Neural Networks

1. Introduction

Spring 2021

Logistics: By now you must have…

- Already watched lecture 0 (logistics)
  – If not do so at once

- Done the zeroth HW

- Been to the course website
  – http://deeplearning.cs.cmu.edu
  – If you have not done so, please visit it at once

- Course objectives, logistics, quiz and homework policies, and grading policies, all have been explained there
  – In both, the logistics lecture and on the course page

- Please familiarize yourself with this information at once

Logistics: Part 2

- You should already have
  - Signed on to piazza
  - Verified you have access to canvas and autolab
  - Ensured you have AWS accounts setup
    - And tested out Google colab

- You have received a note on forming study groups
  - We recommend this; you learn better in teams than you do by yourself
  - **Please sign up for the study groups *immediately!!!!!!!!!!***

A minute for questions…

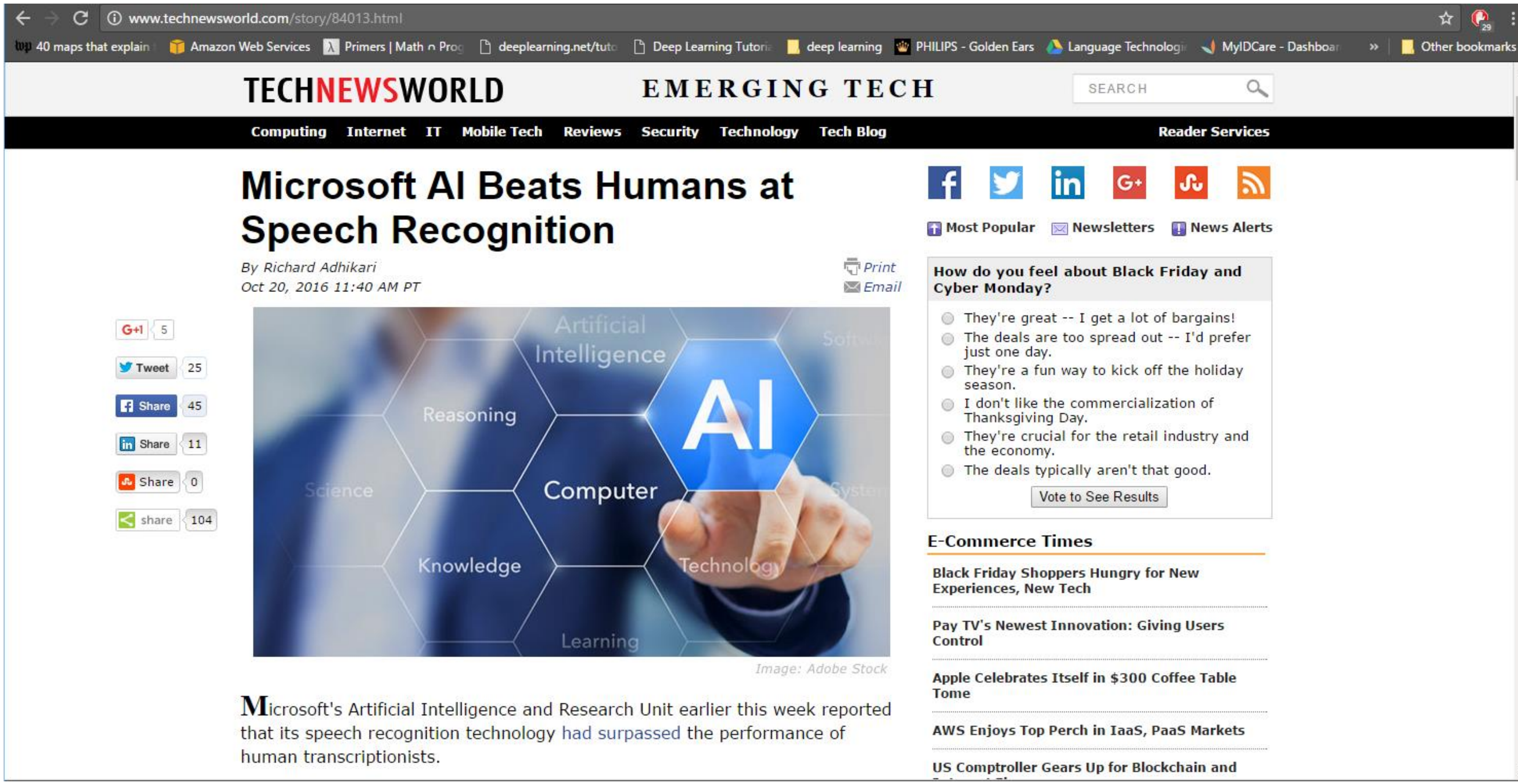
Caveat: Slide deck often have many "hidden" slides that will not be shown during the lecture, but will feature in your weekly quizzes

Neural Networks are taking over!

- Neural networks have become one of *the* main approaches to AI

- They have been successfully applied to various pattern recognition, prediction, and analysis problems
- In many problems they have established the state of the art
  - Often exceeding previous benchmarks by large margins
  - Sometimes solving problems you couldn't solve using earlier ML methods

Breakthroughs with neural networks



www.technewsworld.com/story/84013.html

TECHNEWSWORLD EMERGING TECH

Computing Internet IT Mobile Tech Reviews Security Technology Tech Blog Reader Services

Microsoft AI Beats Humans at Speech Recognition

By Richard Adhikari
Oct 20, 2016 11:40 AM PT

Image: Adobe Stock

Microsoft's Artificial Intelligence and Research Unit earlier this week reported that its speech recognition technology had surpassed the performance of human transcriptionists.

Most Popular Newsletters News Alerts

How do you feel about Black Friday and Cyber Monday?

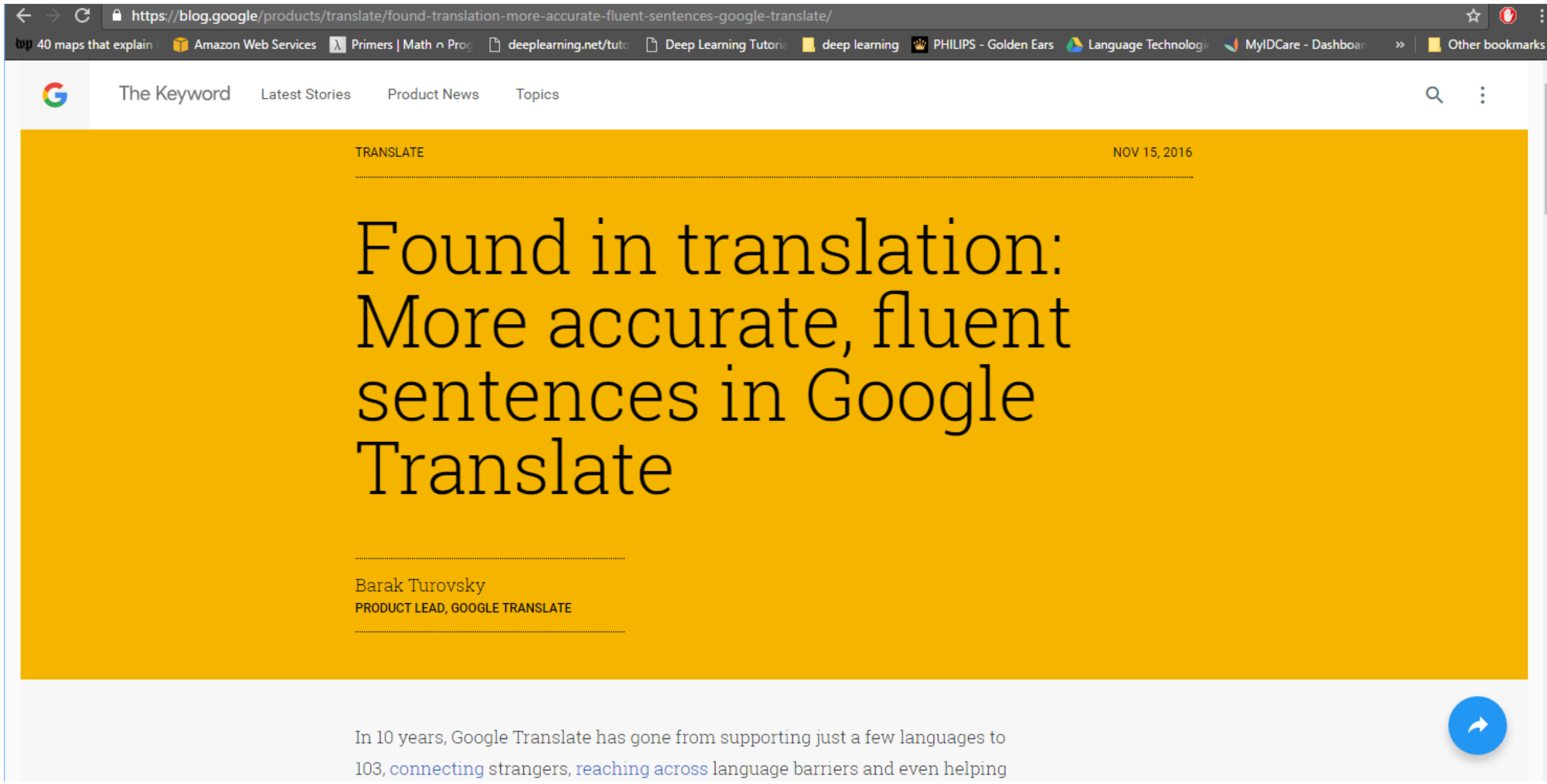
- They're great -- I get a lot of bargains!
- The deals are too spread out -- I'd prefer just one day.
- They're a fun way to kick off the holiday season.
- I don't like the commercialization of Thanksgiving Day.
- They're crucial for the retail industry and the economy.
- The deals typically aren't that good.

Vote to See Results

E-Commerce Times

- Black Friday Shoppers Hungry for New Experiences, New Tech
- Pay TV's Newest Innovation: Giving Users Control
- Apple Celebrates Itself in $300 Coffee Table Tome
- AWS Enjoys Top Perch in IaaS, PaaS Markets
- US Comptroller Gears Up for Blockchain and

Breakthrough with neural networks


https://blog.google/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/

The Keyword Latest Stories Product News Topics

TRANSLATE NOV 15, 2016

Found in translation: More accurate, fluent sentences in Google Translate

Barak Turovsky
PRODUCT LEAD, GOOGLE TRANSLATE

In 10 years, Google Translate has gone from supporting just a few languages to 103, connecting strangers, reaching across language barriers and even helping

Image segmentation and recognition

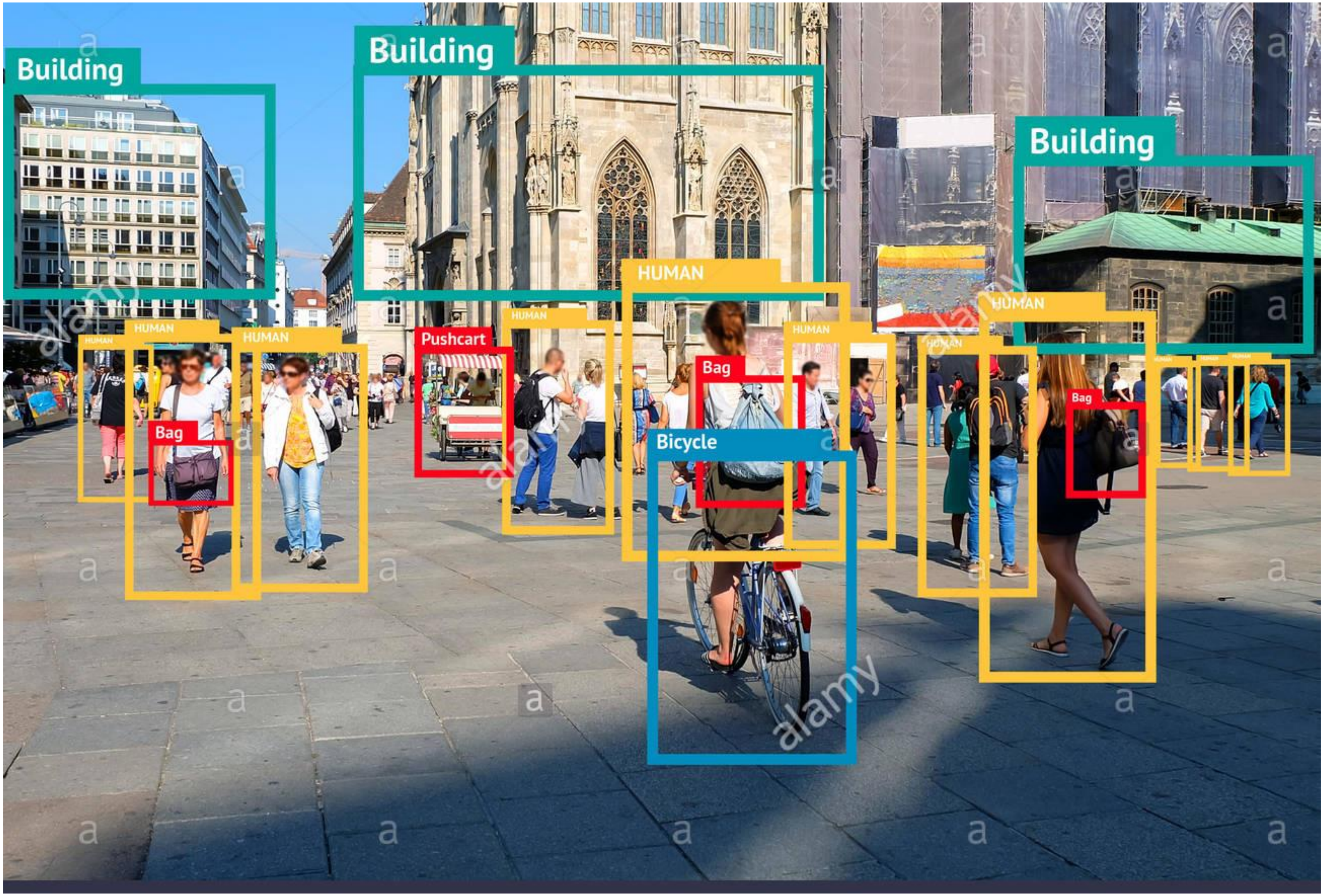
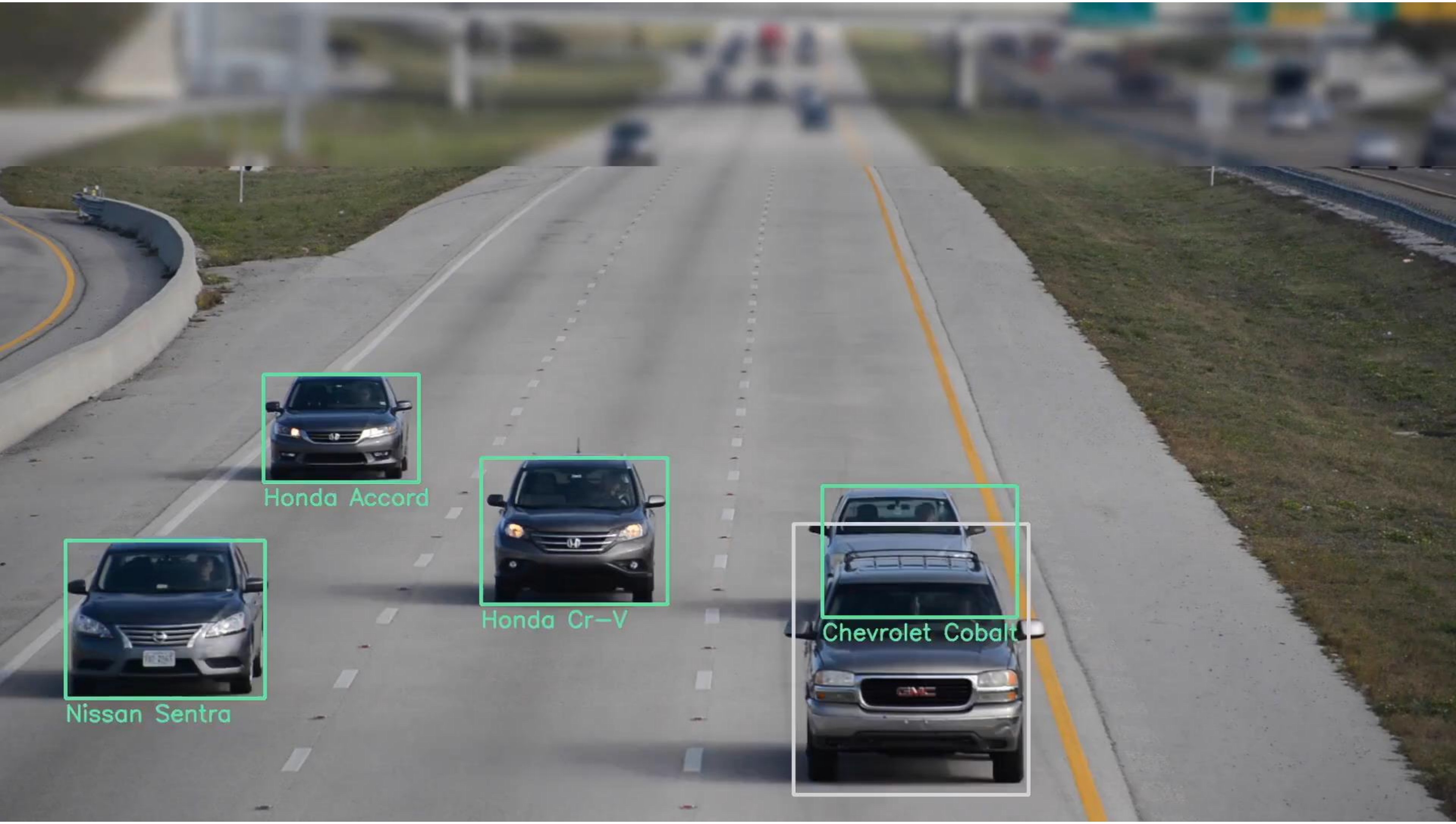
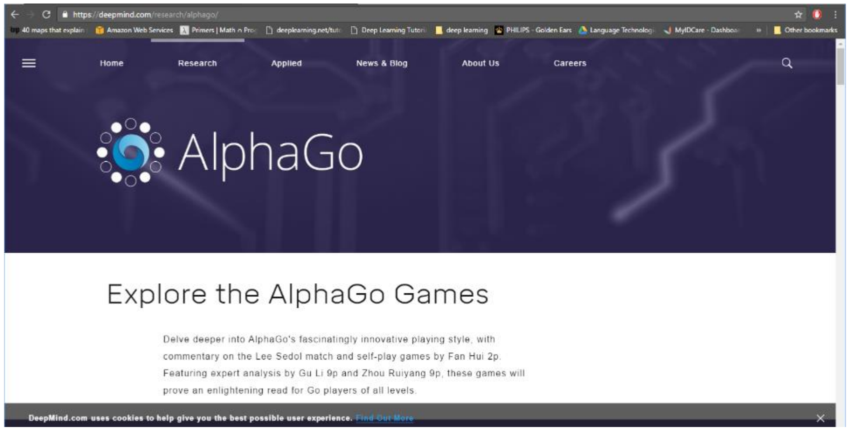


