Introduction to Artificial Neural Networks

Dr. Ratna Babu Chinnam

Industrial & Systems Engineering

Wayne State University

Motivation to Study Artificial Neural Networks (ANNs)

- Humans generally outperform computers in certain tasks (Haykin 1994):
  - Image and video recognition and processing {100-200 ms}
  - Voice recognition
  - Pattern recognition
  - Motor control
  - Heuristic optimization
- Traditional computers are good at tasks based on precise and fast arithmetic operations, outperforming biological systems (Churchland, 1986)

Source: Simon Haykin (2009)

What is Unique About Biological Systems?

"Method" of Information Processing

Biological Systems:

- Massively parallel
  - ~10 Billion neurons and 60 trillion synapses for adults
- Analog in nature
- Adaptive and flexible
- 10^{-16} joules/operation/sec

Computers:

- Serial machines
- Digital processing
- 10^{-6} joules/operation/sec (Foggin, 91)

Source: Simon Haykin (2009)

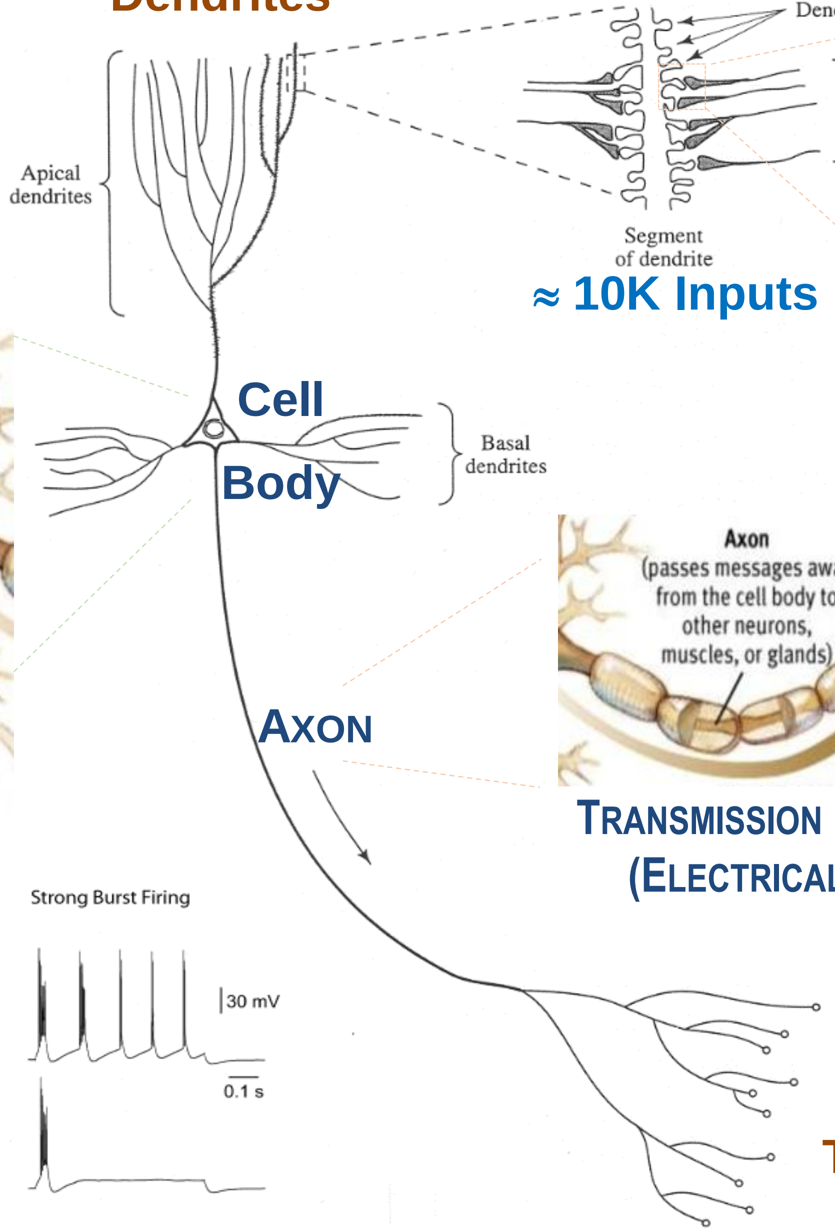
Results from Efforts to Imitate Biology (last century)

