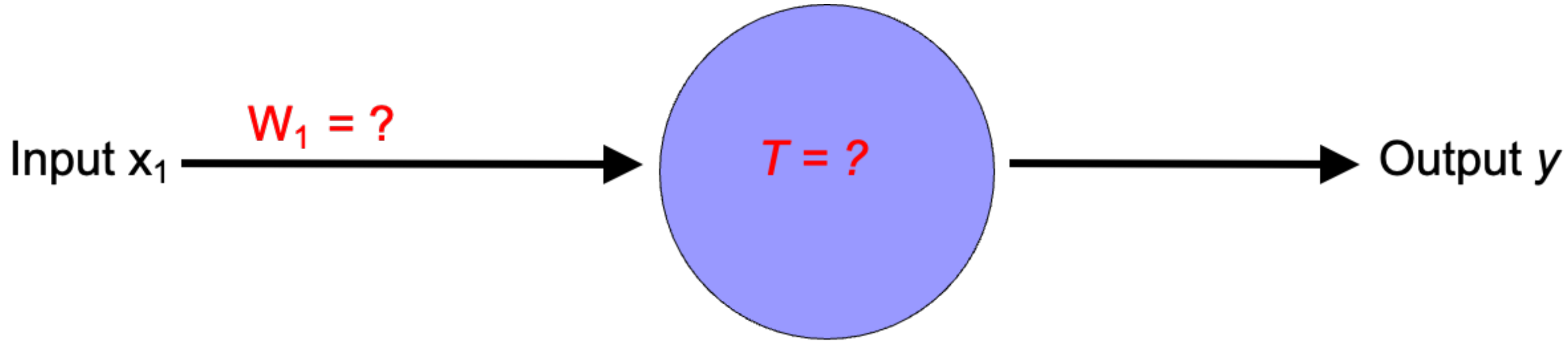
Inputs are either 0 or 1

NOT

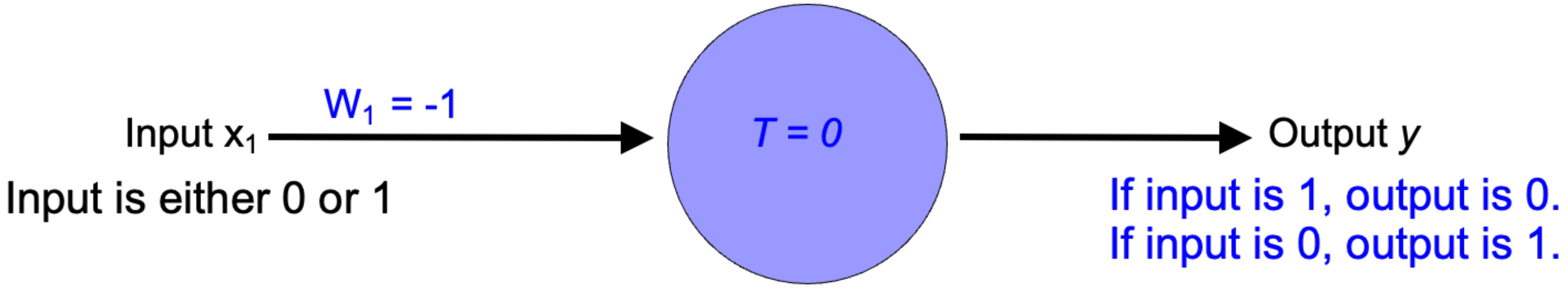
x_1	**not** x_1
0	1
1	0

NOT

x_1	**not** x_1
0	1
1	0

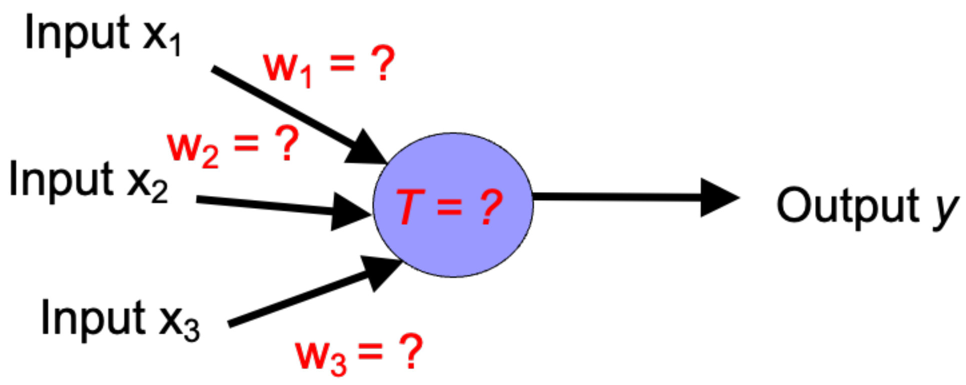


NOT

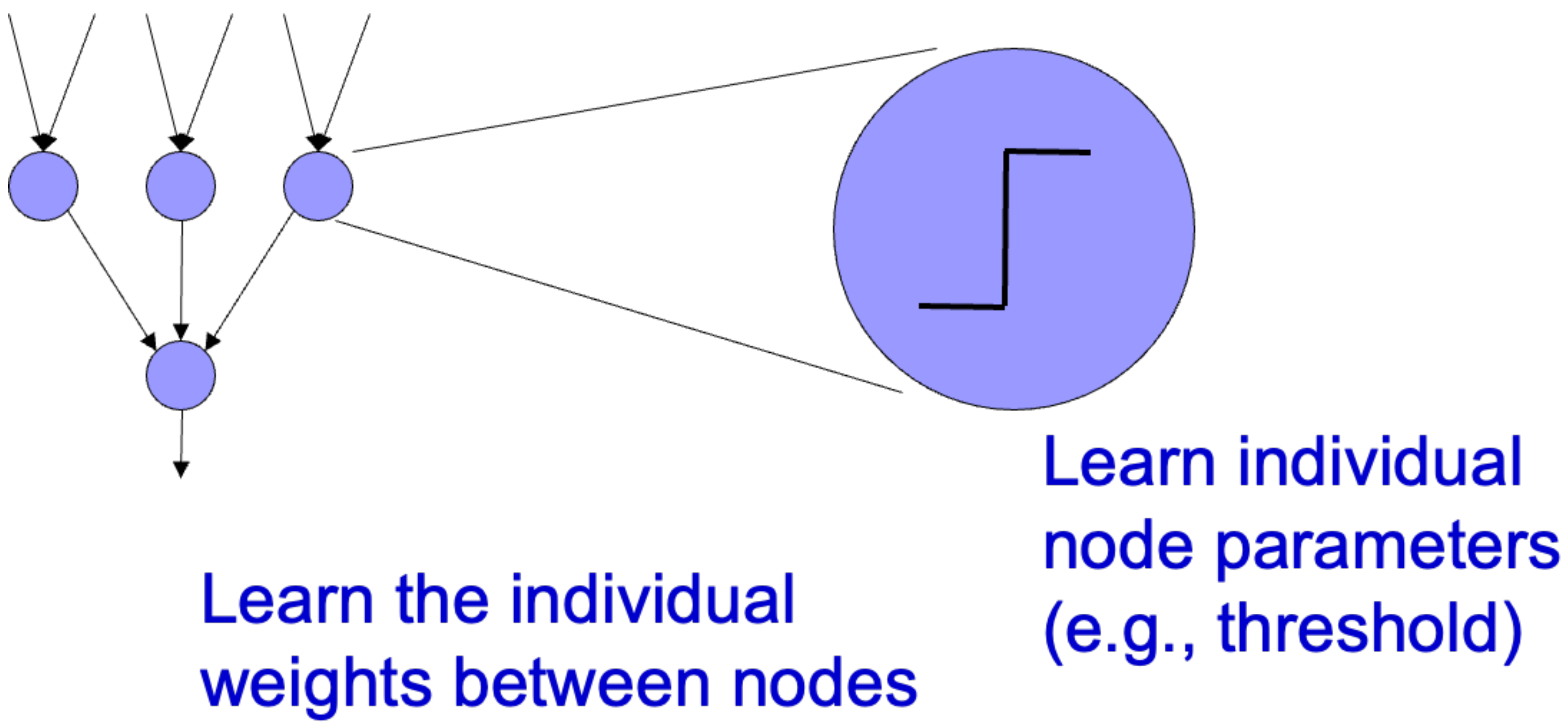

Input x_1
W_1 = -1
Input is either 0 or 1
T = 0
Output y
If input is 1, output is 0.
If input is 0, output is 1.

How about?

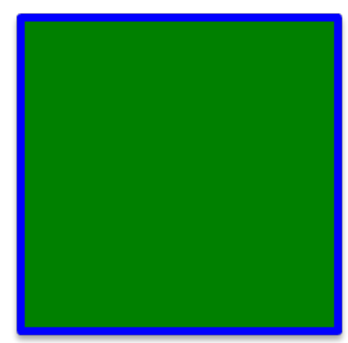
x_1	x_2	x_3	y
0	0	0	1
0	1	0	0
1	0	0	1
1	1	0	0
0	0	1	1
0	1	1	1
1	0	1	1
1	1	1	0



Training neural networks

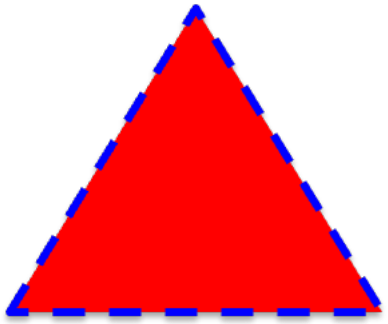


Positive or negative?