Image recognition



https://www.sighthound.com/technology/

Breakthroughs with neural networks



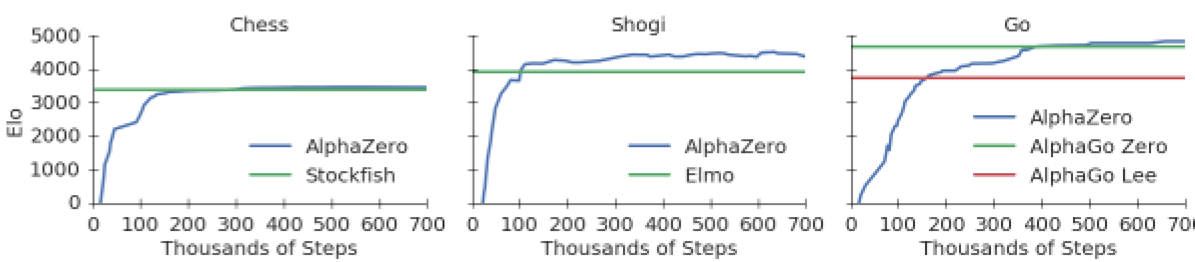
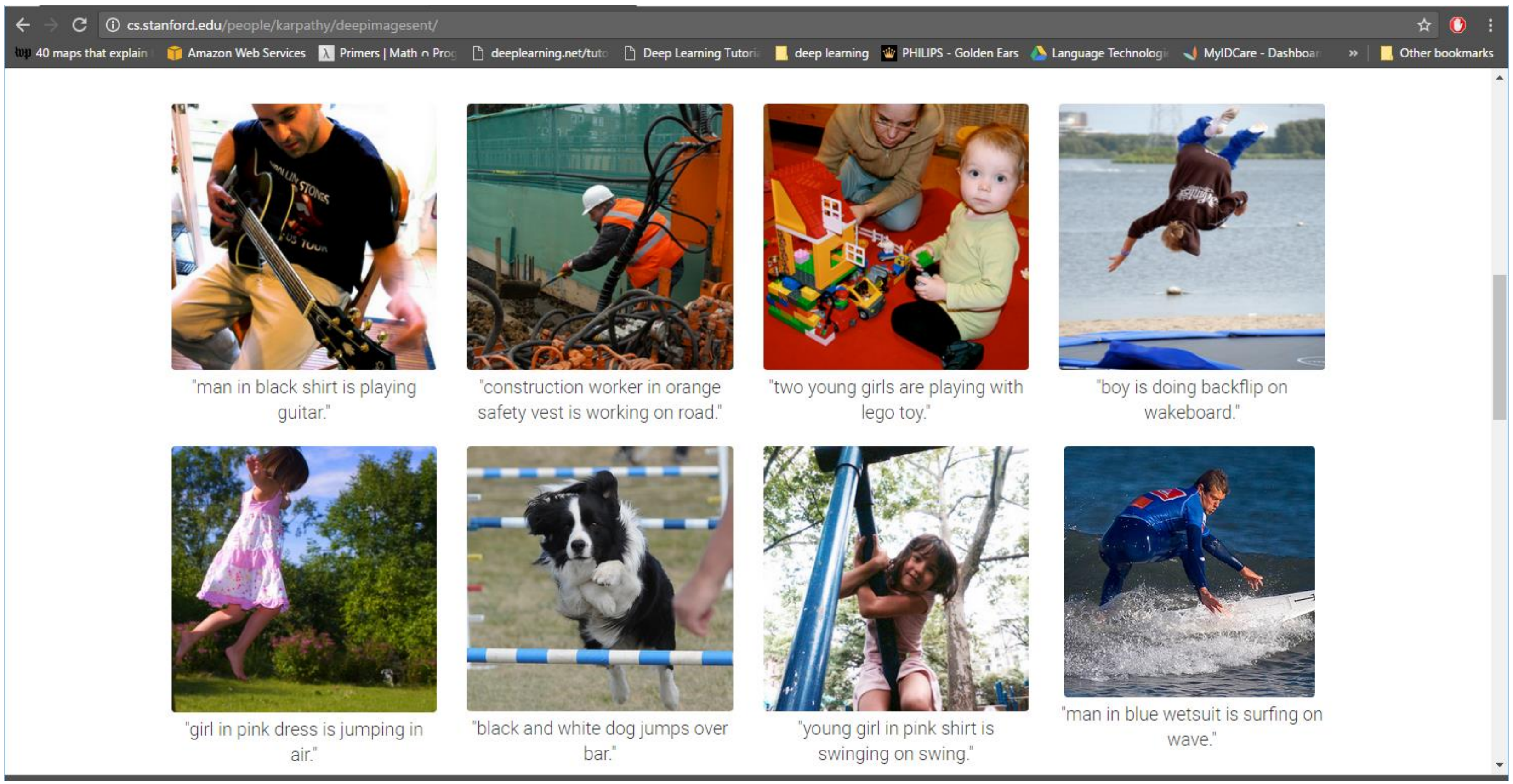


Figure 1: Training *AlphaZero* for 700,000 steps. Elo ratings were computed from evaluation games between different players when given one second per move. **a** Performance of *AlphaZero* in chess, compared to 2016 TCEC world-champion program *Stockfish*. **b** Performance of *AlphaZero* in shogi, compared to 2017 CSA world-champion program *Elmo*. **c** Performance of *AlphaZero* in Go, compared to *AlphaGo Lee* and *AlphaGo Zero* (20 block / 3 day) (*29*).

Success with neural networks



- Captions generated entirely by a neural network

Breakthroughs with neural networks

ThisPersonDoesNotExist.com uses AI to generate endless fake faces

Hit refresh to lock eyes with another imaginary stranger

By James Vincent | Feb 15, 2019, 7:38am EST

SHARE

A few sample faces — all completely fake — created by ThisPersonDoesNotExist.com

– https://www.theverge.com/tldr/2019/2/15/18226005/ai-generated-fake-people-portraits-thispersondoesnotexist-stylegan

Successes with neural networks

- And a variety of other problems:
  - From art to astronomy to healthcare..
  - and even predicting stock markets!

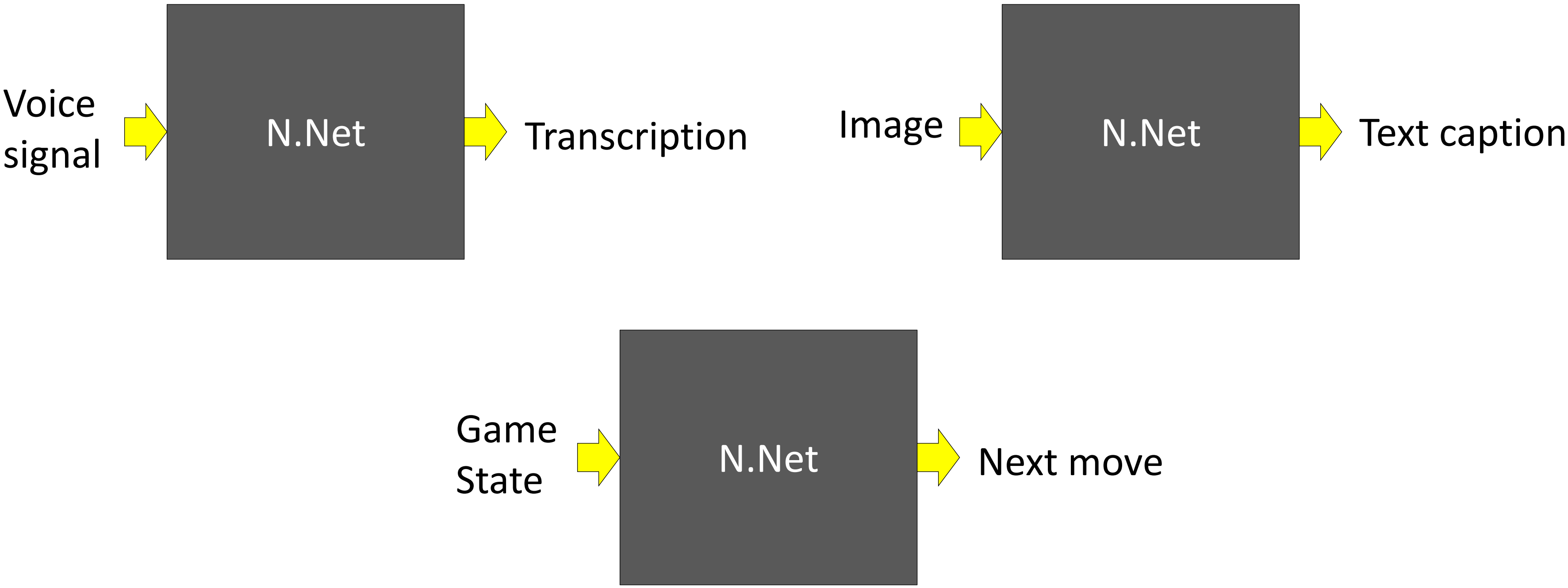
Neural nets can do anything!

Neural nets and the employment market

This guy didn't know about neural networks (a.k.a deep learning)

This guy learned about neural networks (a.k.a deep learning)

So what are neural networks??

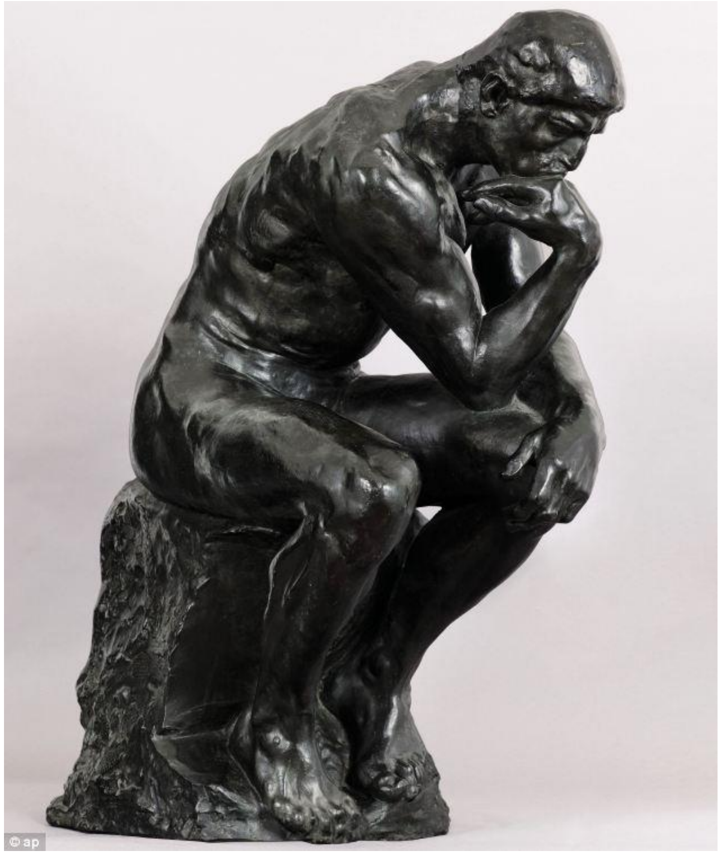


- What are these boxes?

So what are neural networks??

- It begins with this..

So what are neural networks??



"The Thinker!"
by Augustin Rodin

- Or even earlier.. with this..

The magical capacity of humans

- Humans can
  - Learn
  - Solve problems
  - Recognize patterns
  - Create
  - Cogitate
  - …

Dante!

- Worthy of emulation
- But how do humans “work“?

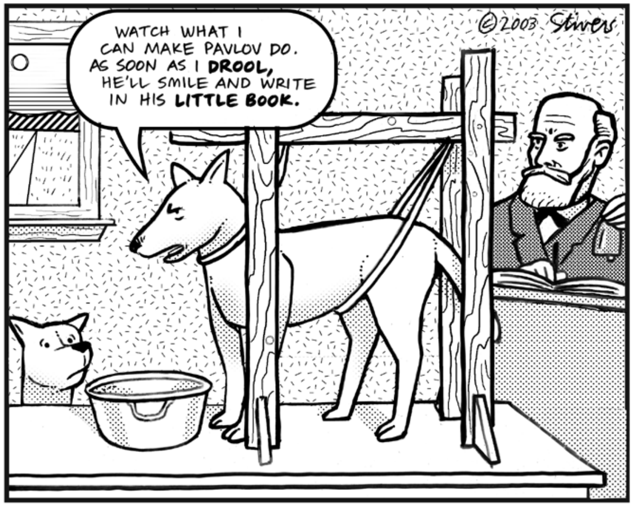
Cognition and the brain..

- “If the brain was simple enough to be understood - we would be too simple to understand it!”
  – Marvin Minsky

Early Models of Human Cognition

- Associationism
  - Humans learn through association
- 400BC-1900AD: Plato, David Hume, Ivan Pavlov..

What are "Associations"



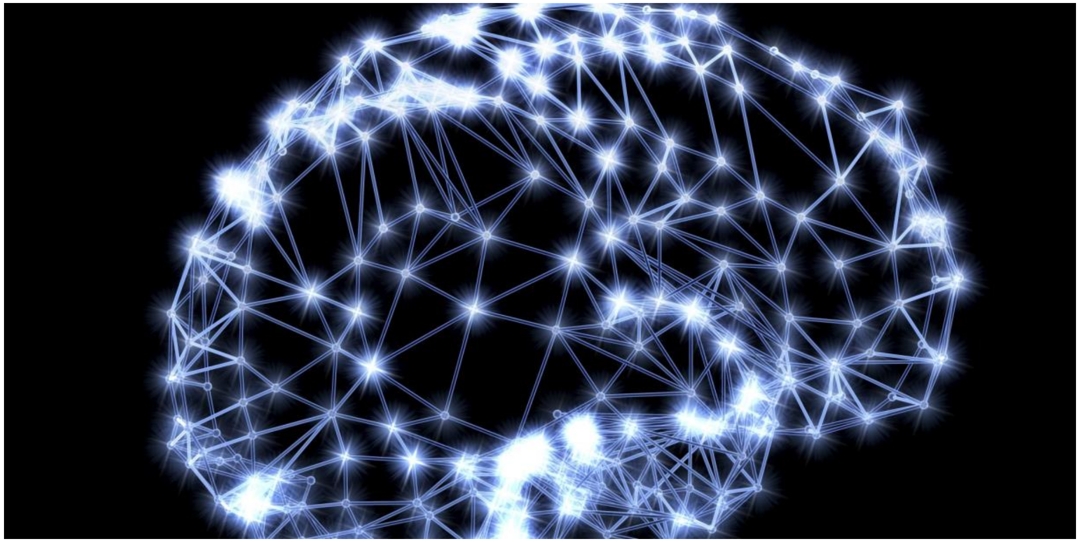
- Lightning is generally followed by thunder
  - Ergo – "hey here's a bolt of lightning, we're going to hear thunder"
  - Ergo – "We just heard thunder; did someone get hit by lightning"?

- Association!

