165B
Machine Learning
Introduction

Lei Li (leili@cs)
UCSB

Acknowledgement: Slides borrowed from Bhiksha Raj's 11485 and Mu Li & Alex Smola's 157 courses on Deep Learning, with modification

About the Course

- This course focus on a sub-field of machine learning -- Deep Learning, with moderate introduction to general learning concepts and methods.
- If you want to take a broader ML course, you may elect 165B for Spring quarter instead.

Course Philosophy

- No student left behind
- Please use the available resources
  - TA
  - Study groups
  - Office hours
  - Me (if additional office hour is needed, just email me)

Prerequisite

- You should have taken the following courses:
  - Calculus: Math 3A, 3B, 6A
    - Integration and derivative
    - Calculate gradients for multiple variables
  - Linear algebra: Math 4A, 4B
    - Vector, Matrix, norm, linear independence
  - Probability: Pstat 120A & 120B
    - Bayes Rule, likelihood, MLE
  - Algorithm & coding: CS 130A & 130B
    - Python, numpy, notebook

Logistics

- Course website:
  - https://www.cs.ucsb.edu/~lilei/course/mldl22w/
- Text
  - Deep Learning, Ian Goodfellow and Yoshua Bengio and Aaron Courville. (available online)
  - (Optional) Dive into Deep Learning, Aston Zhang, Zachary Lipton, Mu Li, Alexander Smola. (available online)

Lecture

- Required to attend
- M/W 11am-12:15pm
- First two weeks: on zoom
  - https://ucsb.zoom.us/j/82443832810?pwd=OWI0KzFMZVNKUndpemswdzFHMUFpdz09
- After Jan 18, wait for university's announcement. Likely to be in-person class.
- In-class quiz at random times
  - You must respond to all. We will mark your participation (but not the correctness)

Logistics

- Assignment
  - 4 Homework
  - HW2, 3, 4 will include both written assignments and machine problems (start early)
  - Submission on Gradescope:
    - https://www.gradescope.com/courses/344145
    - **You should already be added, let me know if not.**
  - Deadline: before class begins on the due date (11am)

Discussion Forum

- Ed platform
  - Join at
  - https://edstem.org/us/join/cPQgXM
  - Post questions or discussion on topics related to course material, assignments
  - Message can be private if only send to instructor & TA
  - We will use the same platform for in-class quiz

Grading

- Homework:
  - 15% each
  - A total of 3 late days for the whole course (applied greedy)
  - Solution submitted after late days will be graded 0.
- Final exam:
  - 30%
- Class participation and in-class quiz
  - 10%
  - We will count the in-class quiz submission
  - If you have legitimate reason to be absent, please email me and TA

Academic integrity is absolutely required

- Allowed:
  - Discussion of lecture and textbook materials
  - Discussion of how to approach assignments, what techniques to consider, what textbook or lecture material is relevant
- Not allowed:
  - Sharing ideas in the form of code, pseudocode, or solutions
  - Turning in someone else's work as your own, even with that person's permission.
  - Allowing someone else to turn in your work as his or her own.
  - Turning in work without proper acknowledgment of the sources of the content (including ideas) contained within the work.
- We will use software to detect plagiarism.
  - It will detect even if change of variables

Recitation Sessions

- Wednesdays:
  - 4-5pm GIRV 2119
  - 5-6pm PHELP1448
  - 6-7pm PHELP2532
  - 7-8pm PHELP2532
- Encourage to attend
- TA will cover
  - Background materials
  - Additional examples
  - Coding assistance
  - Q/A

Computing Resources

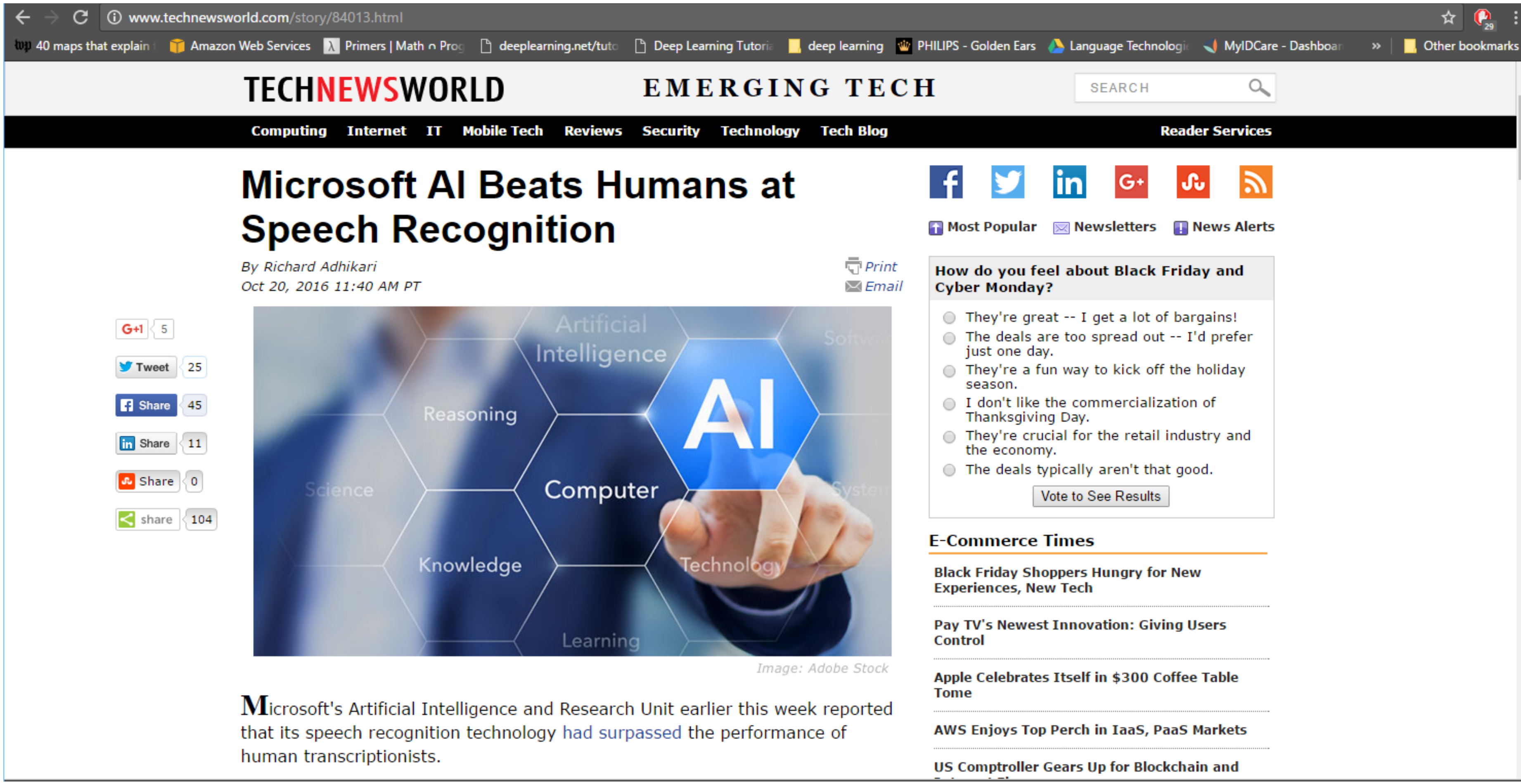
- UCSB supercomputing center
  - http://csc.cnsi.ucsb.edu/acct
- CS Computing Lab
- Google Colab
  - https://colab.research.google.com/
  - Free