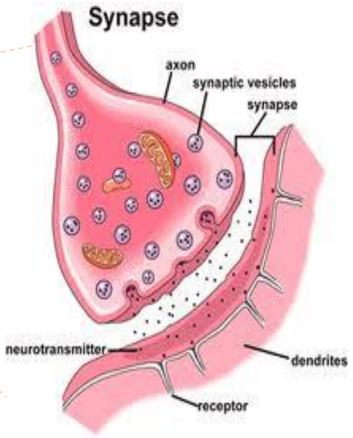
- Advances in applying ANNs for problems found intractable or difficult for traditional computation
  - Examples: Approximators, classifiers, restoration of patterns, optimization
- ANNs can supplement processing power of Von Neuman digital computer
- Tip of the "ice-berg"
  - Most ANN’s are being simulated on serial machines
  - 1k or 2k nodes vs 10 billion neurons and 60 trillion synapses
  - Simplistic structures/connectivity
  - Simplistic neuron models
  - Simplistic learning algorithms

(Founding Father: Raymon Cajal, 1911)

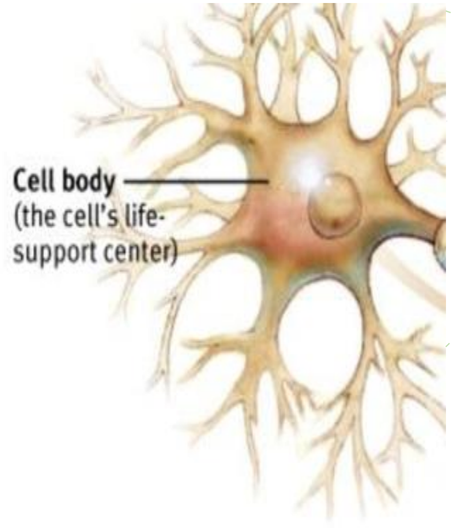
Pyramidal Neuron Cell

Dendrites



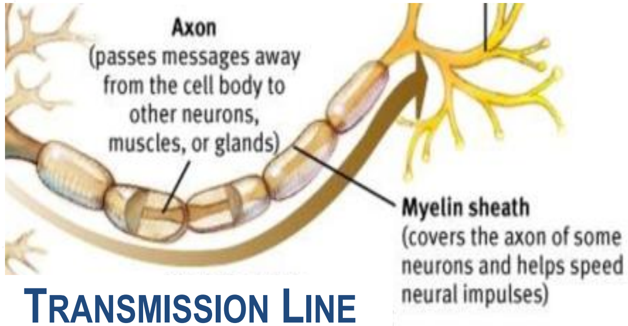


≈ 10K Inputs

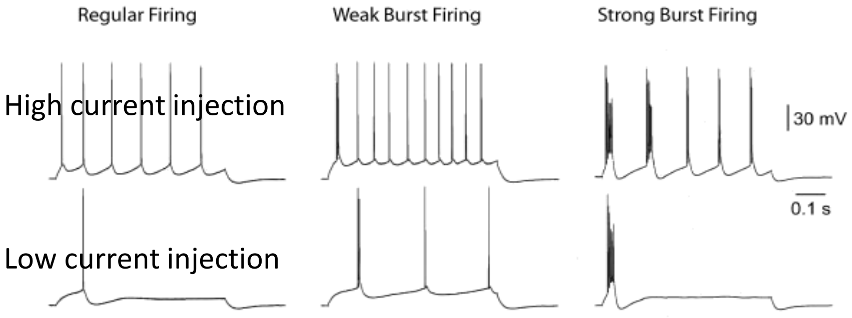


Cell Body

Axon



Transmission Line (Electrical)