- **But where are the associations stored??**

- **And how?**

Observation: ***The Brain***

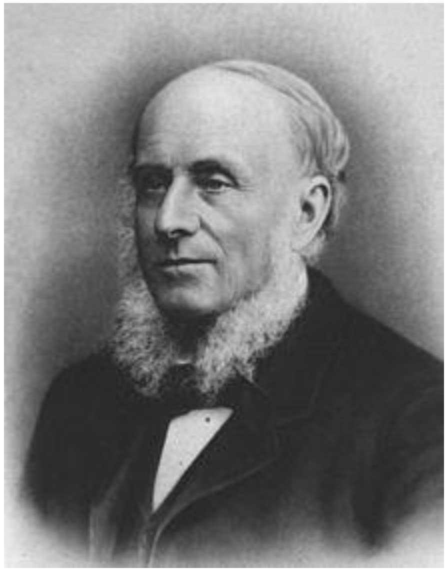


- Mid 1800s: The brain is a mass of interconnected neurons

Brain: Interconnected Neurons

- Many neurons connect *in* to each neuron
- Each neuron connects *out* to many neurons

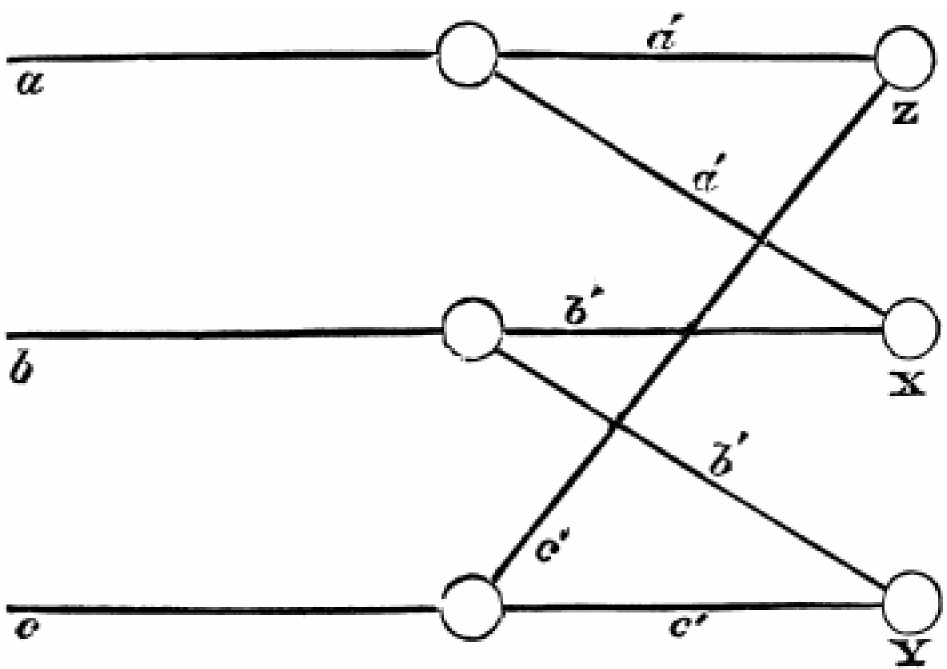
Enter ***Connectionism***



- Alexander Bain, philosopher, psychologist, mathematician, logician, linguist, professor
- 1873: The information is in the ***connections***
  - *Mind and body* (1873)

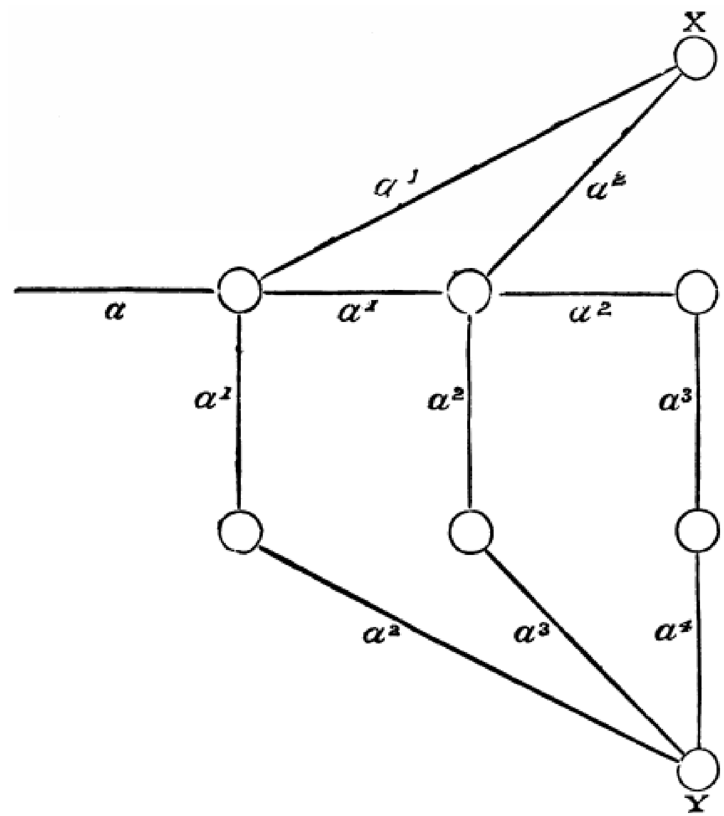
Bain's Idea 1: Neural Groupings

- Neurons excite and stimulate each other
- Different combinations of inputs can result in different outputs



Bain's Idea 1: Neural Groupings

- Different intensities of activation of A lead to the differences in when X and Y are activated



- Even proposed a learning mechanism..

Bain's Idea 2: Making Memories

- "when two impressions concur, or closely succeed one another, the nerve-currents find some bridge or place of continuity, better or worse, according to the abundance of nerve-matter available for the transition."

- Predicts "Hebbian" learning (three quarters of a century before Hebb!)

Bain's Doubts

- *"**The fundamental cause of the trouble is that in the modern world the stupid are cocksure while the intelligent are full of doubt**."*
  - Bertrand Russell

- In 1873, Bain postulated that there must be one million neurons and 5 billion connections relating to 200,000 "acquisitions"

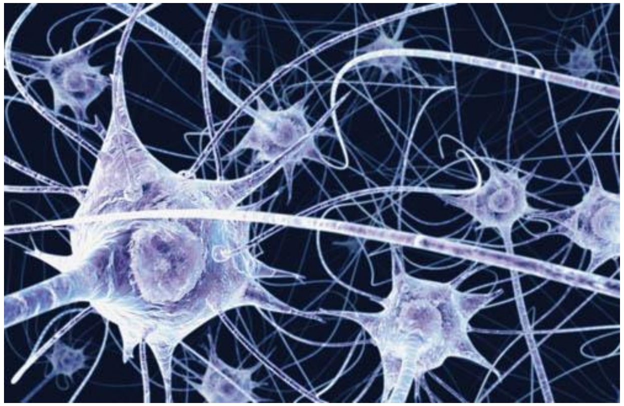
- In 1883, Bain was concerned that he hadn't taken into account the number of "partially formed associations" and the number of neurons responsible for recall/learning

- By the end of his life (1903), recanted all his ideas!
  - Too complex; the brain would need too many neurons and connections

Connectionism lives on..

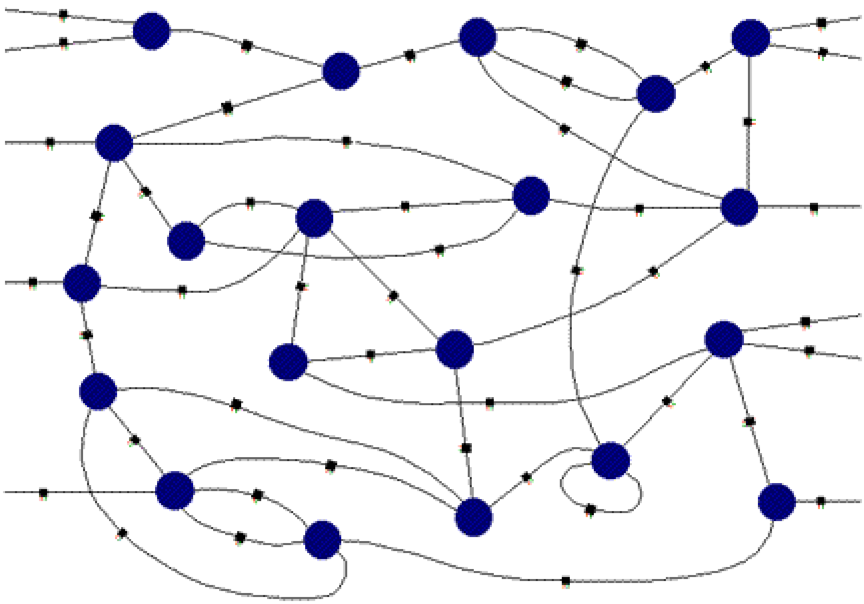
- The human brain is a connectionist machine
  - Bain, A. (1873). *Mind and body. The theories of their relation.* London: Henry King.
  - Ferrier, D. (1876). *The Functions of the Brain.* London: Smith, Elder and Co

- Neurons connect to other neurons. The processing/capacity of the brain is a function of these connections



- Connectionist machines emulate this structure

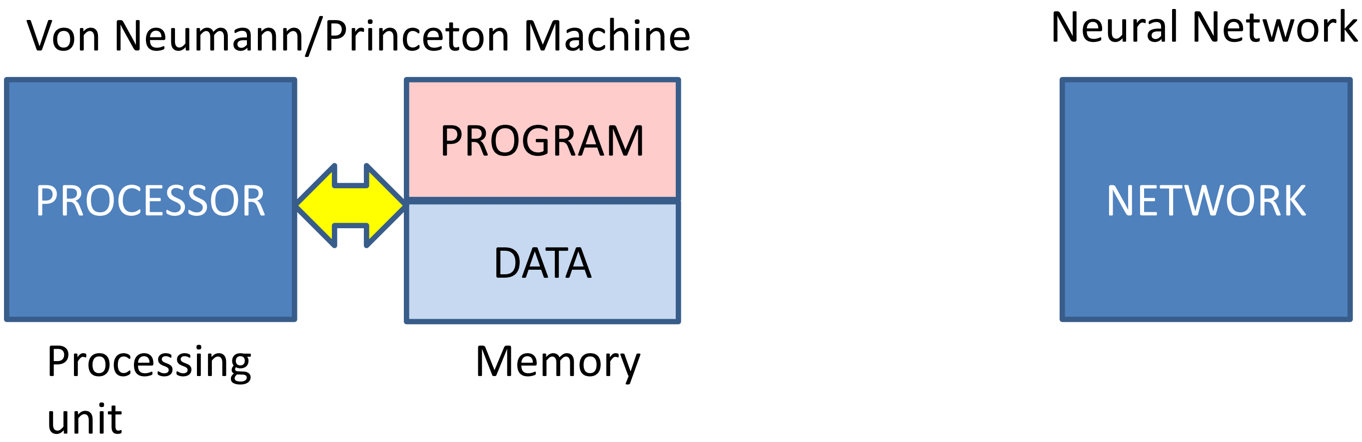
Connectionist Machines



- Network of processing elements
- **All world knowledge is stored in the *connections* between the elements**

Connectionist Machines

- Neural networks are *connectionist* machines
  - As opposed to Von Neumann Machines

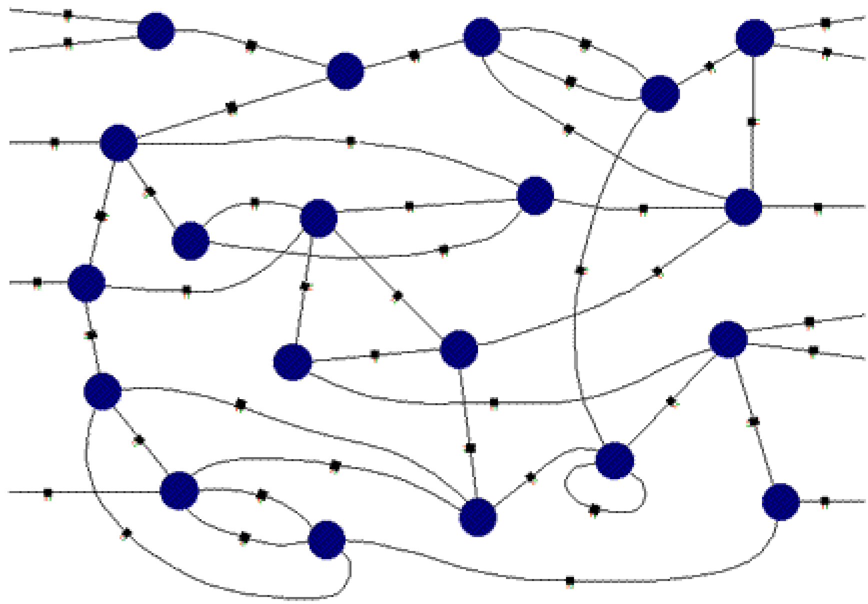


- The machine has many non-linear processing units
  - The program is the connections between these units
    - Connections may also define memory

Recap

- Neural network based AI has taken over most AI tasks
- Neural networks originally began as computational models of the brain
  - Or more generally, models of cognition
- The earliest model of cognition was *associationism*
- The more recent model of the brain is *connectionist*
  - Neurons connect to neurons
  - The workings of the brain are encoded in these connections
- Current neural network models are *connectionist machines*

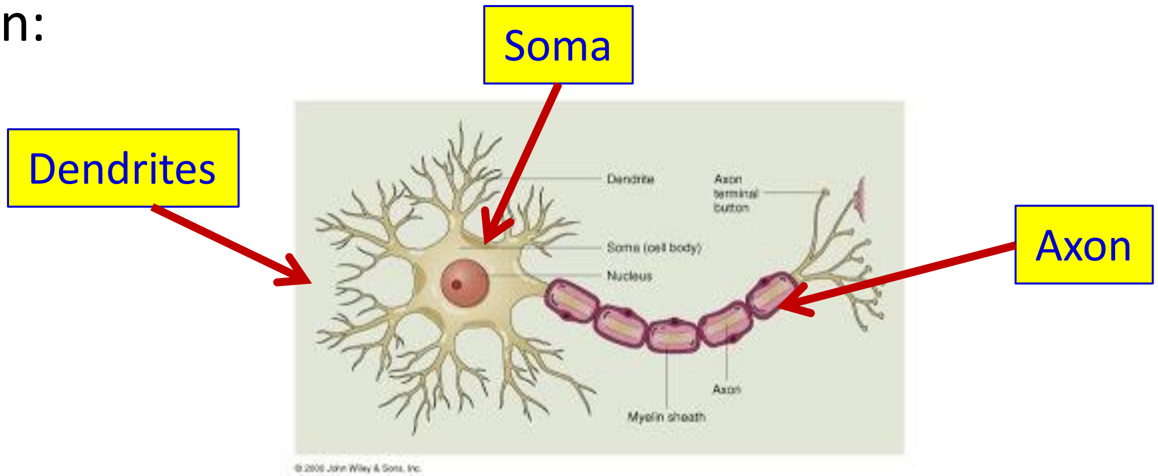
Connectionist Machines



- Network of processing elements
  - All world knowledge is stored in the *connections* between the elements
- *But what are the individual elements?*

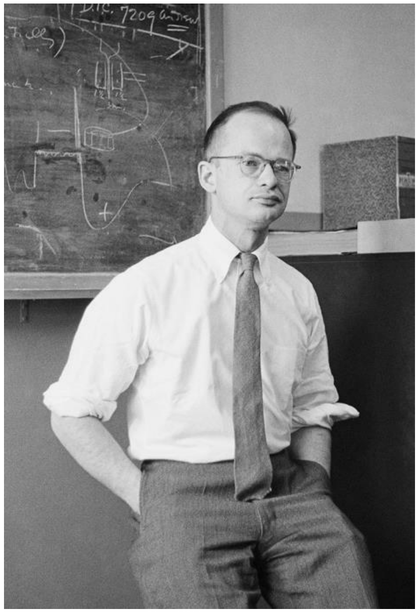
Modelling the brain

- What are the units?
- A neuron:



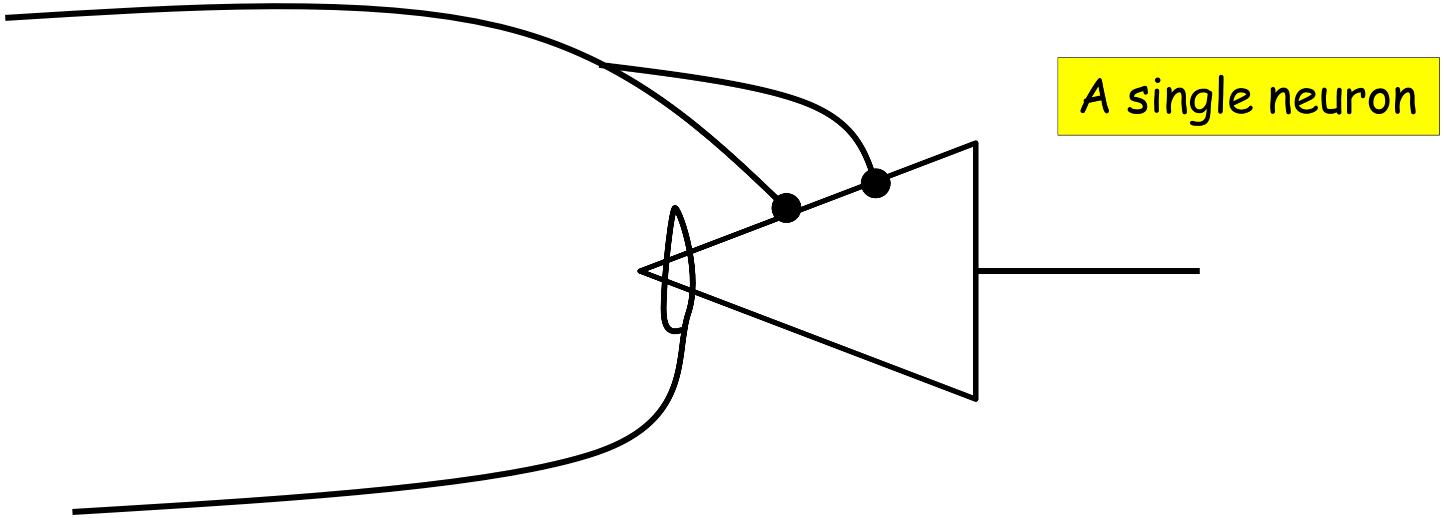
- Signals come in through the dendrites into the Soma
- A signal goes out via the axon to other neurons
  - Only one axon per neuron
- Factoid that may only interest me: Neurons do not undergo cell division
  - Neurogenesis occurs from neuronal stem cells, and is minimal after birth

McCulloch and Pitts



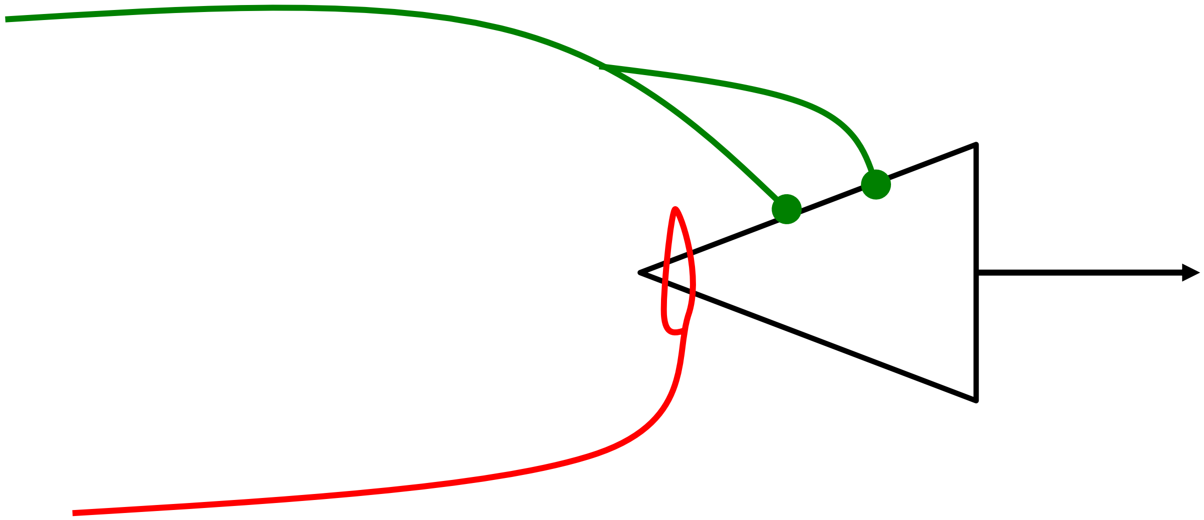
- The Doctor and the Hobo..
  - Warren McCulloch: Neurophysiologist
  - Walter Pitts: Homeless wannabe logician who arrived at his door