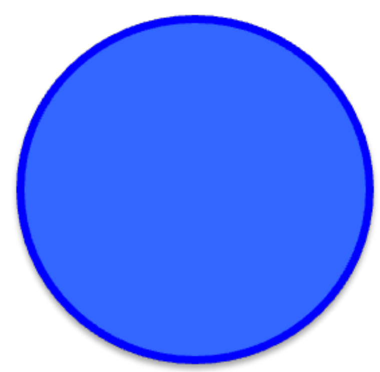
NEGATIVE

Positive or negative?



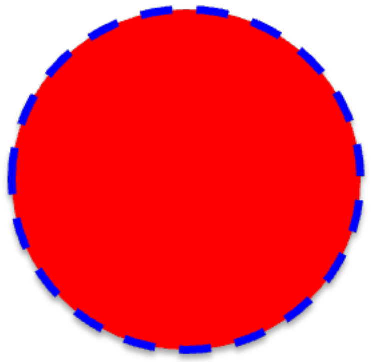
NEGATIVE

Positive or negative?



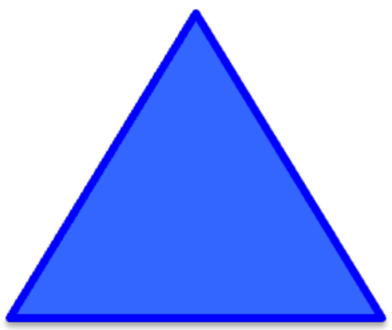
POSITIVE

Positive or negative?



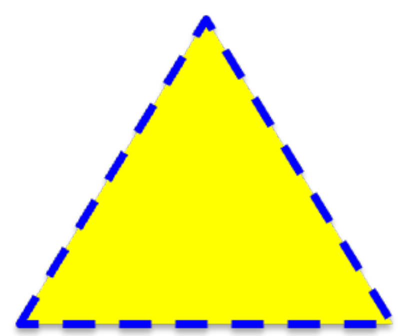
NEGATIVE

Positive or negative?



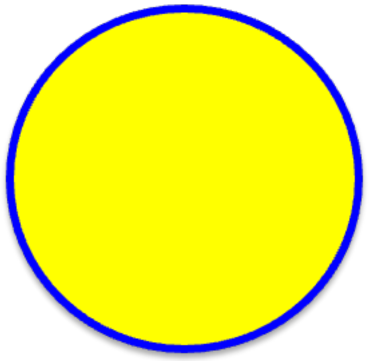
POSITIVE

Positive or negative?



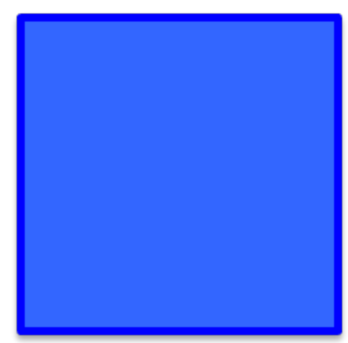
POSITIVE

Positive or negative?



NEGATIVE

Positive or negative?



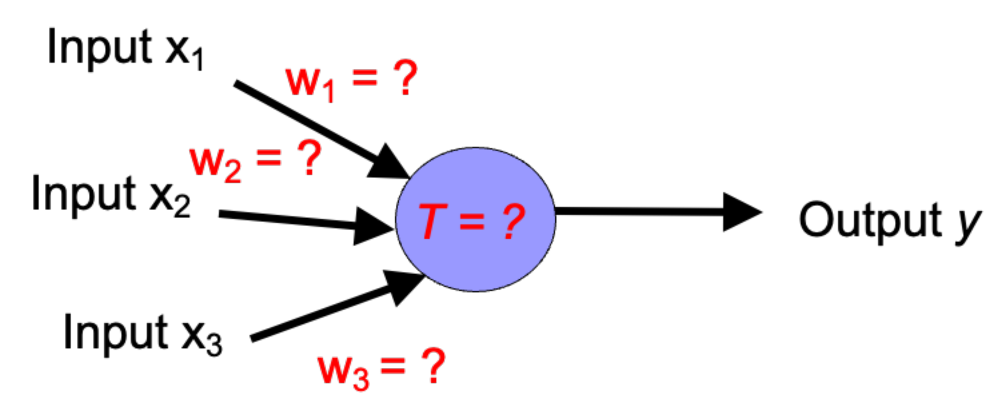
POSITIVE

A method to the madness

- blue = positive
- yellow triangles = positive
- all others negative

Training neural networks

x_1	x_2	x_3	y
0	0	0	1
0	1	0	0
1	0	0	1
1	1	0	0
0	0	1	1
0	1	1	1
1	0	1	1
1	1	1	0



- start with some initial weights and thresholds
- show examples repeatedly to NN
- update weights/thresholds by comparing NN output to actual output

Resources

- https://www.youtube.com/watch?v=2ePf9rue1Ao
- https://www.youtube.com/watch?v=oV74Najm6Nc
- https://bdtechtalks.com/2018/11/12/artificial-intelligence-winter-history/

Homework

- Assignment 6 (sign up for 03/20 presentations)