Cell responses of 3 different neocortical layer V pyramidal neurons (Williams & Stuart, 1999)

Synaptic Terminals

≈ 1K Outputs

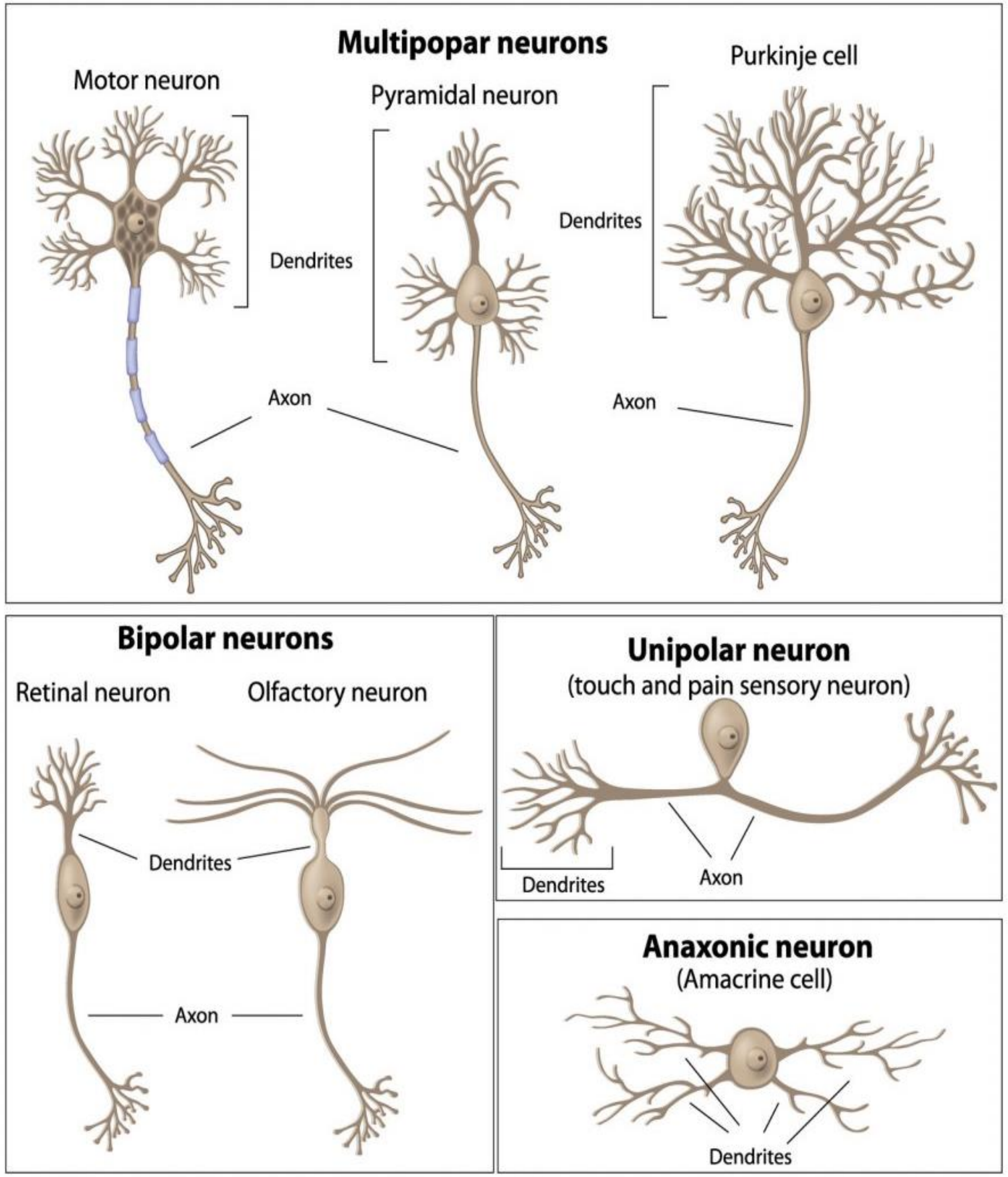
Different Types of Cells

CELLS

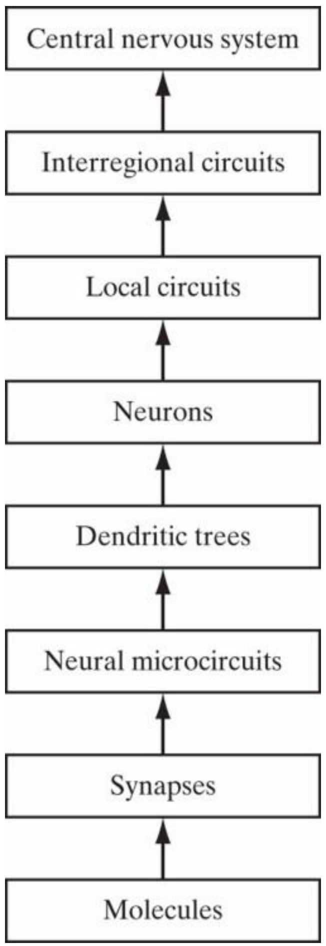
- Come in different sizes and shapes
- Exhibit plasticity

HUMAN CORTEX

- ~10 Billion Neurons
- ~60 Trillion Synapses



Cerebral Cortex



~1m^2
(Wrinkled)