Deep Learning

- Deep Learning 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 Deep Learning



Microsoft AI Beats Humans at Speech Recognition

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

Print
Email

G+1 5
Tweet 25
Share 45
Share 11
Share 0
share 104

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 Deep Learning

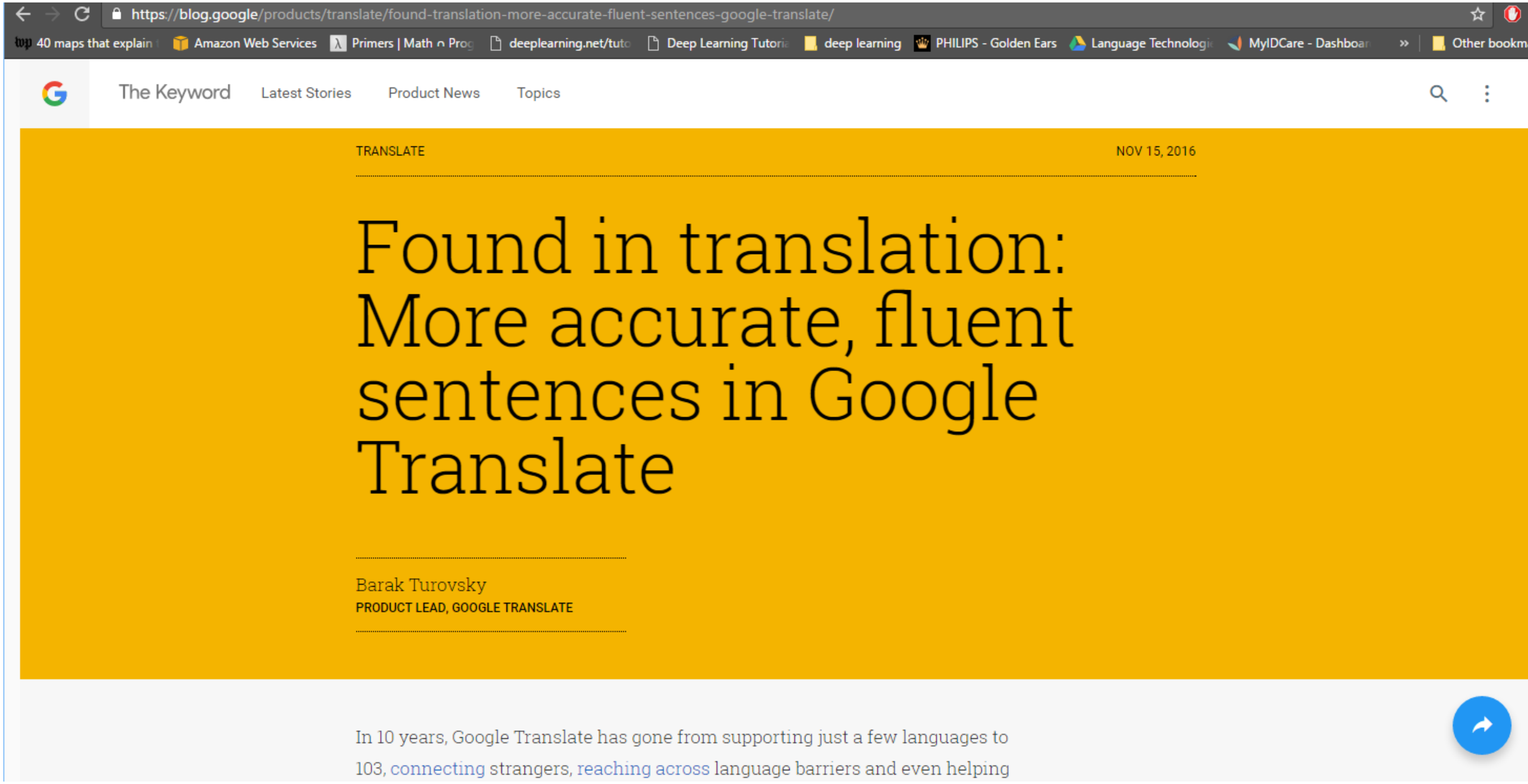
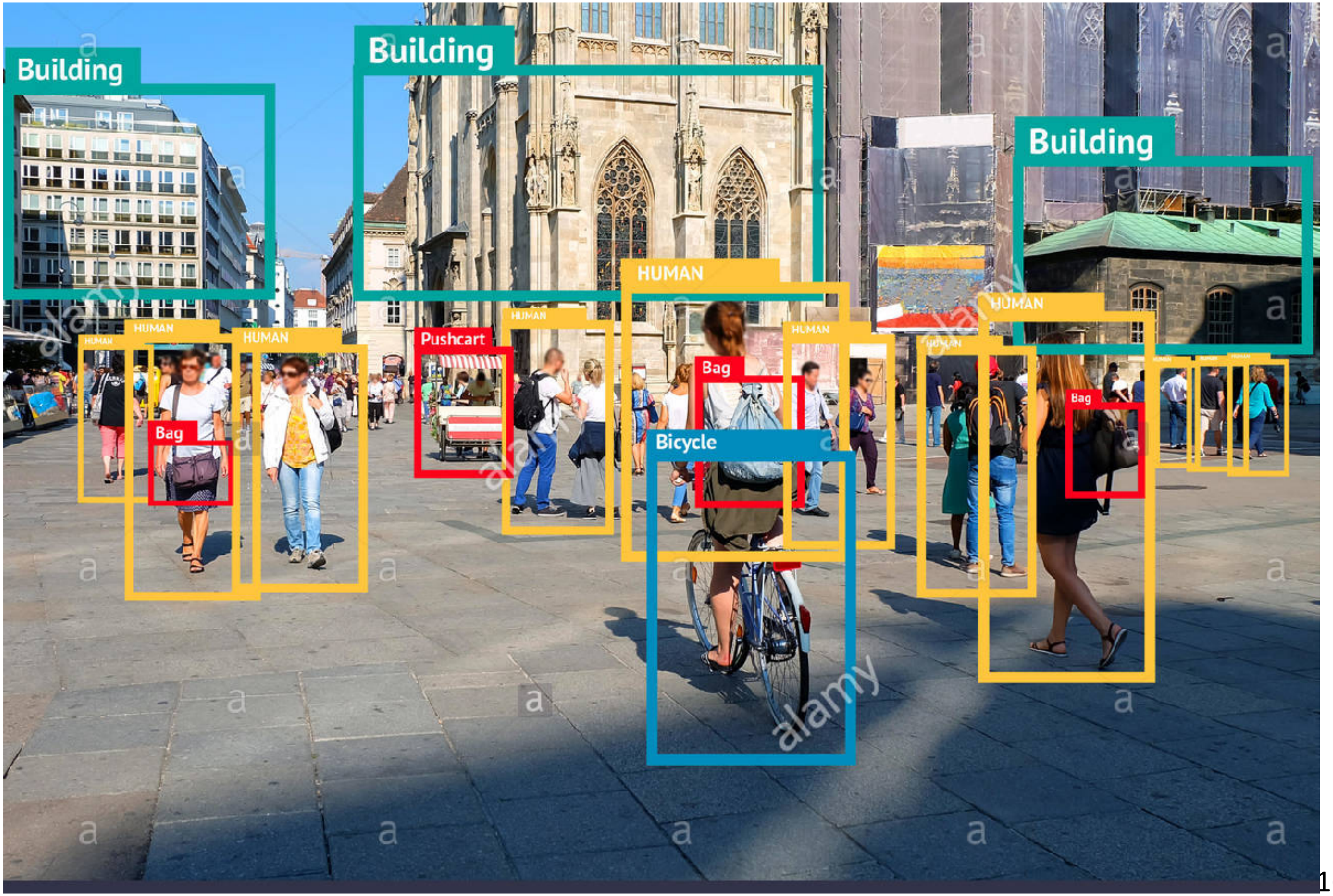
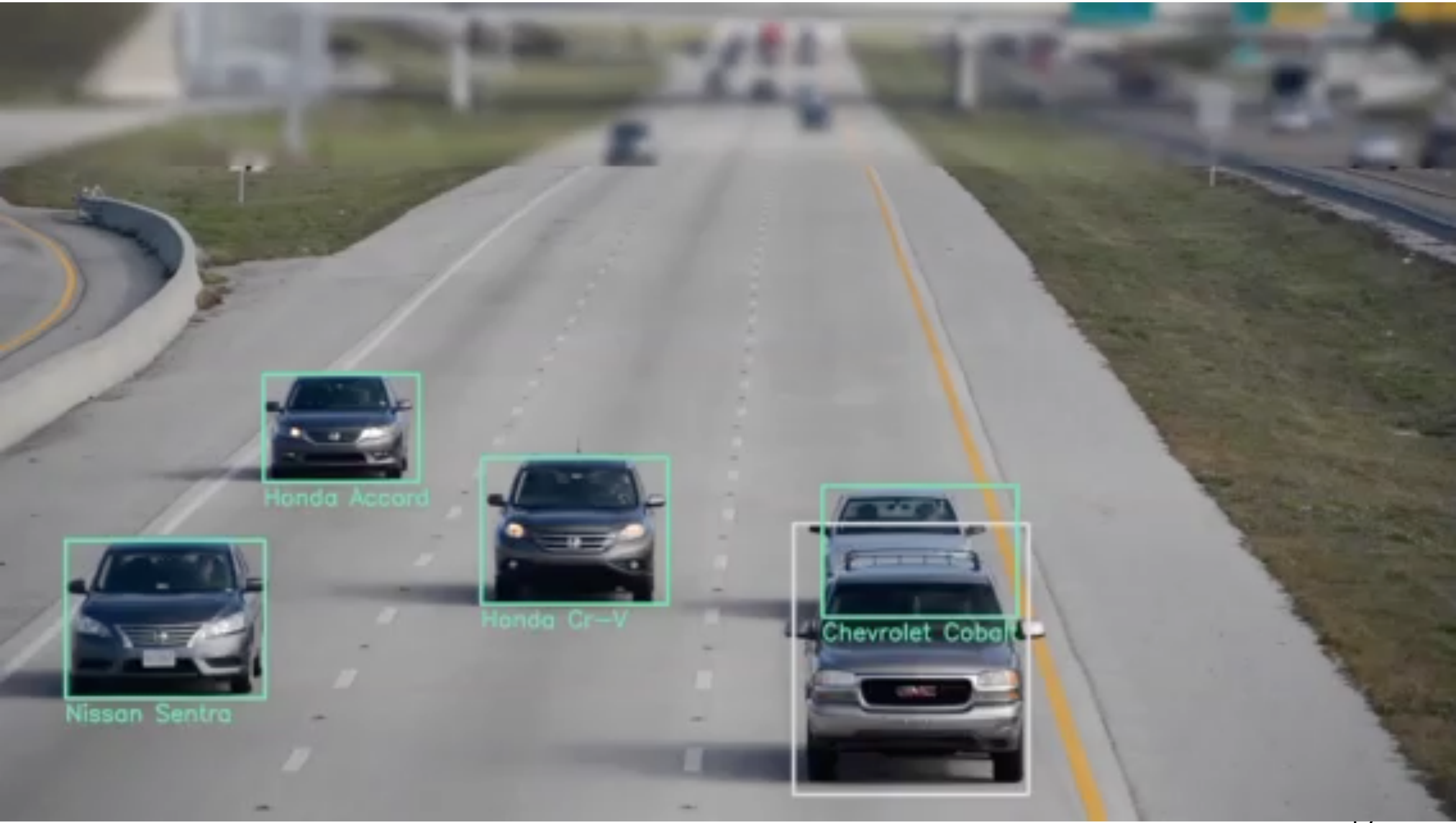