The McCulloch and Pitts model



- A mathematical model of a neuron
  - McCulloch, W.S. & Pitts, W.H. (1943). A Logical Calculus of the Ideas Immanent in Nervous Activity, Bulletin of Mathematical Biophysics, 5:115-137, 1943
    - Pitts was only 20 years old at this time

Synaptic Model



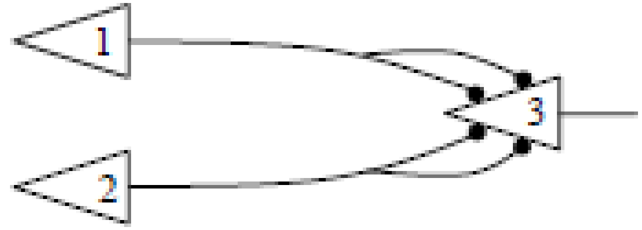
- ***Excitatory synapse:*** Transmits weighted input to the neuron

- ***Inhibitory synapse:*** Any signal from an inhibitory synapse prevents neuron from firing
  - The activity of any inhibitory synapse absolutely prevents excitation of the neuron at that time.
    - Regardless of other inputs

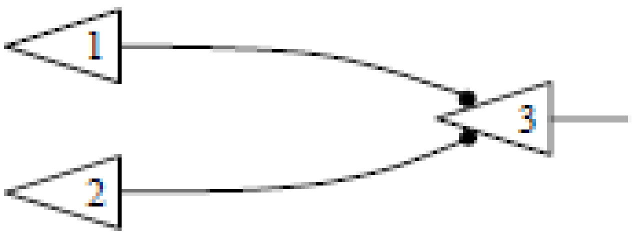
Simple "networks" of neurons can perform Boolean operations

Boolean Gates

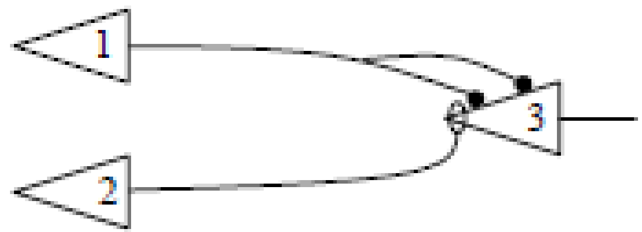
$N_2(t)) \leftrightarrow N_1(t\text{-}1)$
net for temporal predecessor



$N_3(t) \leftrightarrow N_1(t\text{-}1) \vee N_2(t\text{-}1)$
net for disjunction



$N_3(t) \leftrightarrow N_1(t\text{-}1) \& N_2(t\text{-}1)$
net for conjunction



$N_3(t) \leftrightarrow N_1(t\text{-}1) \& {\sim} N_2(t\text{-}1)$
net for conjunction and negation

Figure 1. Diagrams of McCulloch and Pitts nets. In order to send an output pulse, each neuron must receive two excitory inputs and no inhibitory inputs. Lines ending in a dot represent excitatory connections; lines ending in a hoop represent inhibitory connections.

Complex Percepts & Inhibition in action

They can even create illusions of "perception"

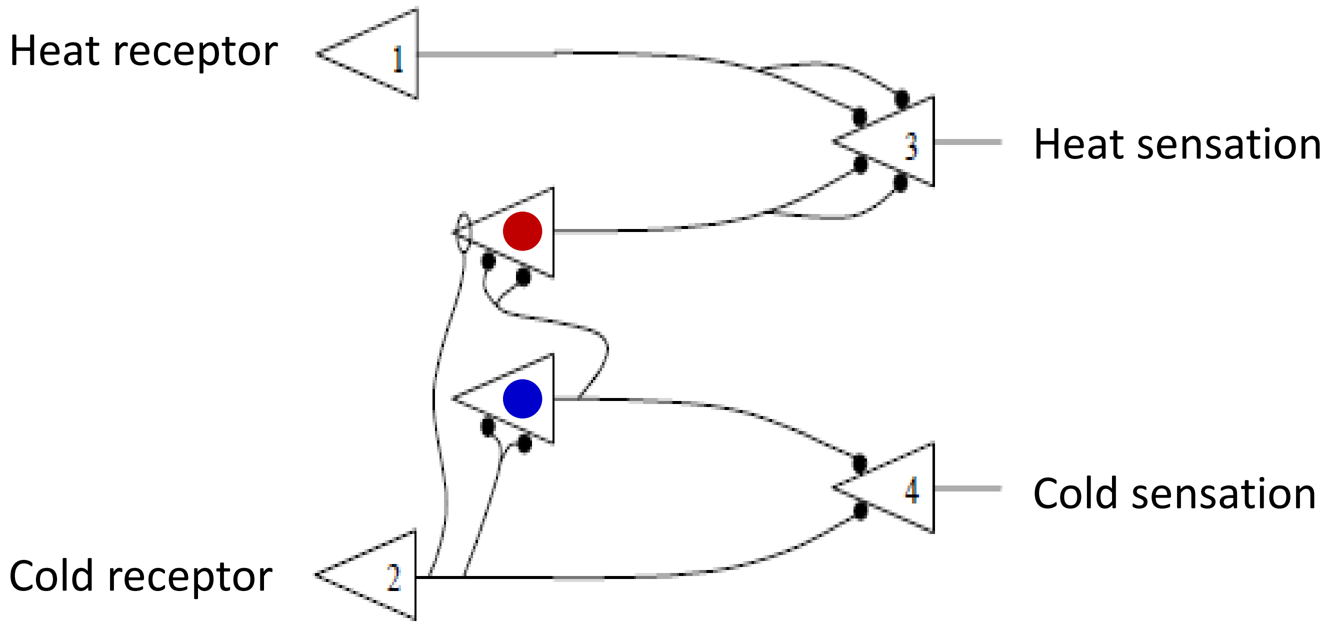


Figure 2. Net explaining the heat illusion. Neuron 3 (heat sensation) fires if and only if it receives two inputs, represented by the lines terminating on its body. This happens when either neuron 1 (heat reception) fires or neuron 2 (cold reception) fires once and then immediately stops firing. When neuron 2 fires twice in a row, the intermediate (unnumbered) neurons excite neuron 4 rather than neuron 3, generating a sensation of cold.

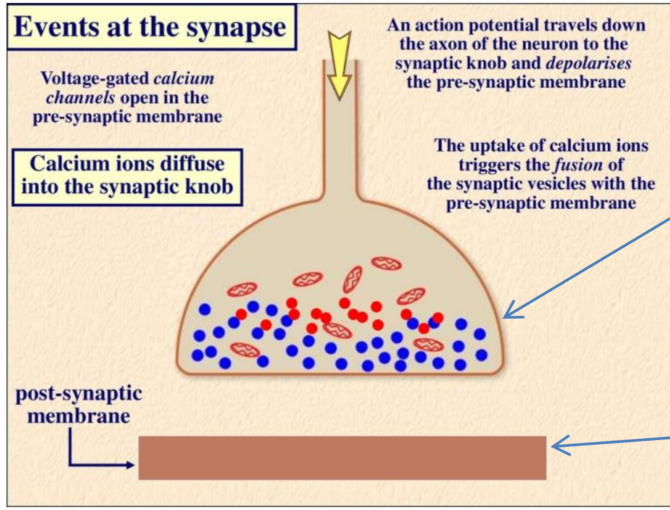
Criticisms

- They claimed that their nets
  - should be able to compute a small class of functions
  - also if tape is provided their nets can compute a richer class of functions.
    - additionally they will be equivalent to Turing machines
    - Dubious claim that they're Turing complete
  - They didn't prove any results themselves
- Didn't provide a learning mechanism..

Donald Hebb

- "Organization of behavior", 1949
- A learning mechanism:
  - "When an axon of cell *A* is near enough to excite a cell *B* and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that *A*'s efficiency, as one of the cells firing *B*, is increased."
    - As *A* repeatedly excites *B*, its *ability* to excite *B* improves
  - *Neurons that fire together wire together*

Hebbian Learning



- If neuron x repeatedly triggers neuron y, the synaptic knob connecting x to y gets larger
- In a mathematical model:

$$w_{xy} = w_{xy} + \eta xy$$

  – Weight of the connection from input neuron x to output neuron y

- This simple formula is actually the basis of many learning algorithms in ML

Hebbian Learning

- **Fundamentally unstable**
  – Stronger connections will enforce themselves
  – No notion of "competition"
  – No *reduction* in weights
  – Learning is unbounded

- Number of later modifications, allowing for weight normalization, forgetting etc.
  – E.g. Generalized Hebbian learning, aka Sanger's rule

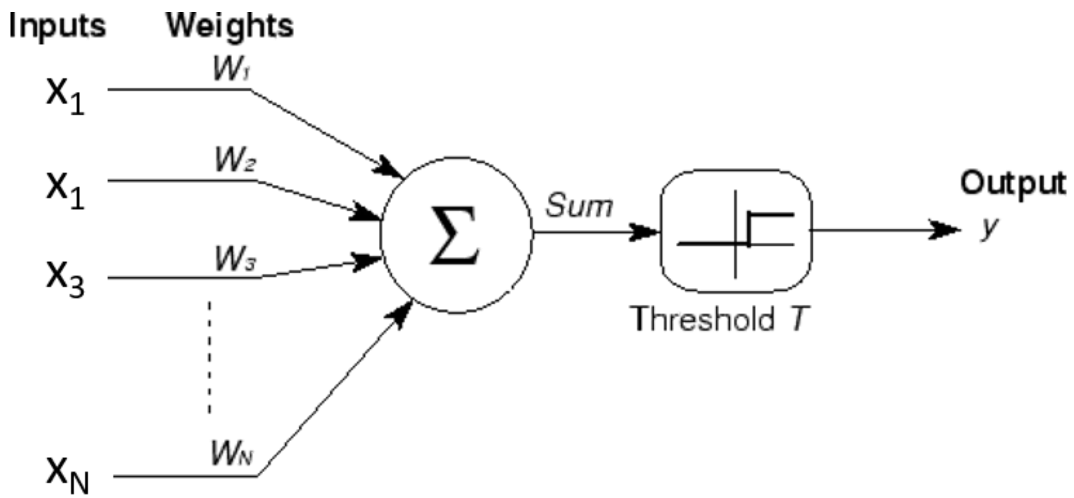
$$w_{ij} = w_{ij} + \eta y_j \left(x_i - \sum_{k=1}^{j} w_{ik} y_k \right)$$

  – The contribution of an input is incrementally *distributed* over multiple outputs..

A better model

- Frank Rosenblatt
  - Psychologist, Logician
  - Inventor of the solution to everything, aka the Perceptron (1958)

Perceptron: Simplified model

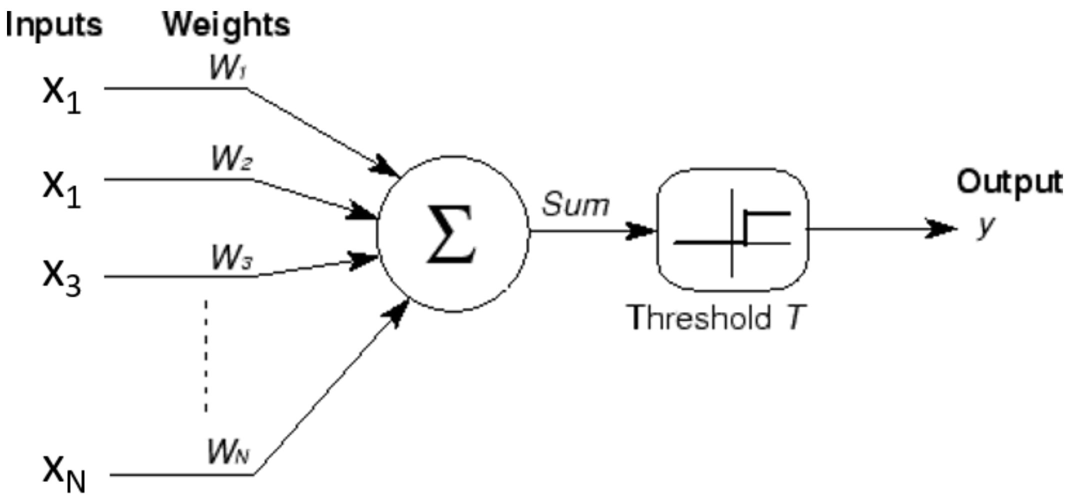


- Number of inputs combine linearly
  - Threshold logic: Fire if combined input exceeds threshold

$$Y = \begin{cases} 1 \;\; if \; \sum_i w_i x_i - T \geq 0 \\ 0 \qquad\qquad\qquad else \end{cases}$$

The Universal Model

- Originally assumed could represent *any* Boolean circuit and perform any logic
  - *"the embryo of an electronic computer that [the Navy] expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence,"* New York Times (8 July) 1958
  - *"Frankenstein Monster Designed by Navy That Thinks,"* Tulsa, Oklahoma Times 1958



Also provided a learning algorithm

$$\mathbf{w} = \mathbf{w} + \eta(d(\mathbf{x}) - y(\mathbf{x}))\mathbf{x}$$

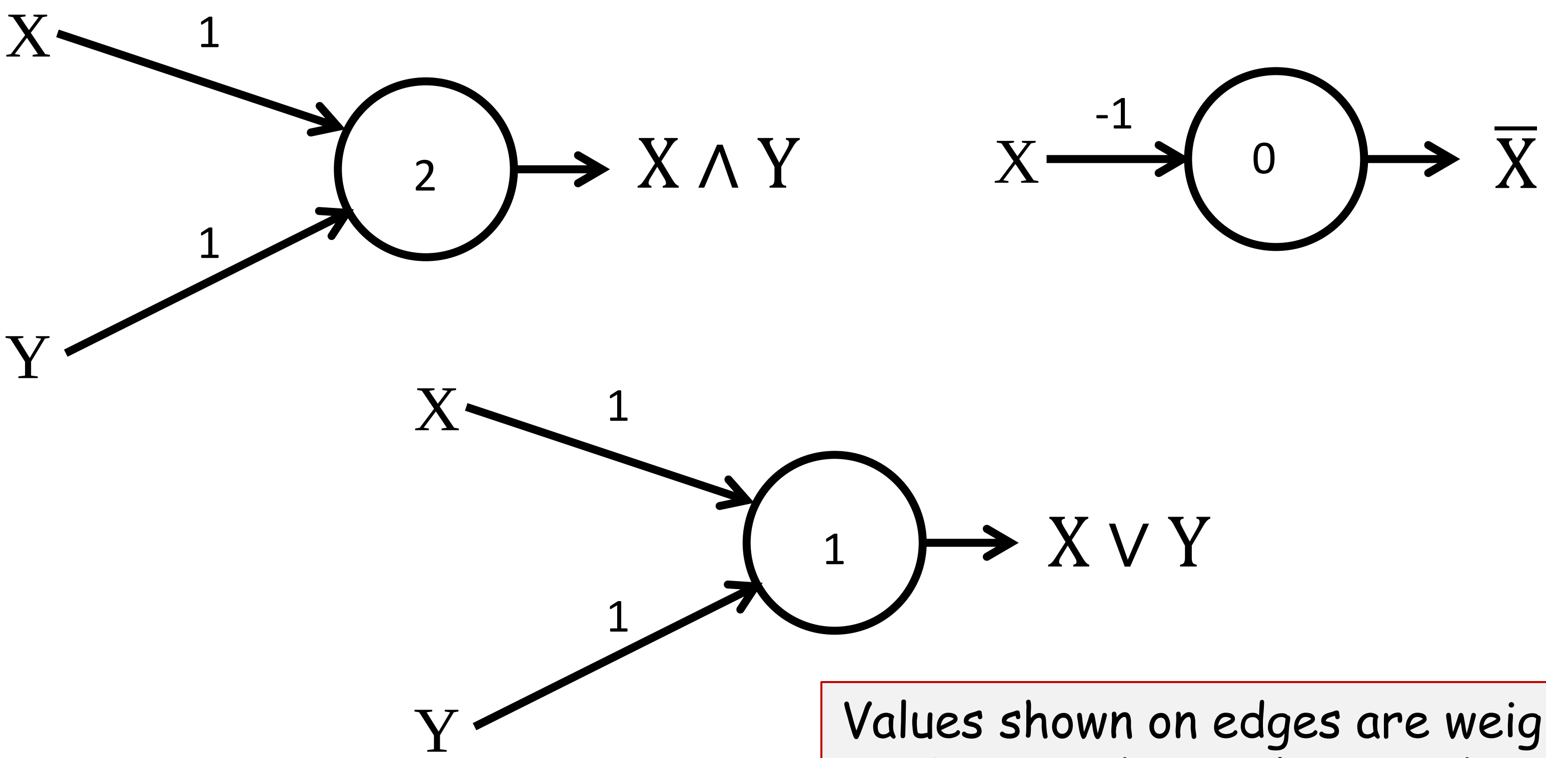
Sequential Learning:

$d(x)$ is the desired output in response to input $\mathbf{x}$

$y(x)$ is the actual output in response to $\mathbf{x}$

- Boolean tasks
- Update the weights whenever the perceptron output is wrong
  - Update the weight by the product of the input and the *error* between the desired and actual outputs
- Proved convergence for linearly separable classes

Perceptron

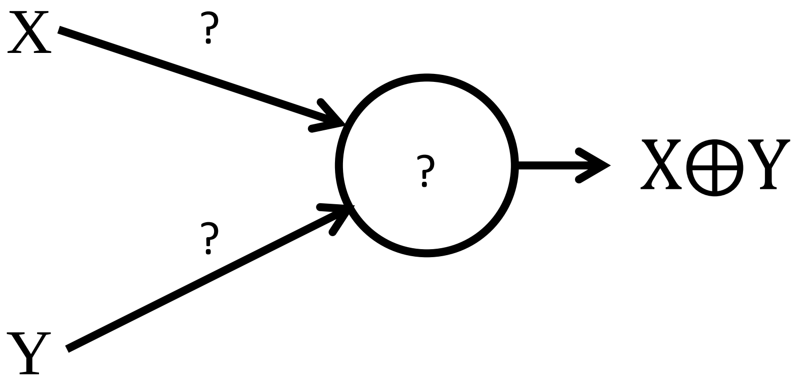


Values shown on edges are weights, numbers in the circles are thresholds

- Easily shown to mimic any Boolean gate
- But…

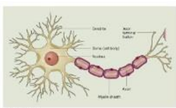
Perceptron

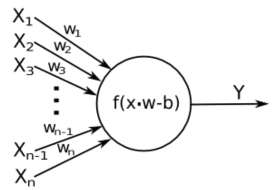
No solution for XOR!
Not universal!