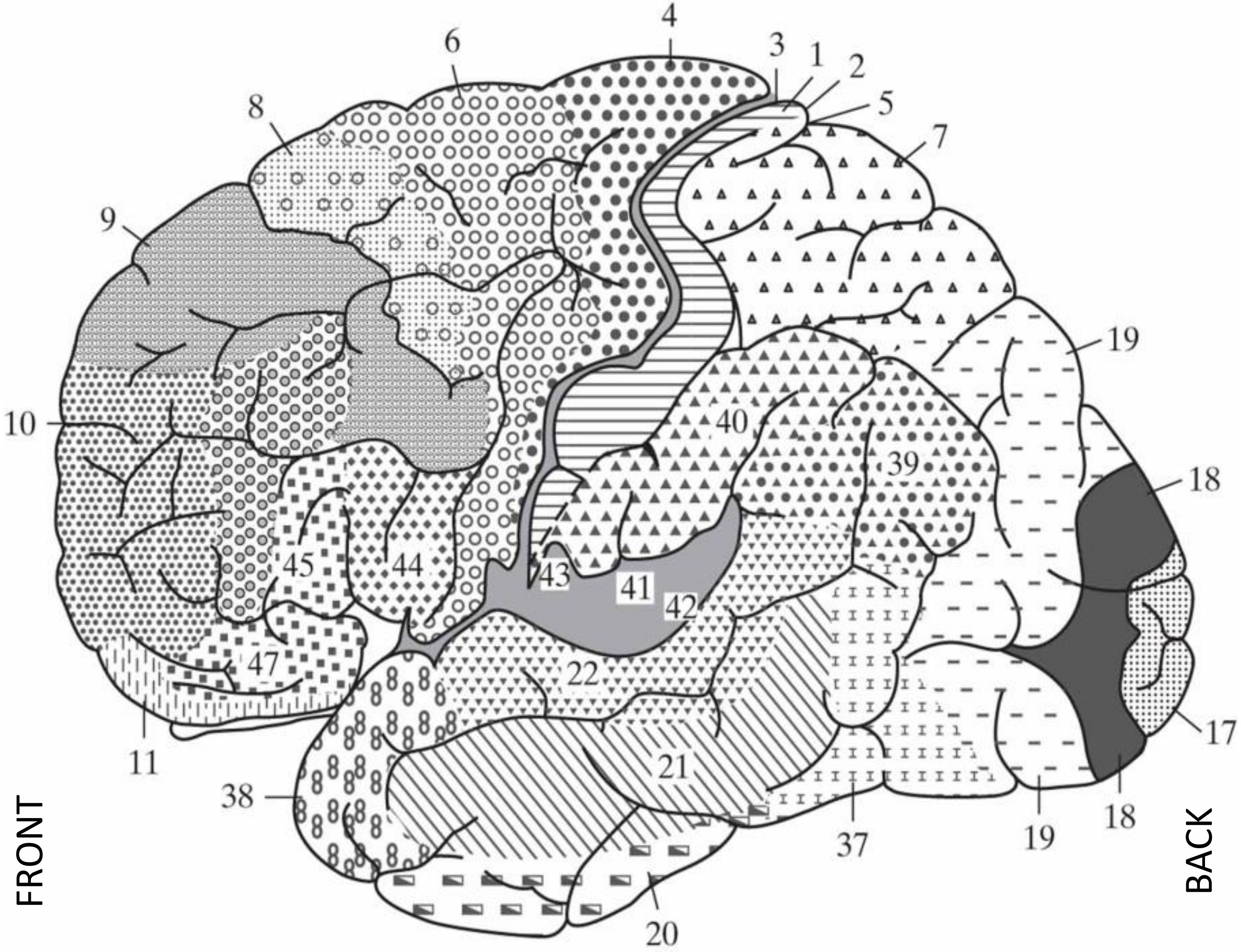
~1mm

~100μm

~1μm

~1nm

Structural Organization of Levels in the Brain



Different areas identified by thickness of layers and types of cells within.
Some key sensory areas are as follows: Motor cortex: motor strip, area 4; premotor area, area 6; frontal eye fields, area 8. Somatosensory cortex: areas 3, 1, and 2. Visual cortex: areas 17, 18, and 19. Auditory cortex: areas 41 and 42.

Characteristics of Most ANNs

- Function as parallel distributed computing networks
- Most basic characteristic is their architecture
- Provide either "instantaneous" or "dynamic" responses
- Contrast to conventional computers (which are programmed to perform specific tasks), most ANNs must be taught or trained (supervised learning or unsupervised learning)