Image segmentation and Object recognition

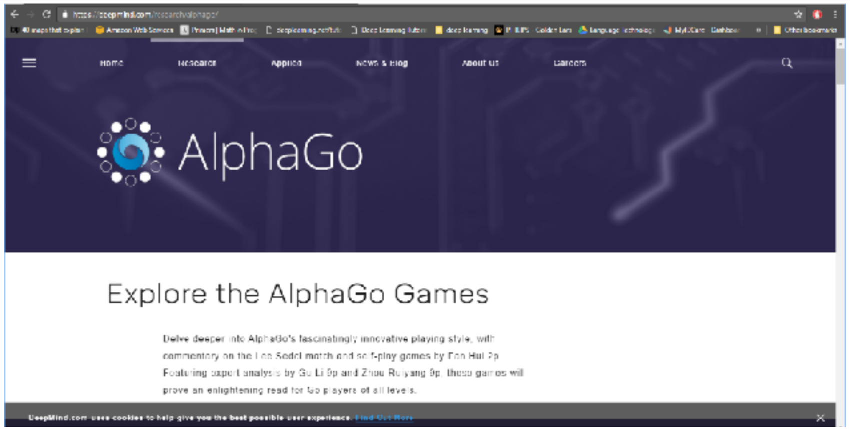


Autonomous Driving



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

Achieving Master Level in GO



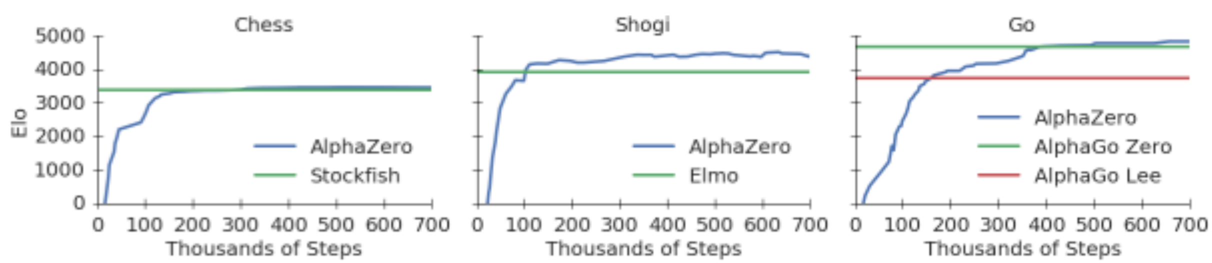
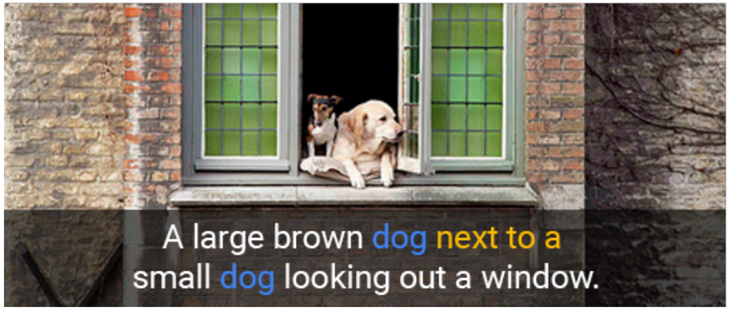


Figure 1: Training *AlphaZero* for 700,000 steps. Elo ratings were computed from evaluation games between different players when given one second per move. **a** Performance of *AlphaZero* in chess, compared to 2016 TCEC world-champion program *Stockfish*. **b** Performance of *AlphaZero* in shogi, compared 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*).