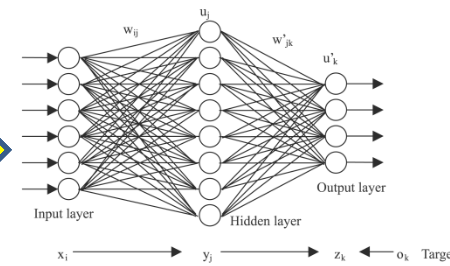


- Minsky and Papert, 1968

A single neuron is not enough

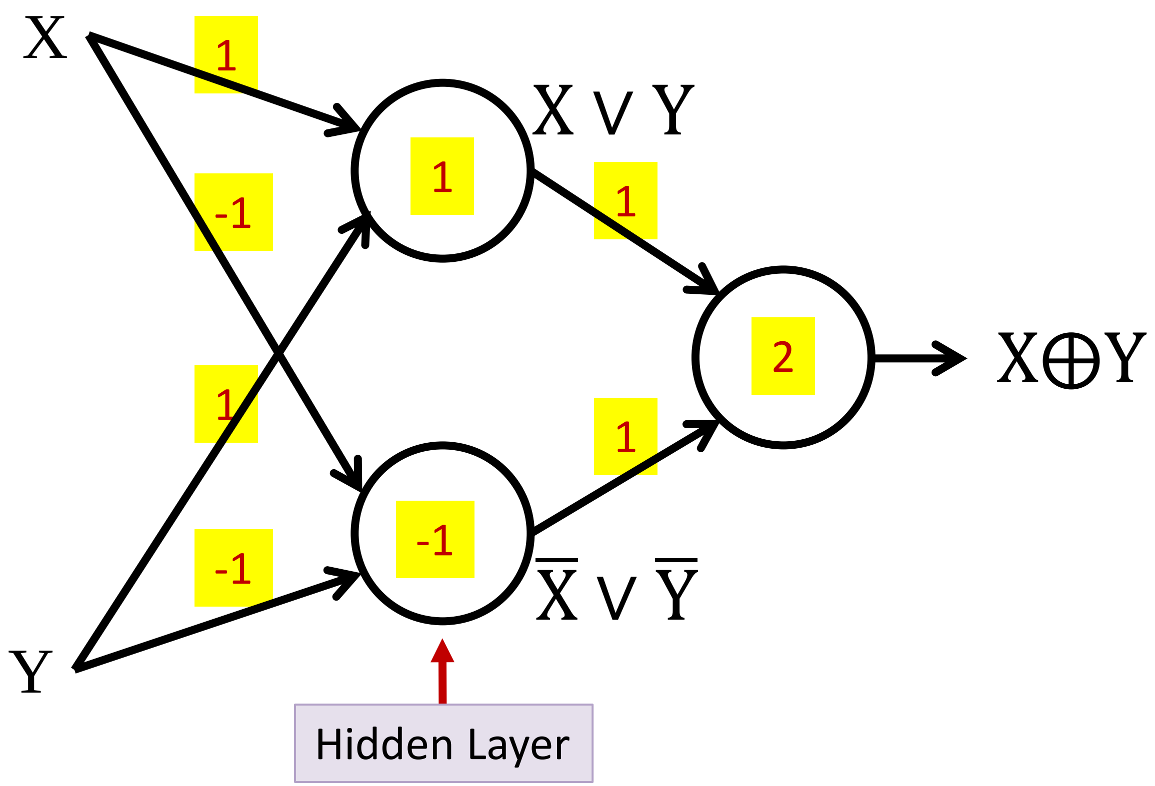






- Individual elements are weak computational elements
  - Marvin Minsky and Seymour Papert, 1969, *Perceptrons: An Introduction to Computational Geometry*
- *Networked* elements are required

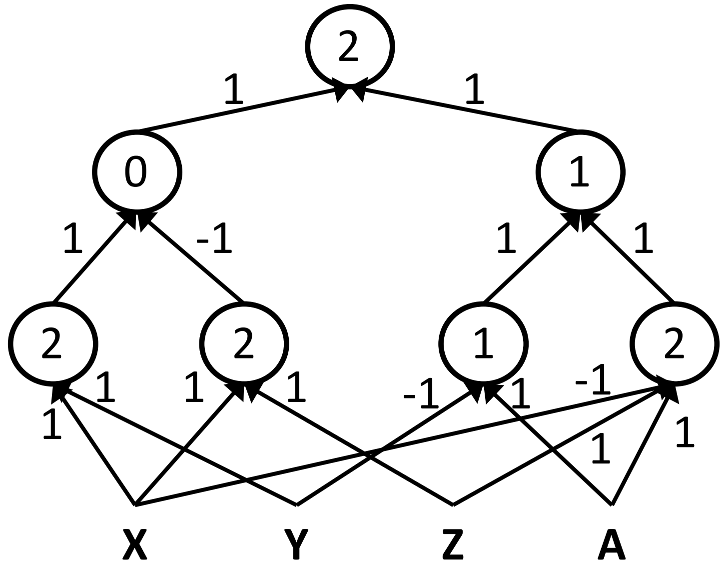
Multi-layer Perceptron!



- XOR
  - The first layer is a "hidden" layer
  - Also originally suggested by Minsky and Papert 1968

A more generic model

$$((A\&\bar{X}\&Z)|(A\&\bar{Y}))\&((X\ \&\ Y)|\overline{(X\&Z)})$$

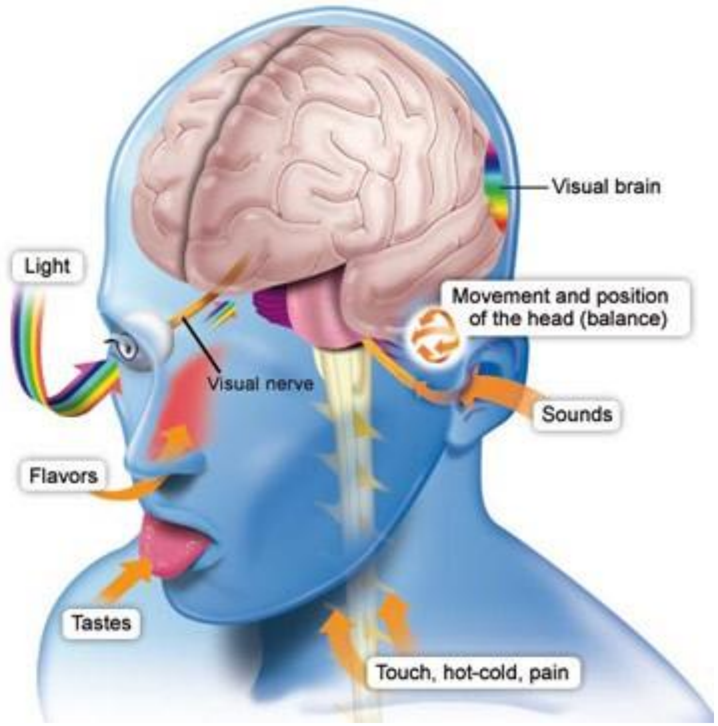


- A "multi-layer" perceptron
- Can compose arbitrarily complicated Boolean functions!
  - In cognitive terms: Can compute arbitrary Boolean functions over sensory input
  - More on this in the next class

Story so far

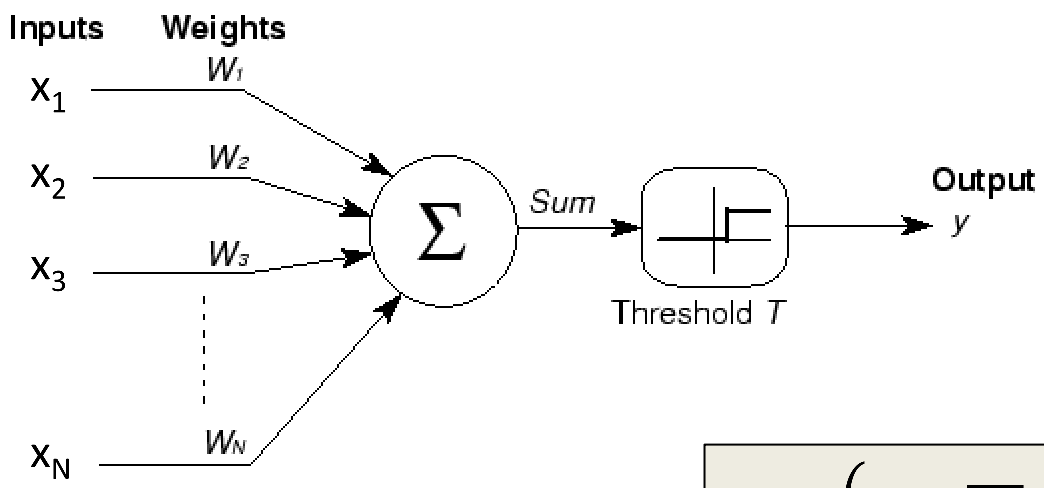
- Neural networks began as computational models of the brain
- Neural network models are *connectionist machines*
  - The comprise networks of neural units
- McCullough and Pitt model: Neurons as Boolean threshold units
  - Models the brain as performing propositional logic
  - But no learning rule
- Hebb's learning rule: Neurons that fire together wire together
  - Unstable
- Rosenblatt's perceptron : A variant of the McCulloch and Pitt neuron with a provably convergent learning rule
  - But individual perceptrons are limited in their capacity (Minsky and Papert)
- Multi-layer perceptrons can model arbitrarily complex Boolean functions

But our brain is not Boolean



- We have real inputs
- We make non-Boolean inferences/predictions

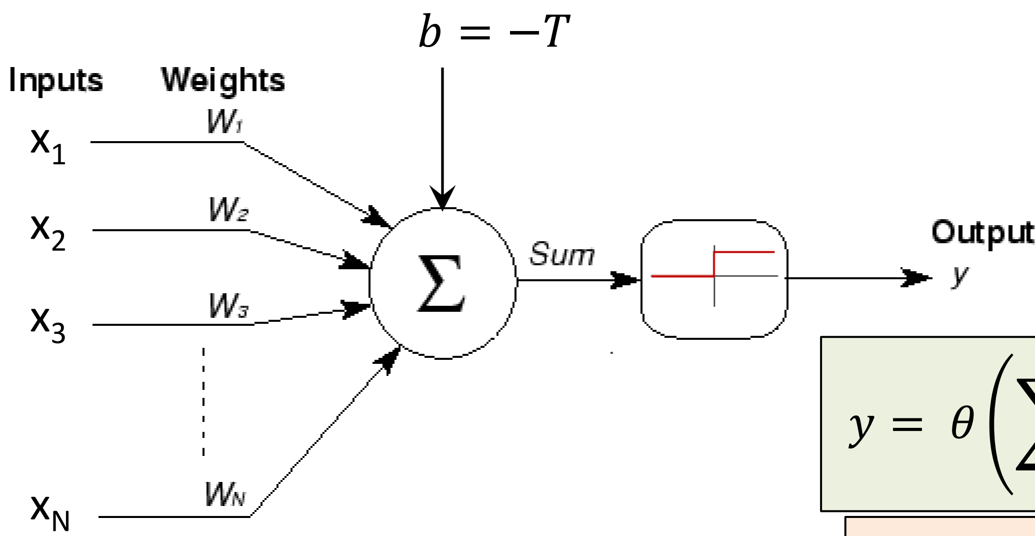
The perceptron with *real* inputs



$$y = \begin{cases} 1 \; if \; \sum_i w_i x_i - T \geq 0 \\ 0 \; else \end{cases}$$

- $x_1 \ldots x_N$ are real valued
- $w_1 \ldots w_N$ are real valued
- Unit "fires" if weighted input matches (or exceeds) a threshold

The perceptron with *real* inputs

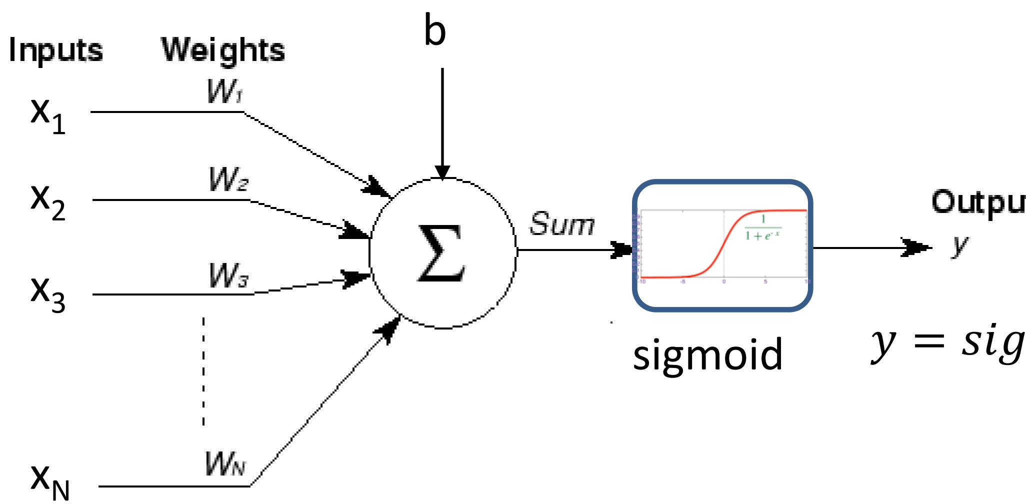


$$y = \theta\left(\sum_i w_i x_i + b\right)$$

$$\theta(z) = \begin{cases} 1 \; if \; z \geq 0 \\ 0 \; else \end{cases}$$

- Alternate view:
  - A threshold "activation" $\theta(z)$ operates on the weighted sum of inputs plus a bias
    - An *affine* function of the inputs
  - $\theta(z)$ outputs a 1 if z is non-negative, 0 otherwise
- Unit "fires" if weighted input matches or exceeds a threshold

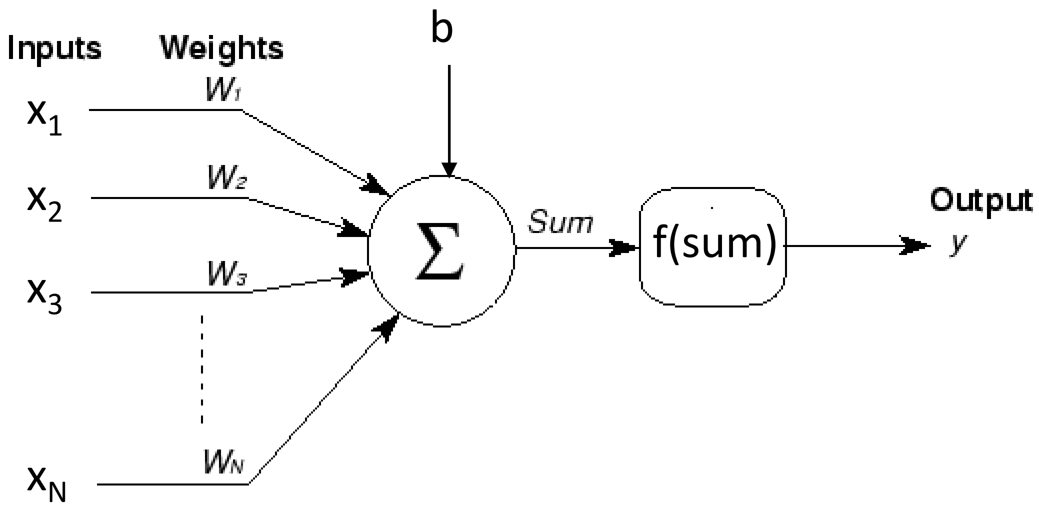
The perceptron with *real* inputs and a real *output*



$$y = sigmoid\left(\sum_i w_i x_i + b\right)$$

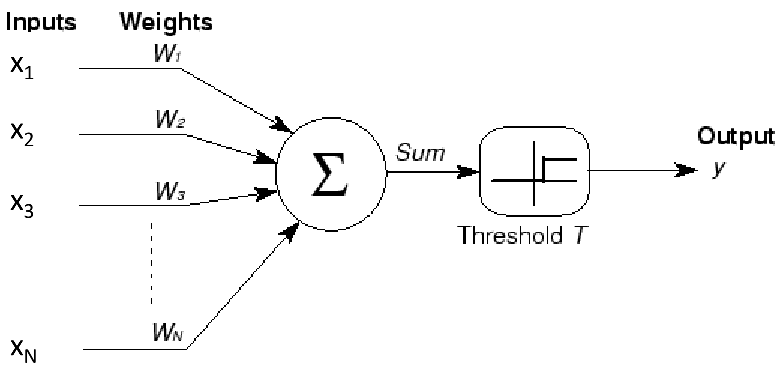
- $x_1 \ldots x_N$ are real valued
- $w_1 \ldots w_N$ are real valued
- The output y can also be real valued
  - Sometimes viewed as the "probability" of firing

The "real" valued perceptron

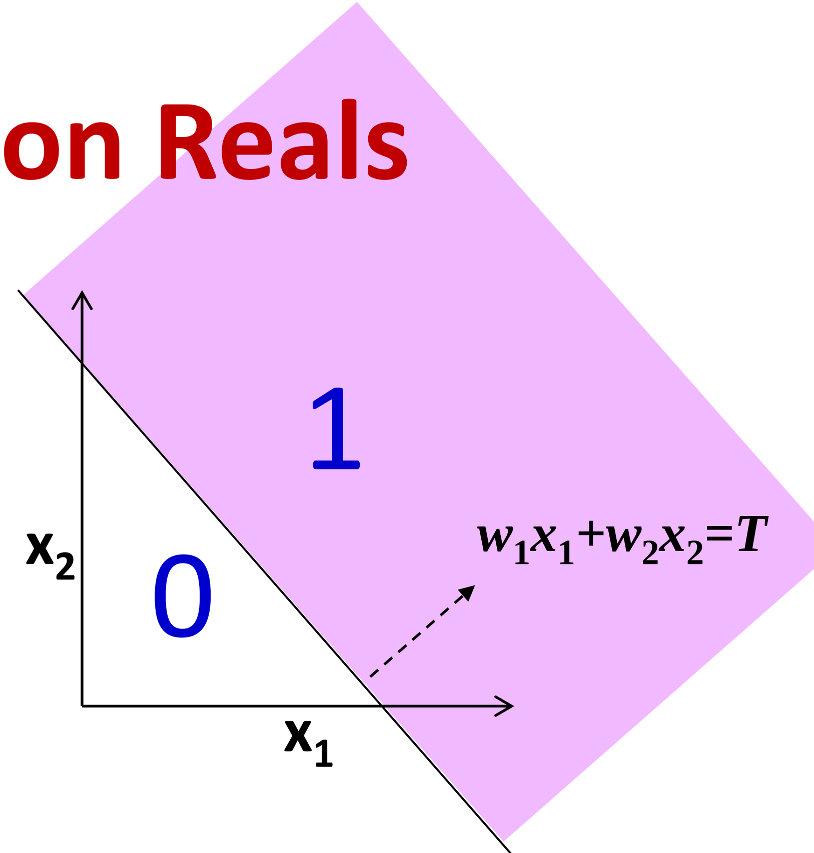


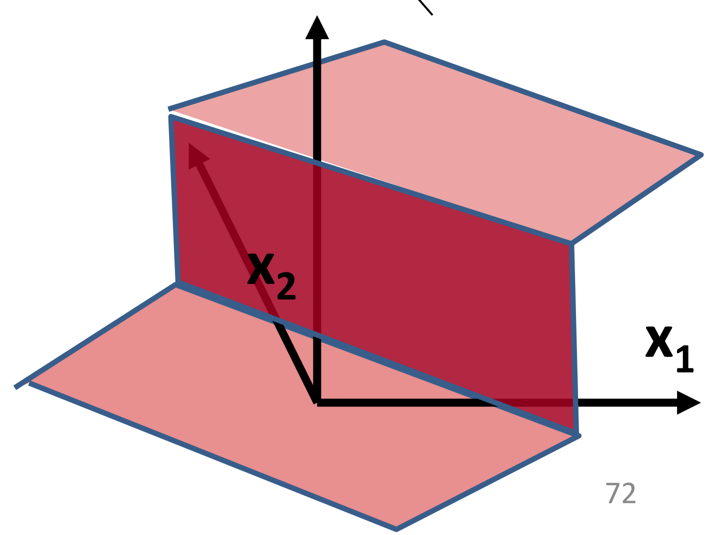
- Any real-valued "activation" function may operate on the weighted-sum input
  - We will see several later
  - Output will be real valued
- The perceptron maps real-valued inputs to real-valued outputs
- *Is useful to continue assuming Boolean outputs though, for interpretation*

A Perceptron on Reals



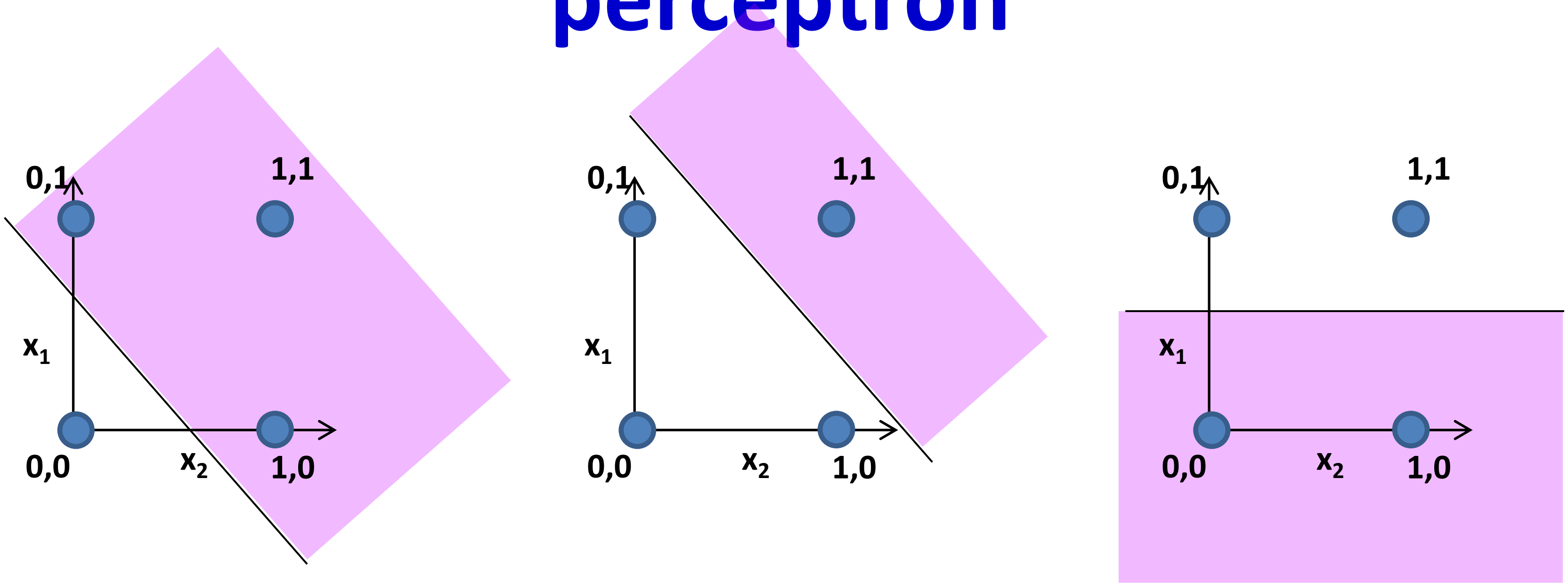
$$y = \begin{cases} 1 \; if \; \sum_i w_i x_i \geq T \\ 0 \; else \end{cases}$$





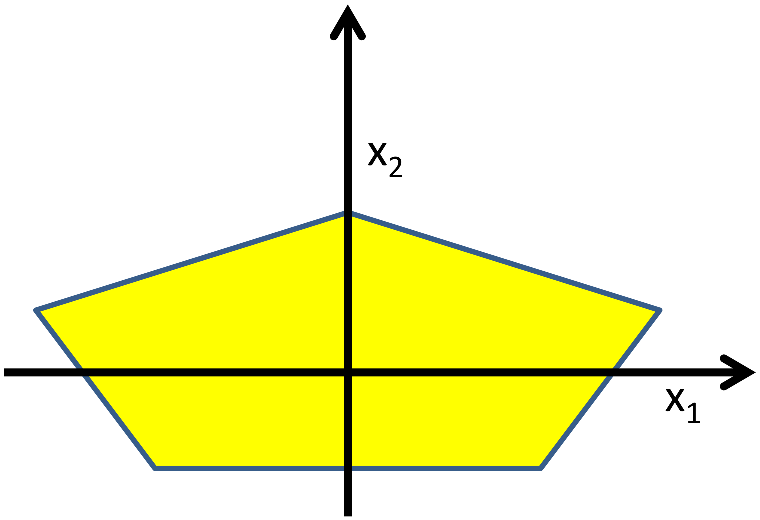
- A perceptron operates on *real*-valued vectors
  - **This is a *linear classifier***

Boolean functions with a real perceptron



- Boolean perceptrons are also linear classifiers
  - Purple regions have output 1 in the figures
  - What are these functions
  - Why can we not compose an XOR?

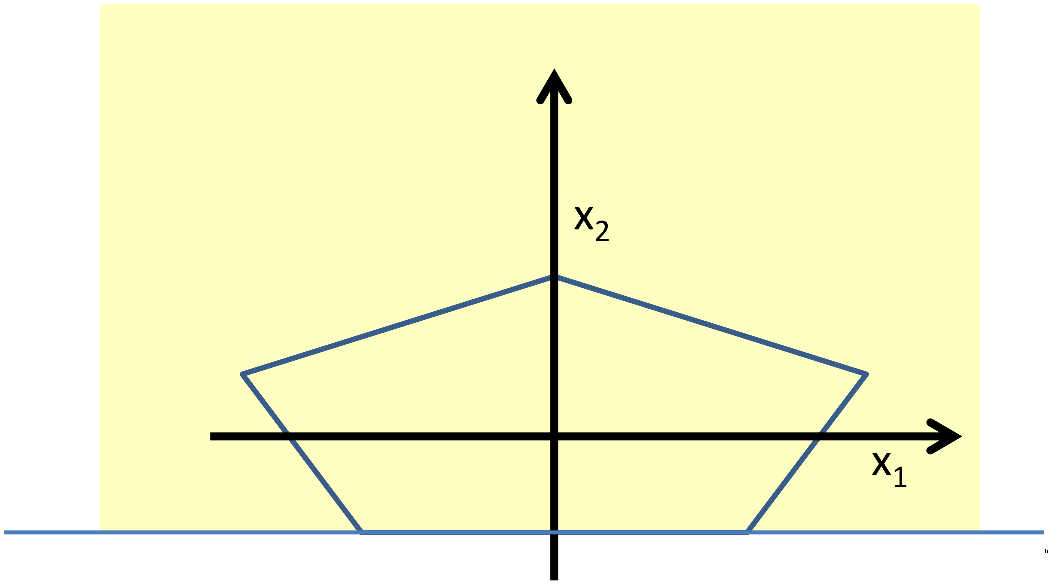
Composing complicated "decision" boundaries

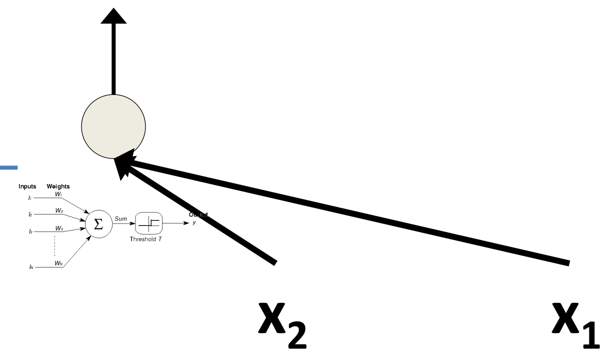


Can now be composed into "networks" to compute arbitrary classification "boundaries"

- Build a network of units with a single output that fires if the input is in the coloured area

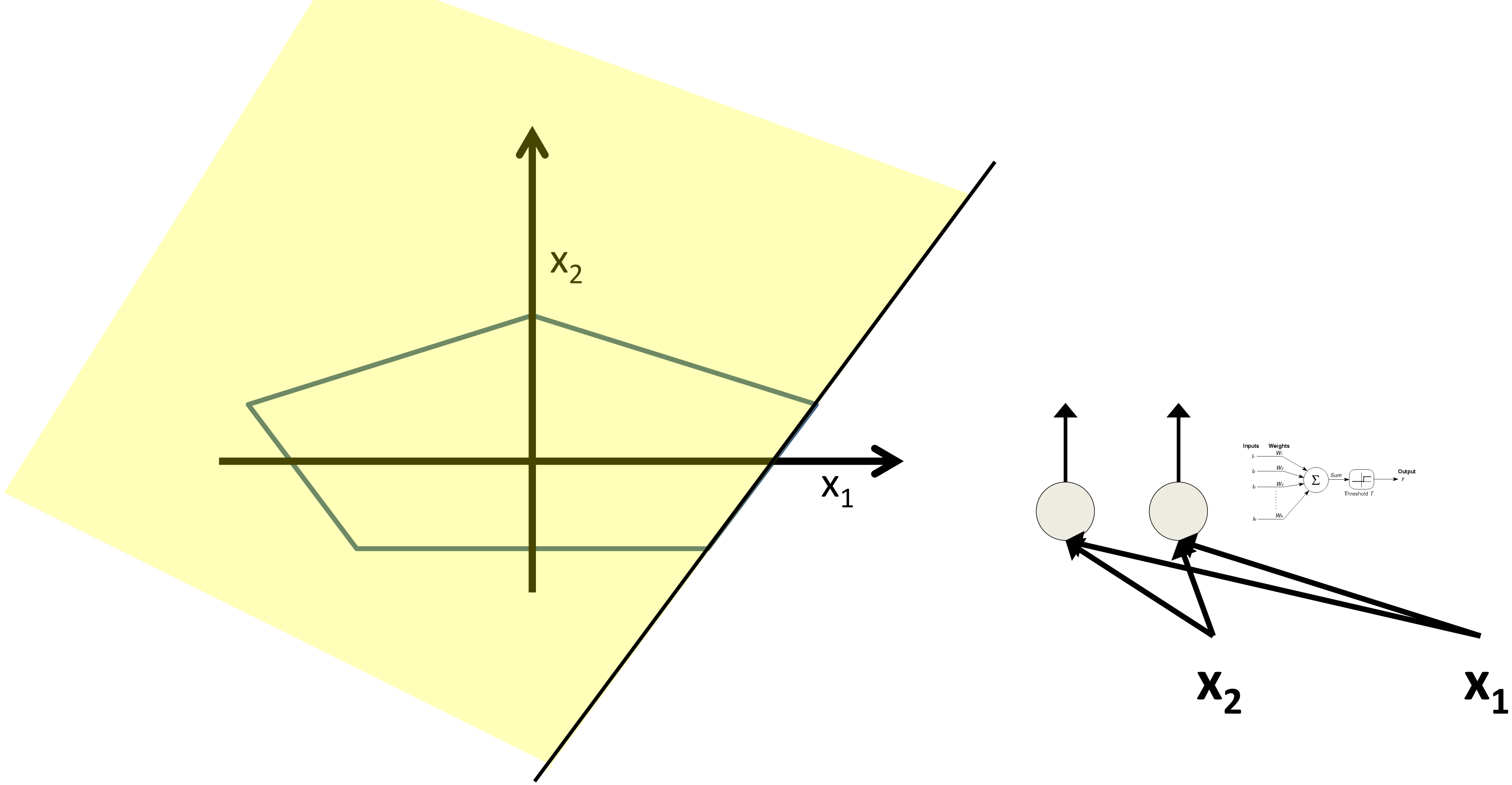
Booleans over the reals





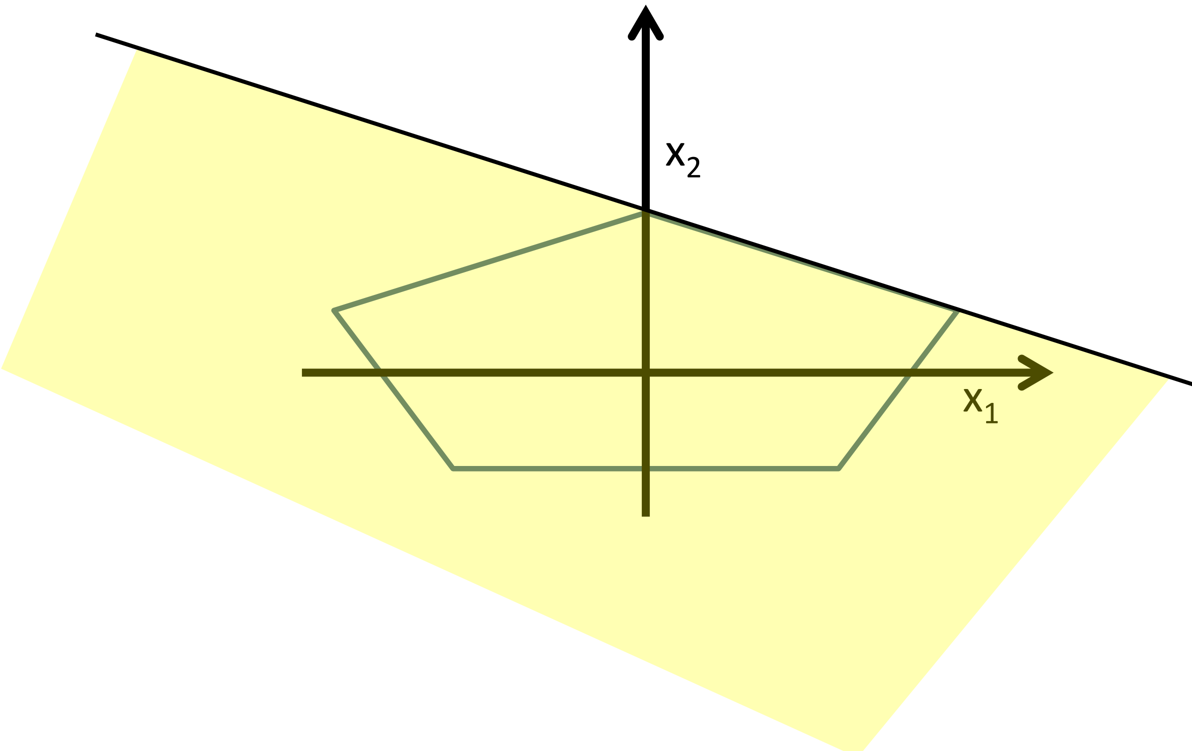
- The network must fire if the input is in the coloured area