Image Captioning

Human captions from the training set



Automatically captioned

Successes with Deep Learning

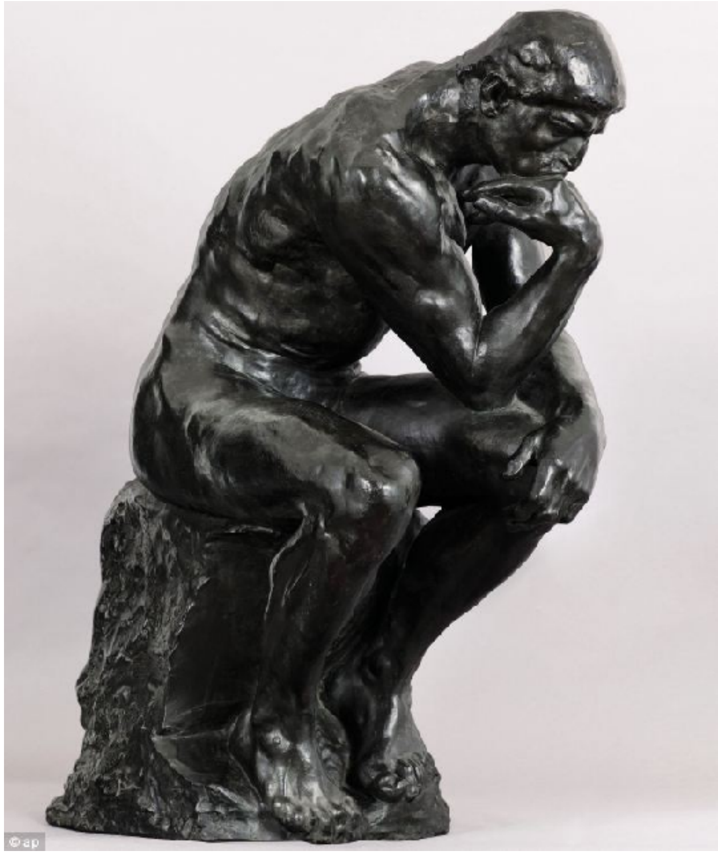
- And a variety of other problems:
  - From art to astronomy to healthcare..
  - and even predicting stock markets!

So what are neural networks??

- It begins with this..

So what are neural networks??

- Or even earlier.. with this..



"The Thinker!"
by Augustin Rodin

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

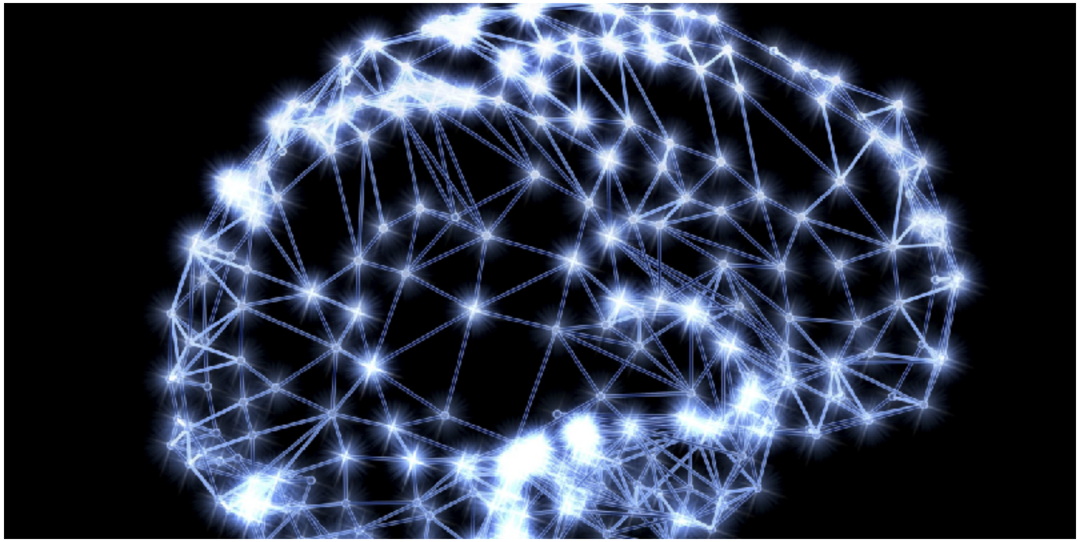
Early Models of Human Cognition

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

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

Observation: *The Brain*

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



Brain: Interconnected Neurons

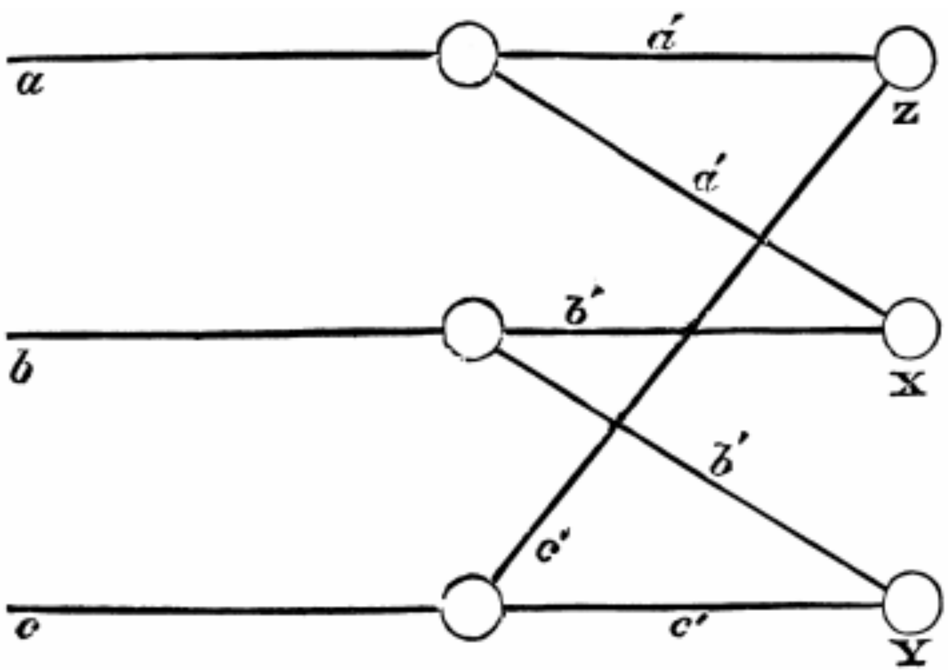
- Many neurons connect *in* to each neuron
- Each neuron connects *out* to many neurons