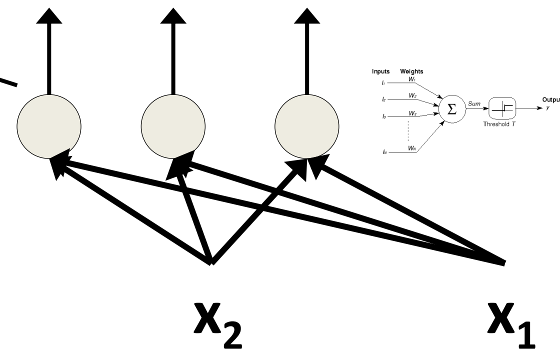
Booleans over the reals



- The network must fire if the input is in the coloured area

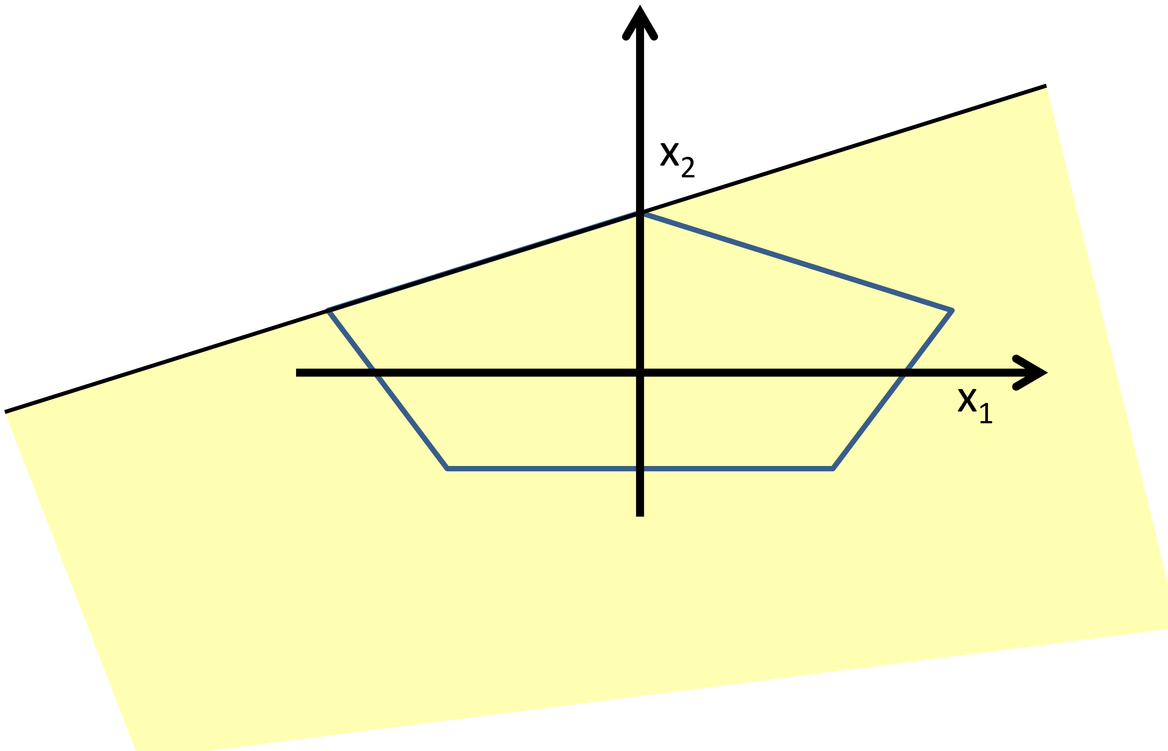
Booleans over the reals

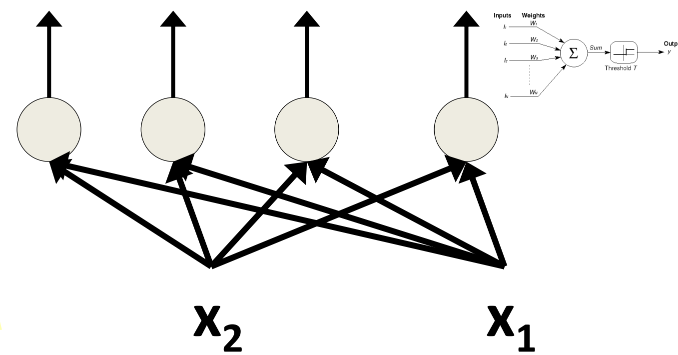




- The network must fire if the input is in the coloured area

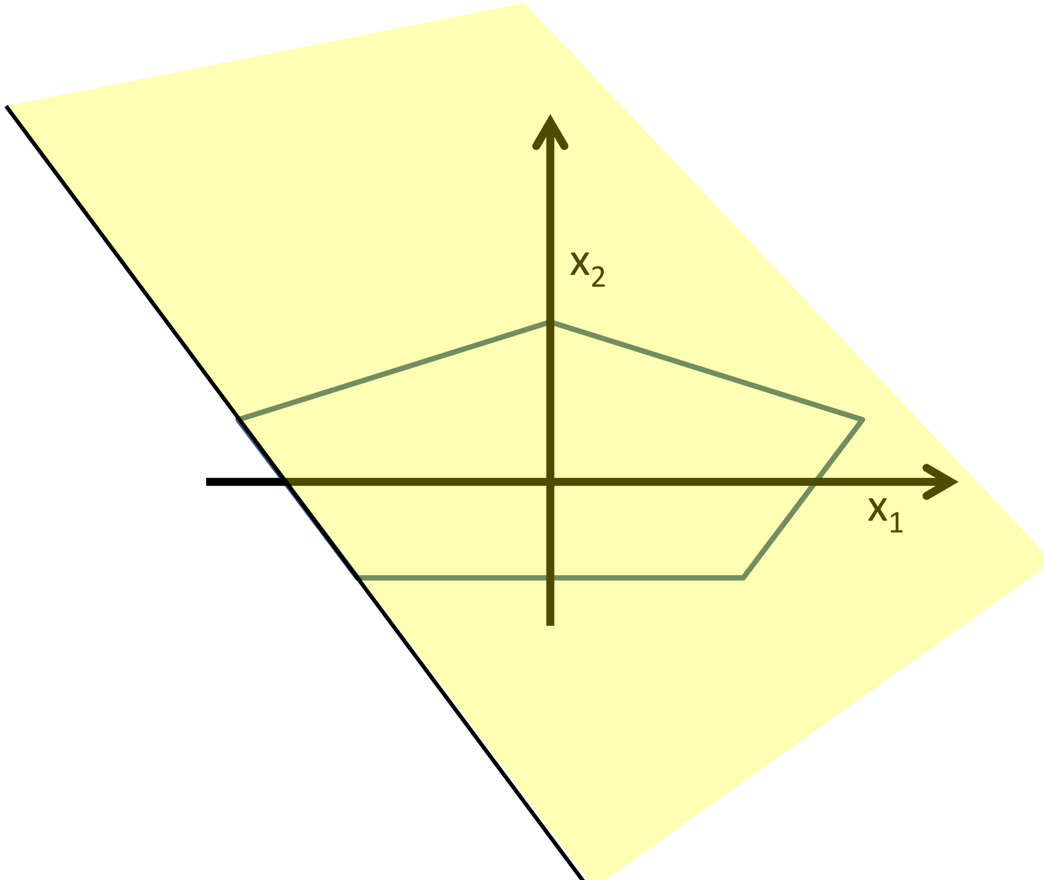
Booleans over the reals

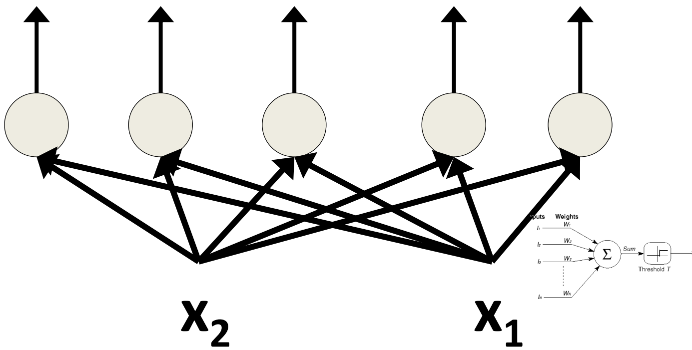




- The network must fire if the input is in the coloured area

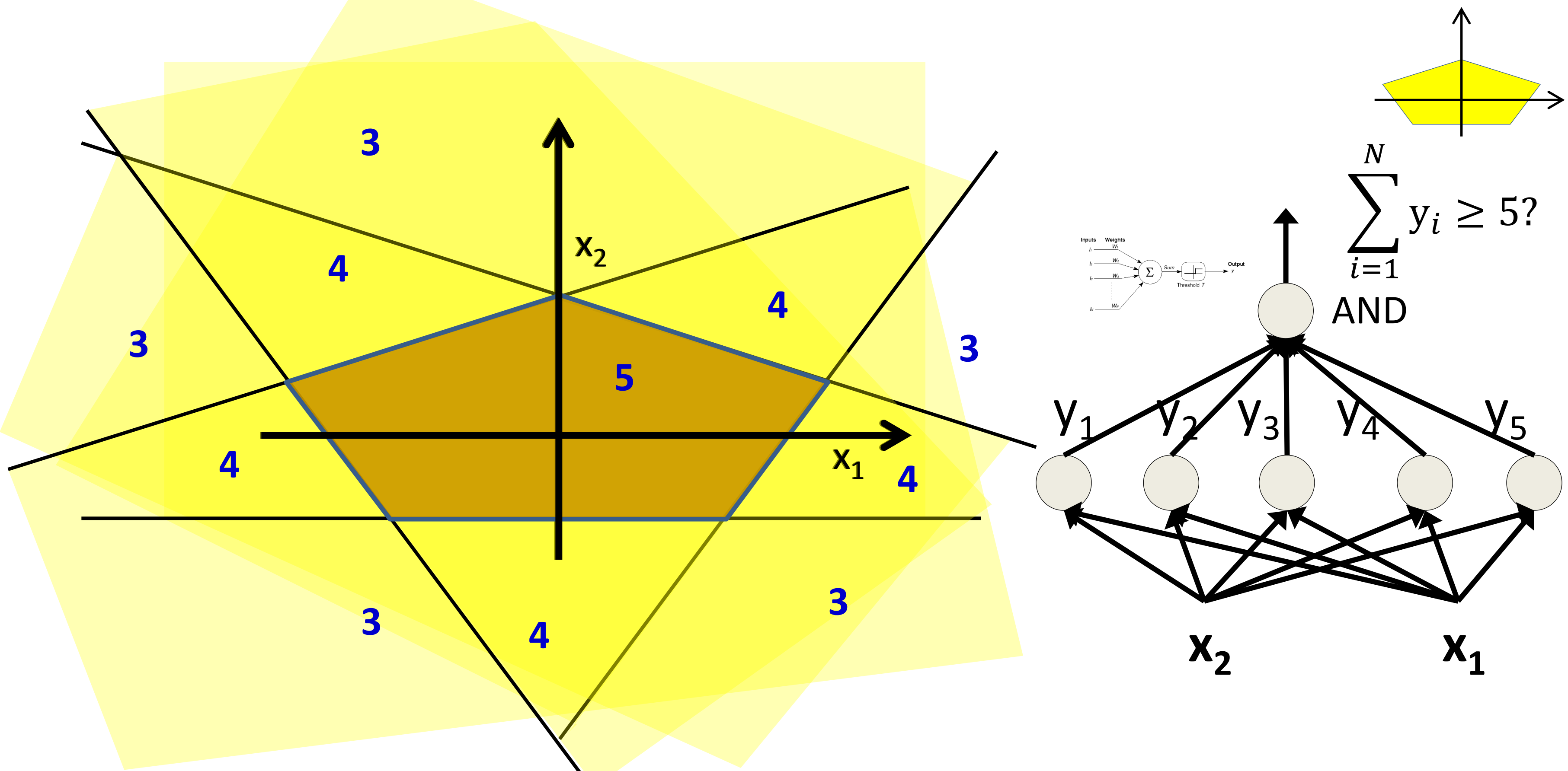
Booleans over the reals





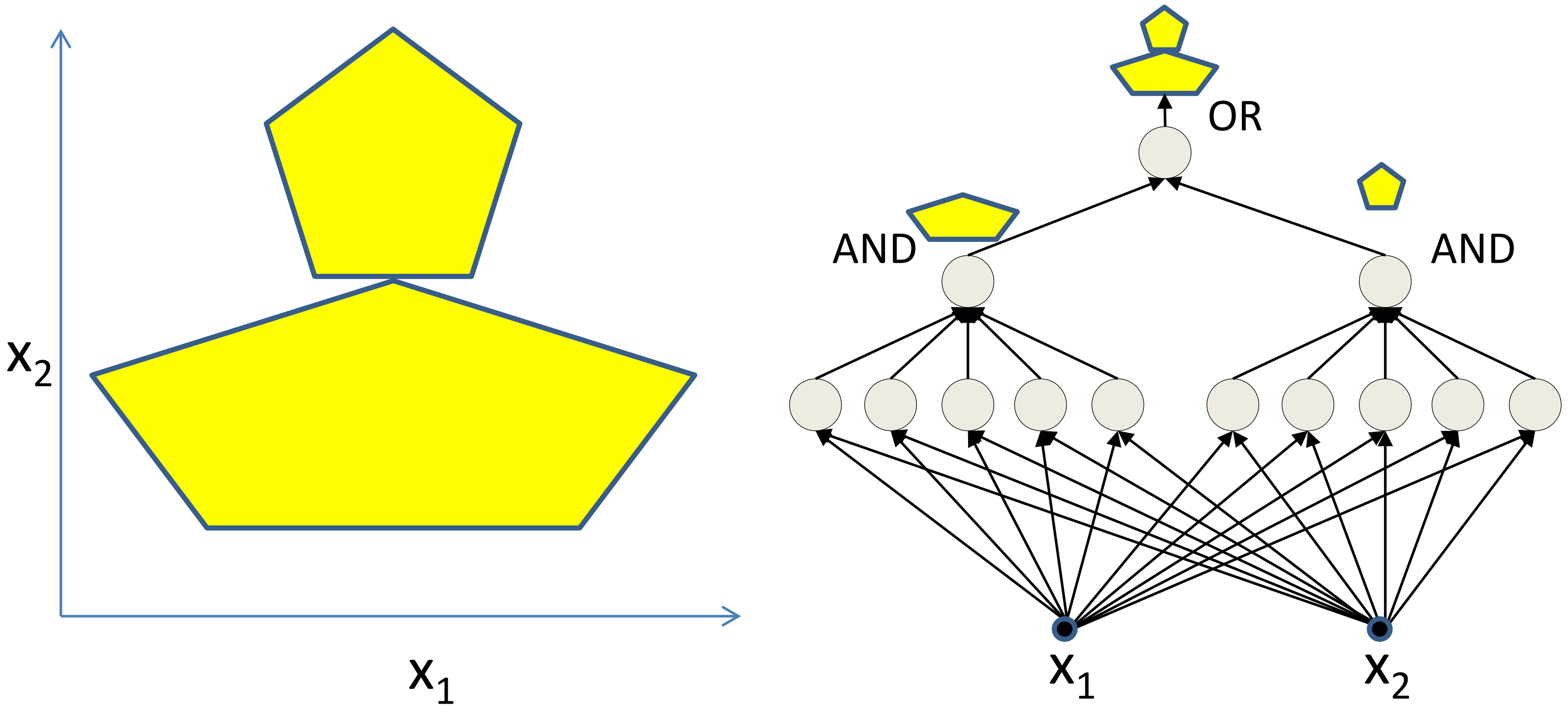
- The network must fire if the input is in the coloured area

Booleans over the reals



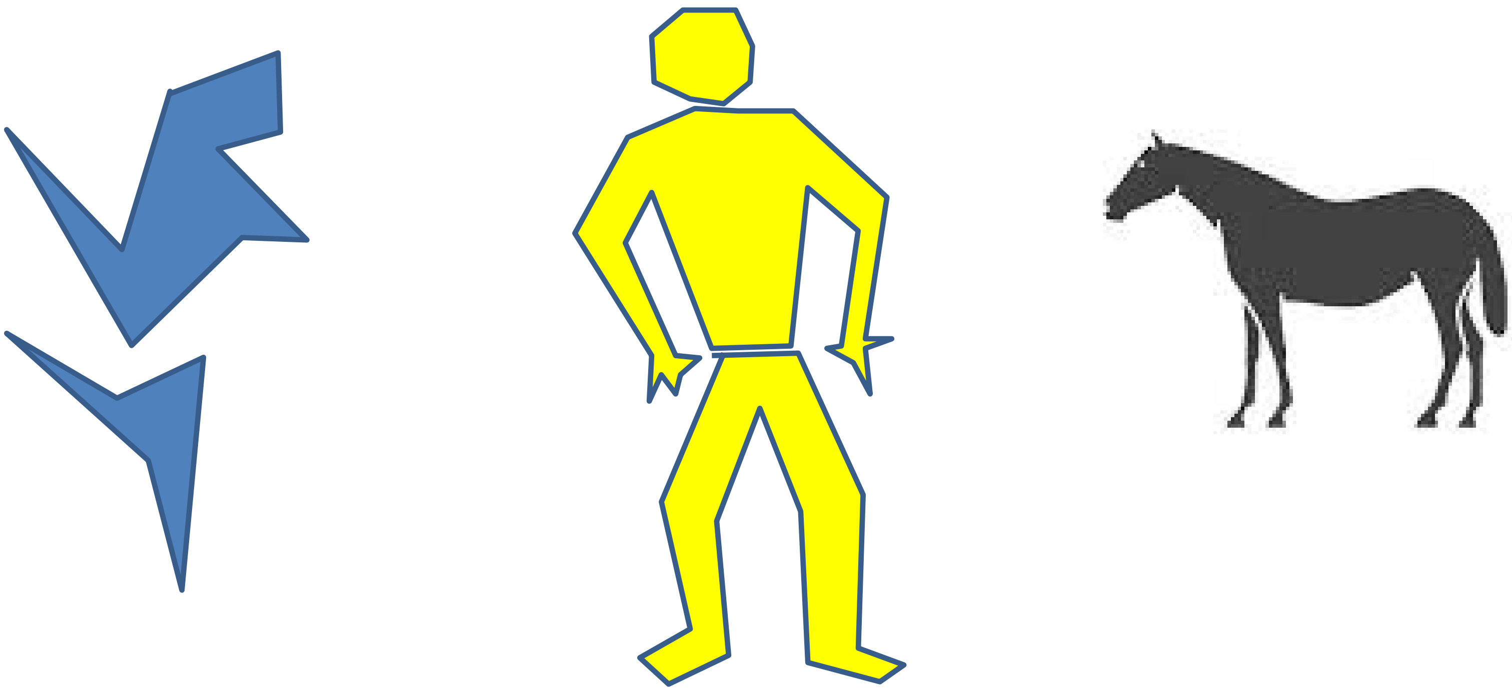
- The network must fire if the input is in the coloured area