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

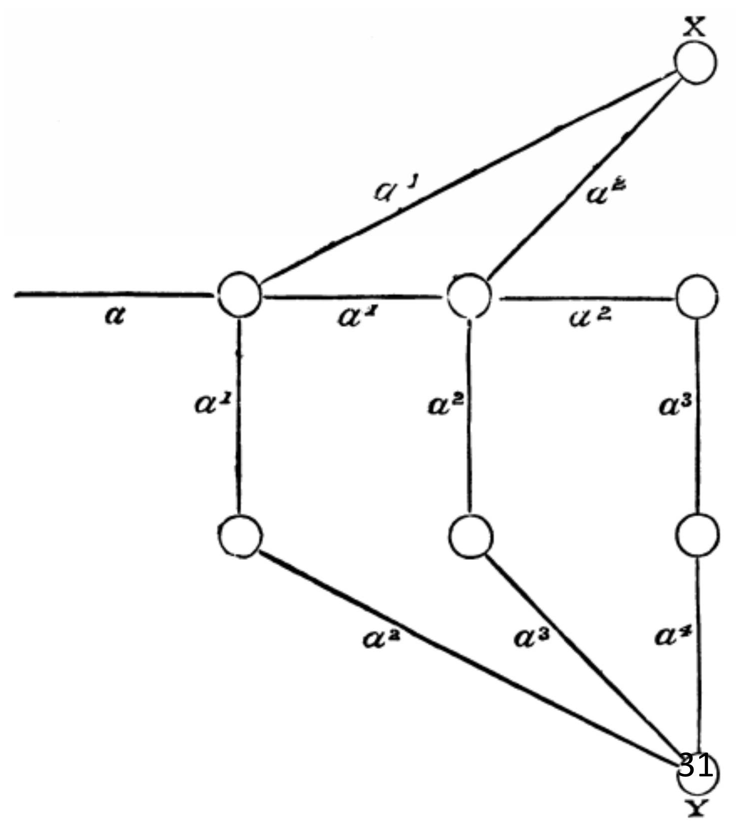
- Neurons excite and stimulate each other
- Different combinations of inputs can result in different outputs



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..

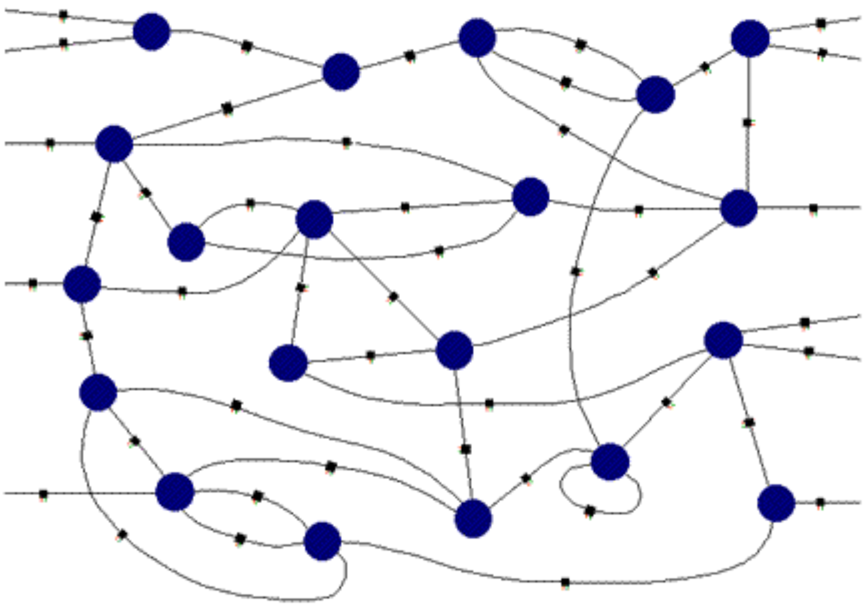
- The human brain is a connectionist machine
  - Bain, A. (1873). *Mind and body. The theories of their relation.* London: Henry King.
  - Ferrier, D. (1876). *The Functions of the Brain.* London: Smith, Elder and Co

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

- Connectionist machines emulate this structure

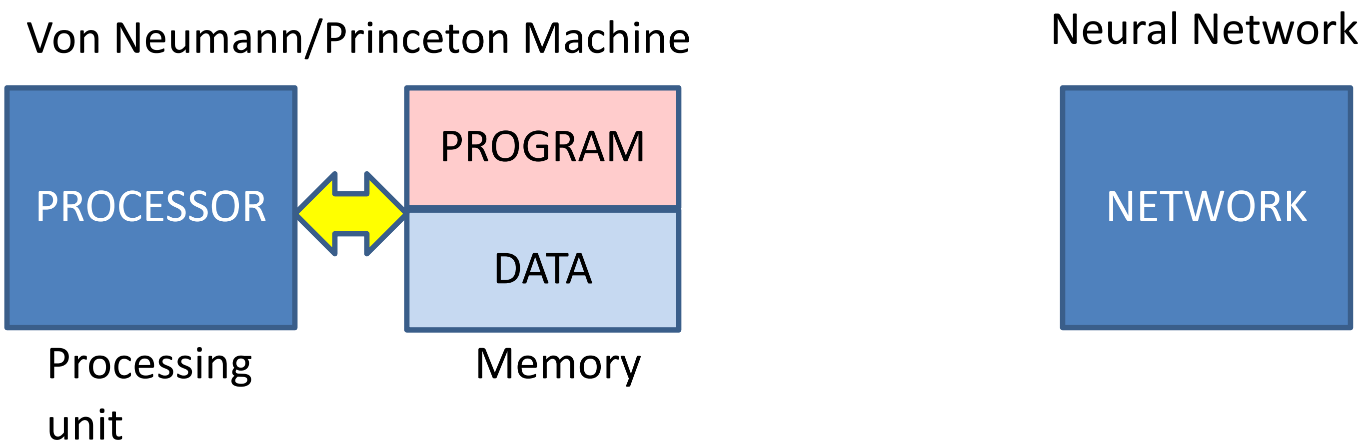
Connectionist Machines

- 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



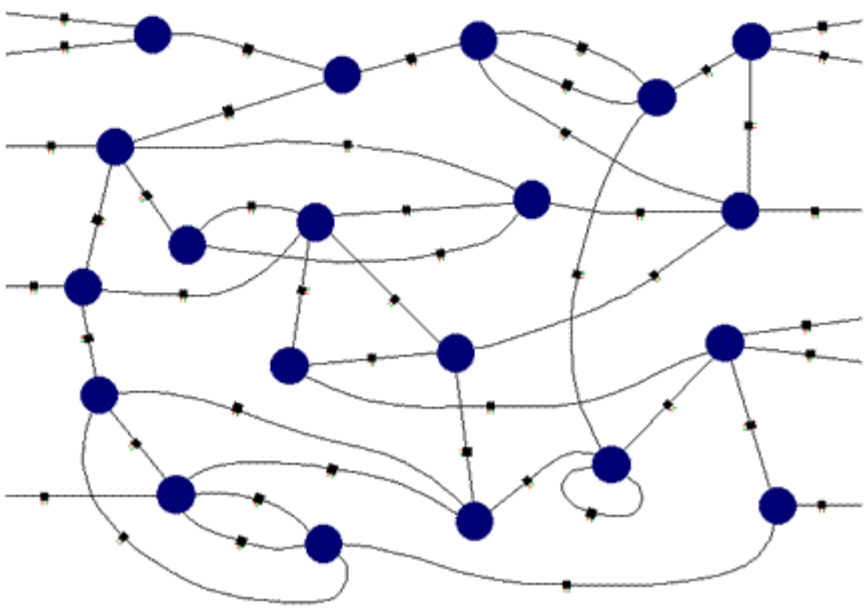
- The machine has many non-linear processing units
  - The program is the connections between these units
    - Connections may also define memory

Recap

- Deep Learning 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*

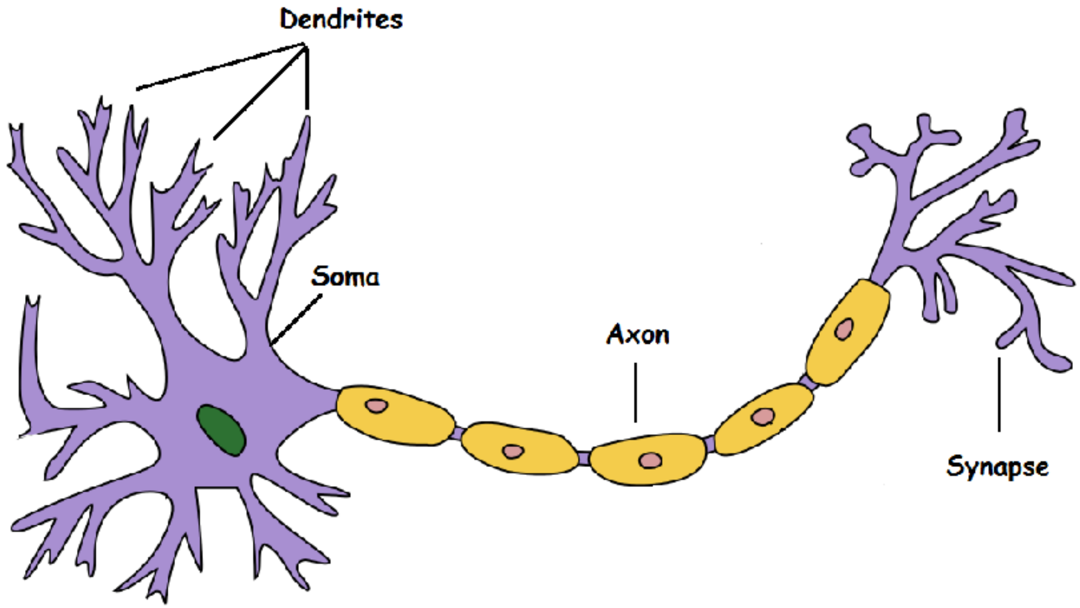
Connectionist Machines

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



Modelling the brain

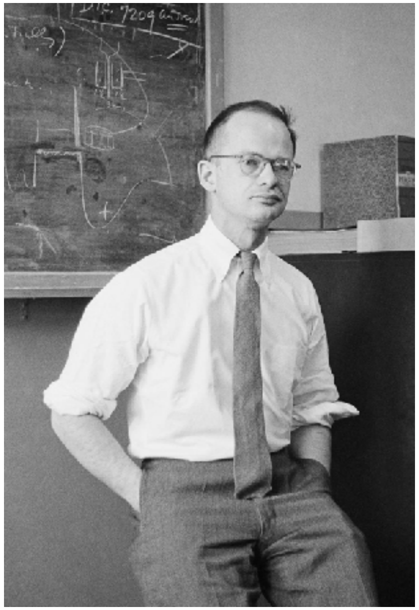
- A neuron:



- 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

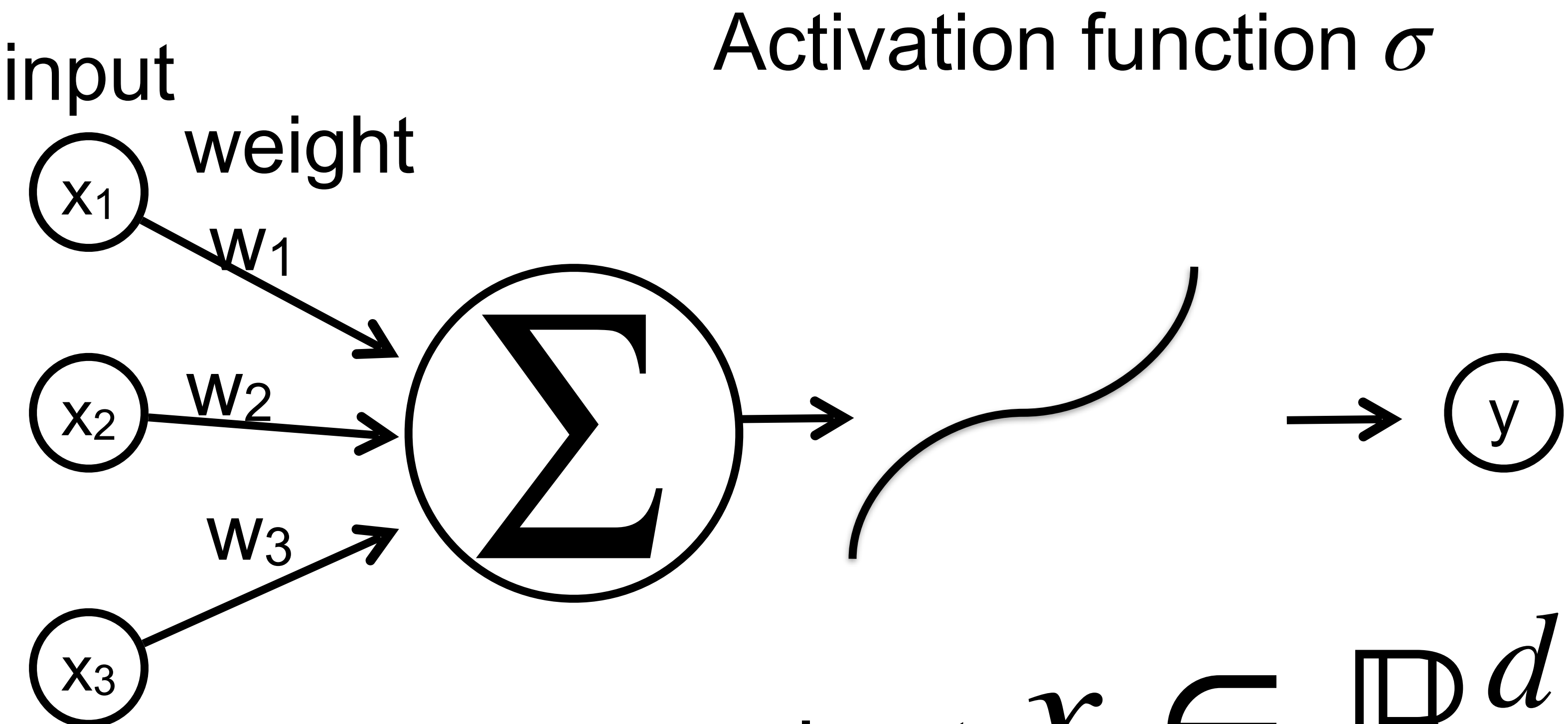
- The Doctor and the Hobo..
  - Warren McCulloch: Neurophysiologist
  - Walter Pitts: Homeless wannabe logician who arrived at his door



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

A single Artificial Neuron



Input: $x \in \mathbb{R}^d$

Weight:

$$w \in \mathbb{R}^d, b \in \mathbb{R}$$

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

Novelist, farmer, hobo, schoolteacher psychologist

- "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:

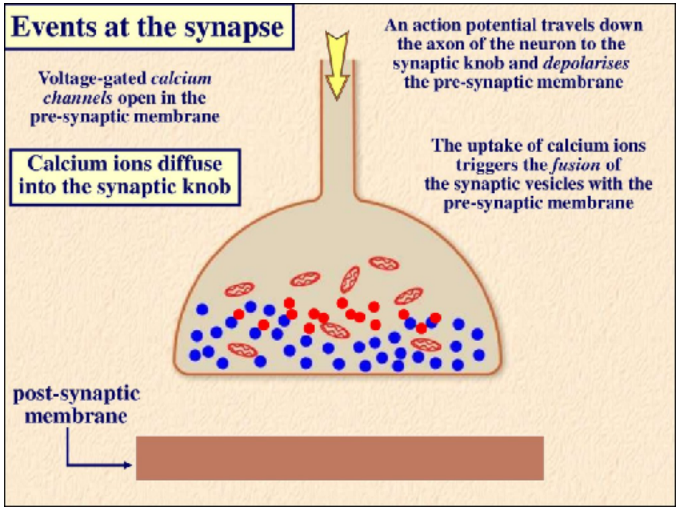
Axonal connection from neuron X

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

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

Dendrite of 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

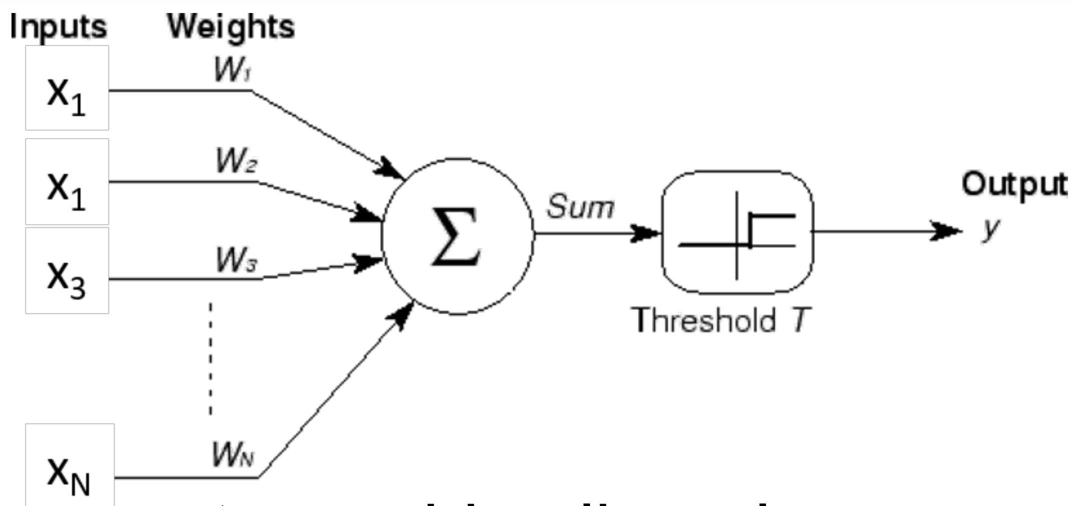
$$w_{ij} = w_{ij} + \eta y_j \left(x_i - \sum_{k=1}^{j} w_{ik} y_k \right)$$

  - 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

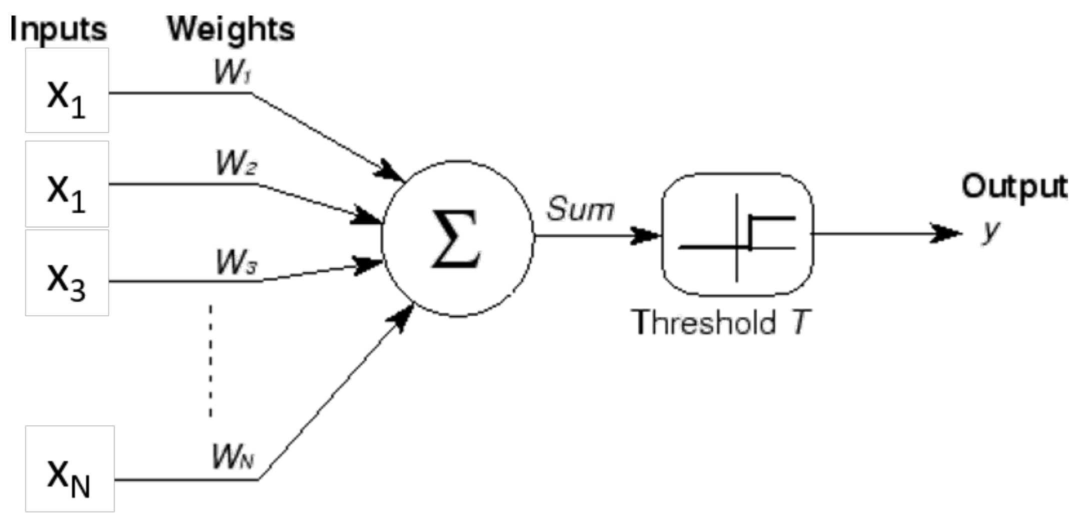


- Number of inputs combine linearly
  - Threshold logic: Fire if combined input exceeds threshold

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

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

- 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

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

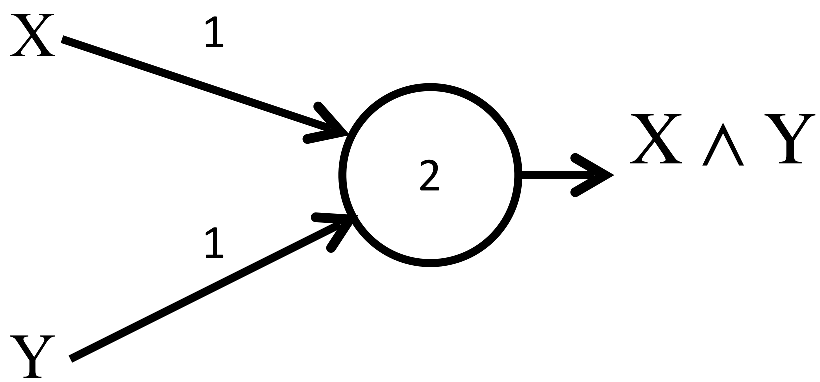
Sequential Learning:

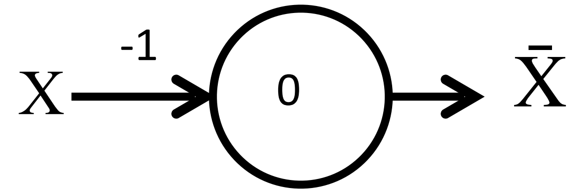
$d(x)$ is the desired output in response to input $\mathbf{x}$

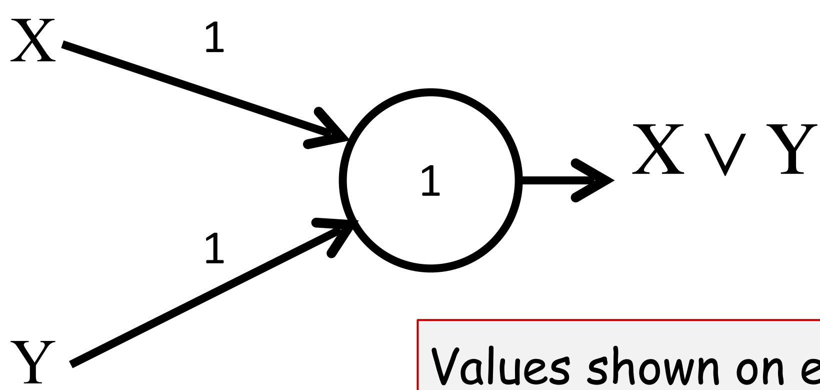
$y(x)$ is the actual output in response to $\mathbf{x}$

Perceptron

- Easily shown to mimic any Boolean gate







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

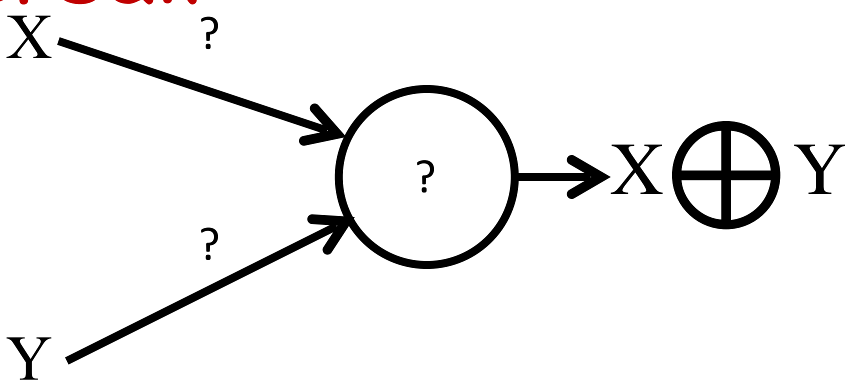
- But…

Perceptron

- Minsky and Papert, 1968

No solution for XOR!