More complex decision boundaries



- Network to fire if the input is in the yellow area
  - "OR" two polygons
  - A third layer is required

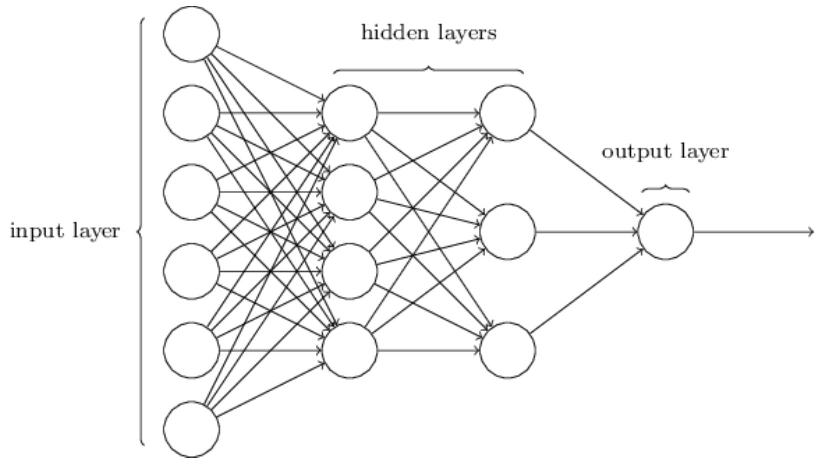
Complex decision boundaries

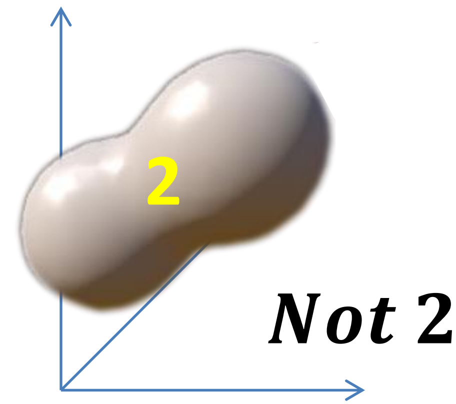


- Can compose very complex decision boundaries
  - How complex exactly? More on this in the next class

Complex decision boundaries

784 dimensions
(MNIST)





784 dimensions

- Classification problems: finding decision boundaries in high-dimensional space
  - Can be performed by an MLP
- MLPs can *classify* real-valued inputs

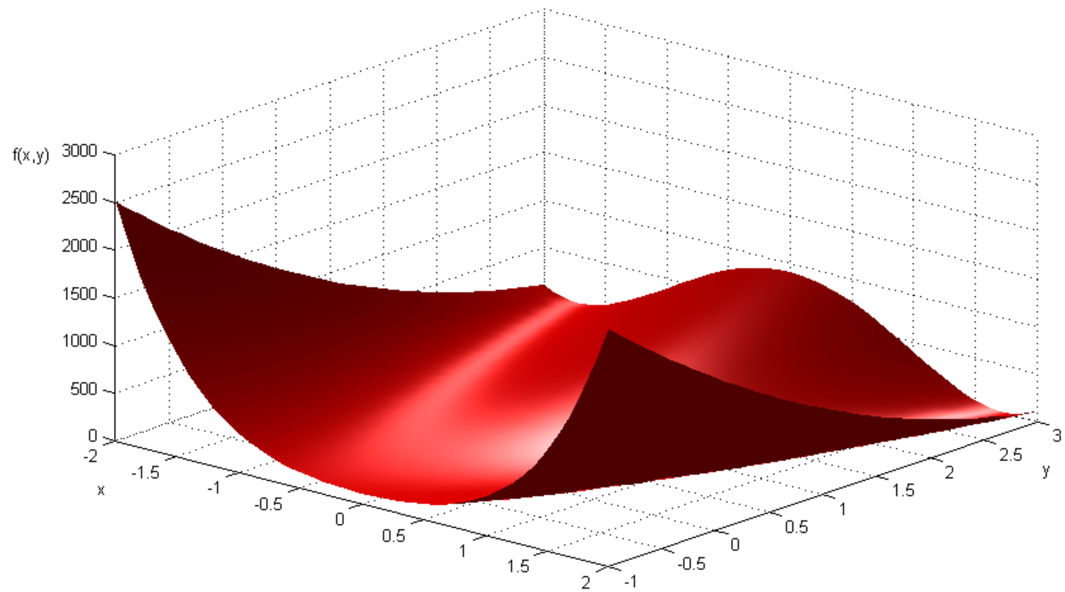
Story so far

- **MLPs are connectionist computational models**
  - Individual perceptrons are computational equivalent of neurons
  - The MLP is a layered composition of many perceptrons

- **MLPs can model Boolean functions**
  - Individual perceptrons can act as Boolean gates
  - Networks of perceptrons are Boolean functions

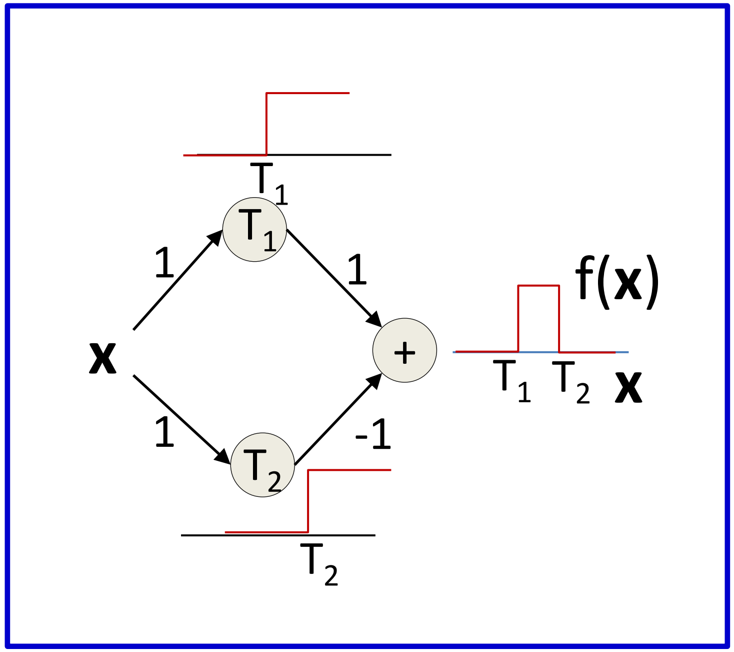
- **MLPs are Boolean *machines***
  - They represent Boolean functions over linear boundaries
  - They can represent arbitrary decision boundaries
  - They can be used to *classify* data

But what about continuous valued *outputs?*



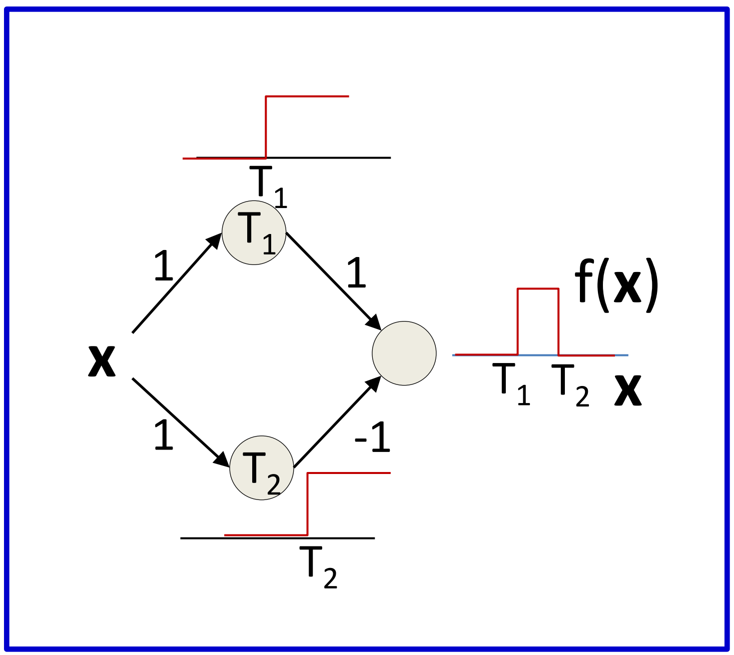
- Inputs may be real-valued
- Can outputs be continuous-valued too?

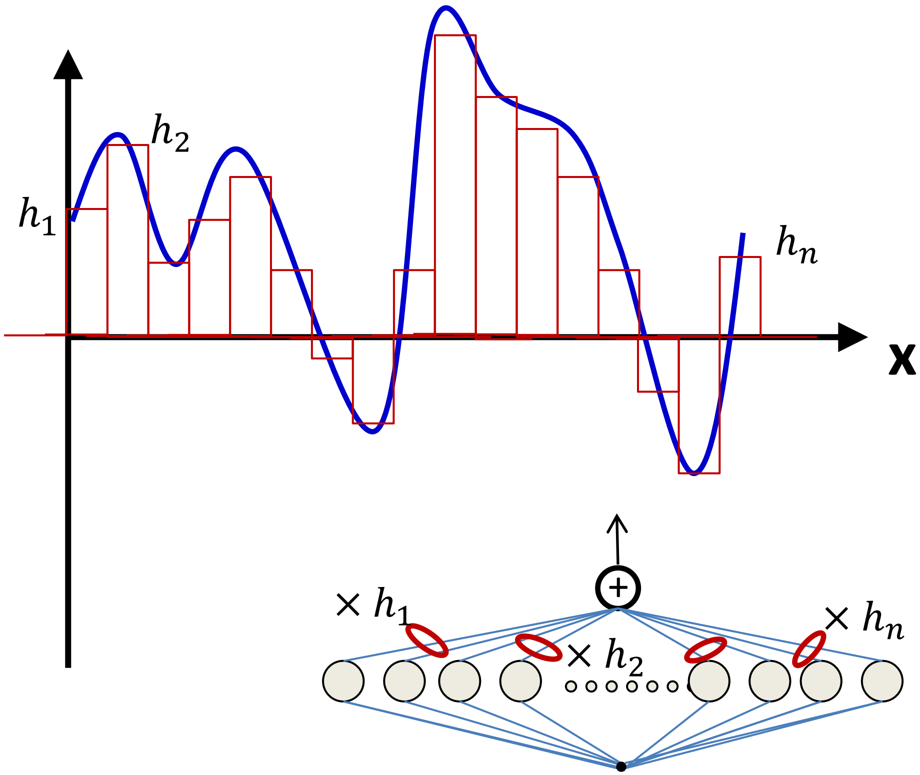
MLP as a continuous-valued regression



- A simple 3-unit MLP with a "summing" output unit can generate a "square pulse" over an input
  - Output is 1 only if the input lies between T_1 and T_2
  - T_1 and T_2 can be arbitrarily specified

MLP as a continuous-valued regression





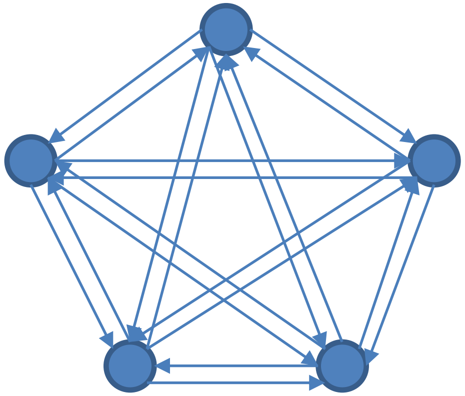
- A simple 3-unit MLP can generate a "square pulse" over an input
- **An MLP with many units can model an arbitrary function over an input**
  - To arbitrary precision
    - Simply make the individual pulses narrower
- This generalizes to functions of any number of inputs (next class)

Story so far

- **Multi-layer perceptrons are connectionist computational models**
- **MLPs are *classification engines***
  – They can identify classes in the data
  – Individual perceptrons are feature detectors
  – The network will fire if the combination of the detected basic features matches an "acceptable" pattern for a desired class of signal
- **MLP can also model continuous valued functions**

Other things MLPs can do

- Model memory
  - Loopy networks can “remember” patterns
    - Proposed by Lawrence Kubie in 1930, as a model for memory in the CNS



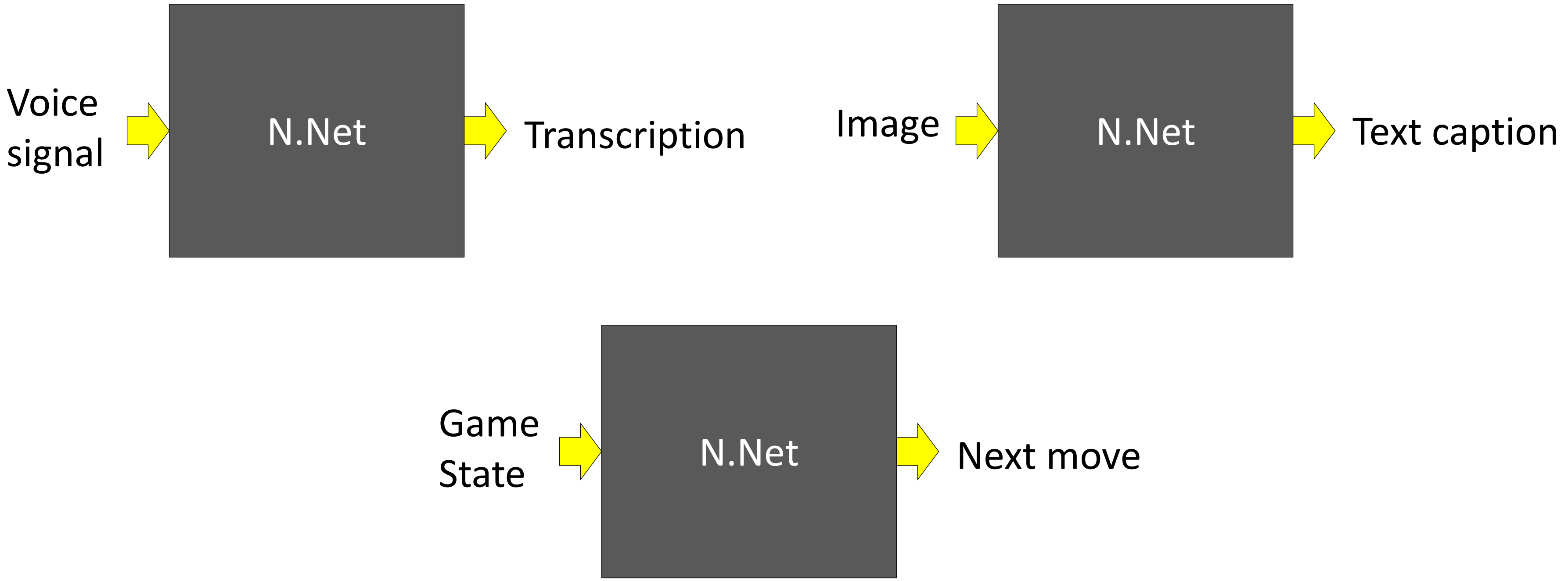
- Represent probability distributions
  - Over integer, real and complex-valued domains
  - MLPs can model both *a posteriori* and *a priori* distributions of data
    - A posteriori conditioned on other variables
  - MLPs can *generate* data from complicated, or even unknown distributions

- They can rub their stomachs and pat their heads at the same time..

NNets in AI

- The network is a function
  - Given an input, it computes the function layer wise to predict an output
    - More generally, given one or more inputs, predicts one or more outputs

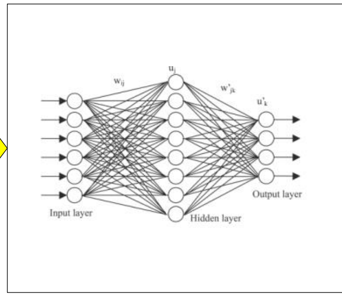
These tasks are *functions*



- Each of these boxes is actually a function
  - E.g f: Image → Caption

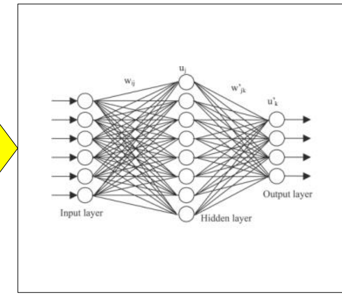
These tasks are *functions*

Voice signal



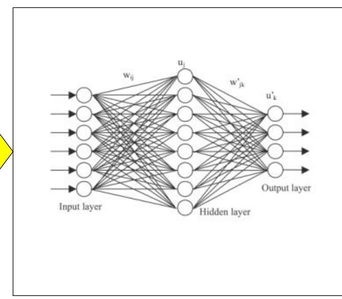
Transcription

Image



Text caption

Game State



Next move

- Each box is actually a function
  - E.g f: Image → Caption
  - It can be approximated by a neural network

Story so far

- Multi-layer perceptrons are connectionist computational models
- MLPs are *classification engines*
- MLP can also model continuous valued functions
- Interesting AI tasks are functions that can be modelled by the network

Next Up

- More on neural networks as universal approximators
  - And the issue of depth in networks