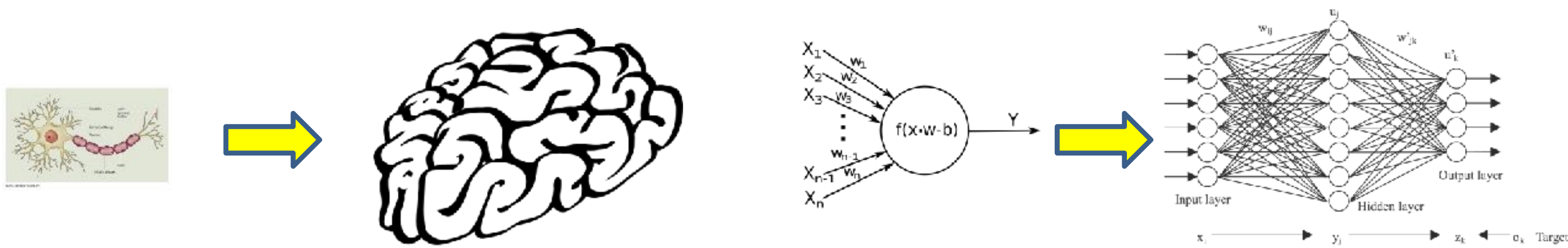
Not universal!



A single neuron is not enough

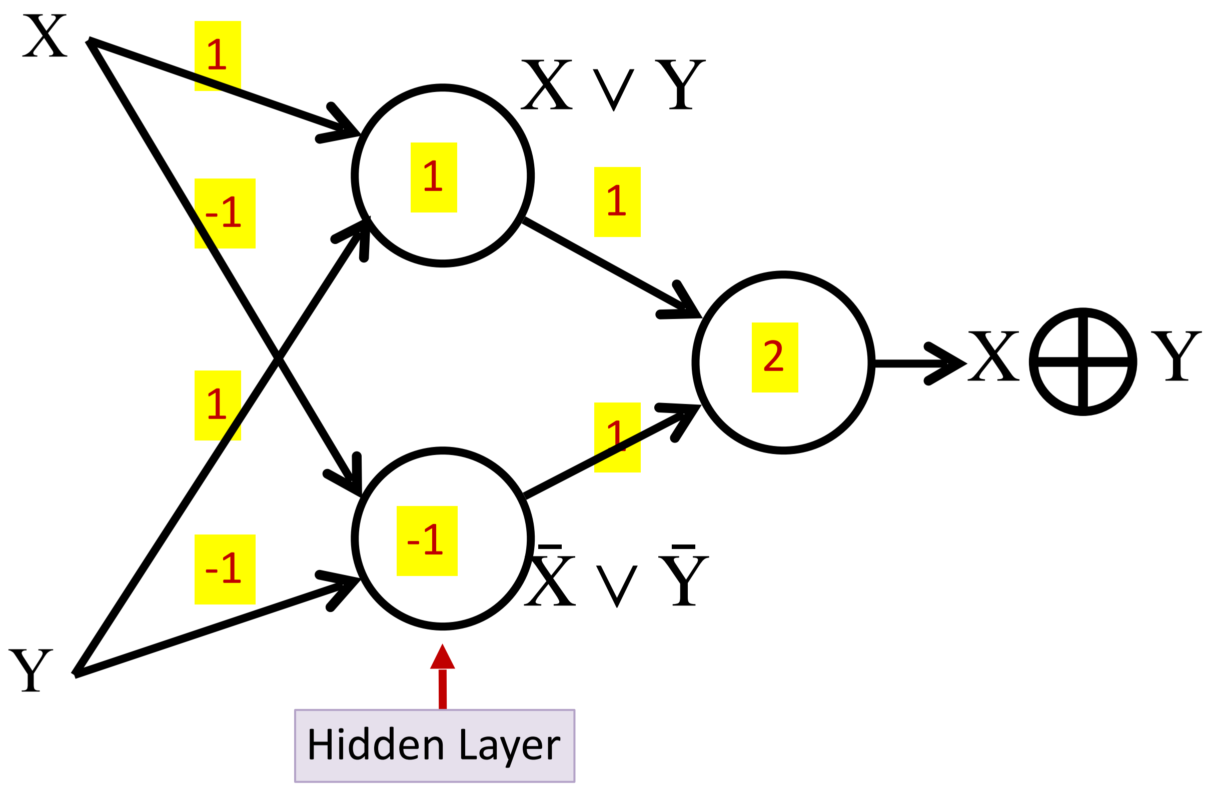
- Individual elements are weak computational elements
  – Marvin Minsky and Seymour Papert, 1969, *Perceptrons: An Introduction to Computational Geometry*

- *Networked* elements are required



Multi-layer Perceptron!

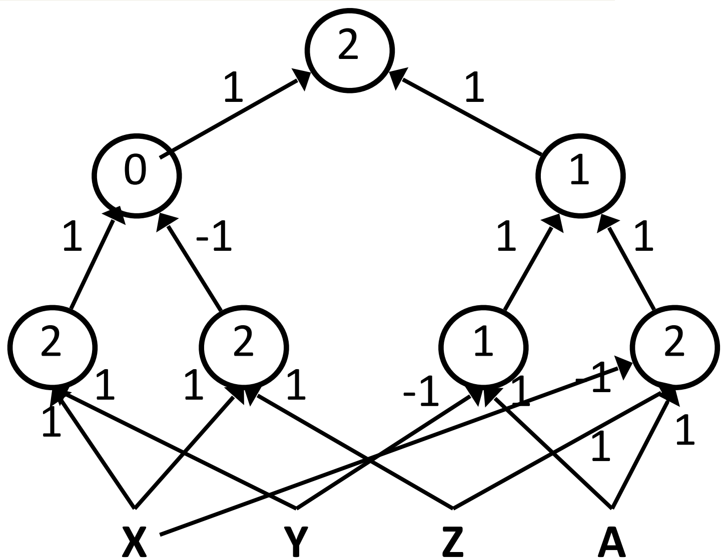
- XOR
  - The first layer is a "hidden" layer
  - Also originally suggested by Minsky and Papert 1968



A more generic model

- A “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

$$((A\bar{X}Z) \mid (A\bar{Y}))((X \ Y) \mid (\bar{X}Z))$$



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

Next Up

- What is Machine learning
- Linear Models
- More on neural networks as universal approximators
  - And the issue of depth in networks
  - How to train neural network from data