- Resort to different "learning algorithms" and "learning modes"
- Exhibit different speeds, efficiency of learning, and effectiveness
- Learn new associations, new patterns, functional dependencies
- Different types of neuron models are popular

Source: Simon Haykin (2009)

Benefits of ANNs

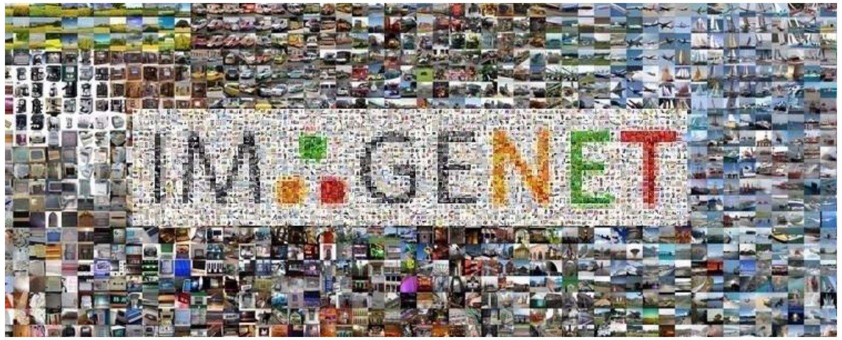
- Non-linearity
- Non-parametric
- Input-Output Mapping or Approximation
- VLSI Implementability
- Graceful degradation if implemented in hardware
- Uniformity of Analysis and Design
- Adaptivity while maintaining overall system "stability"
  - Stability-Plasticity dilemma (Grossberg, 1988)

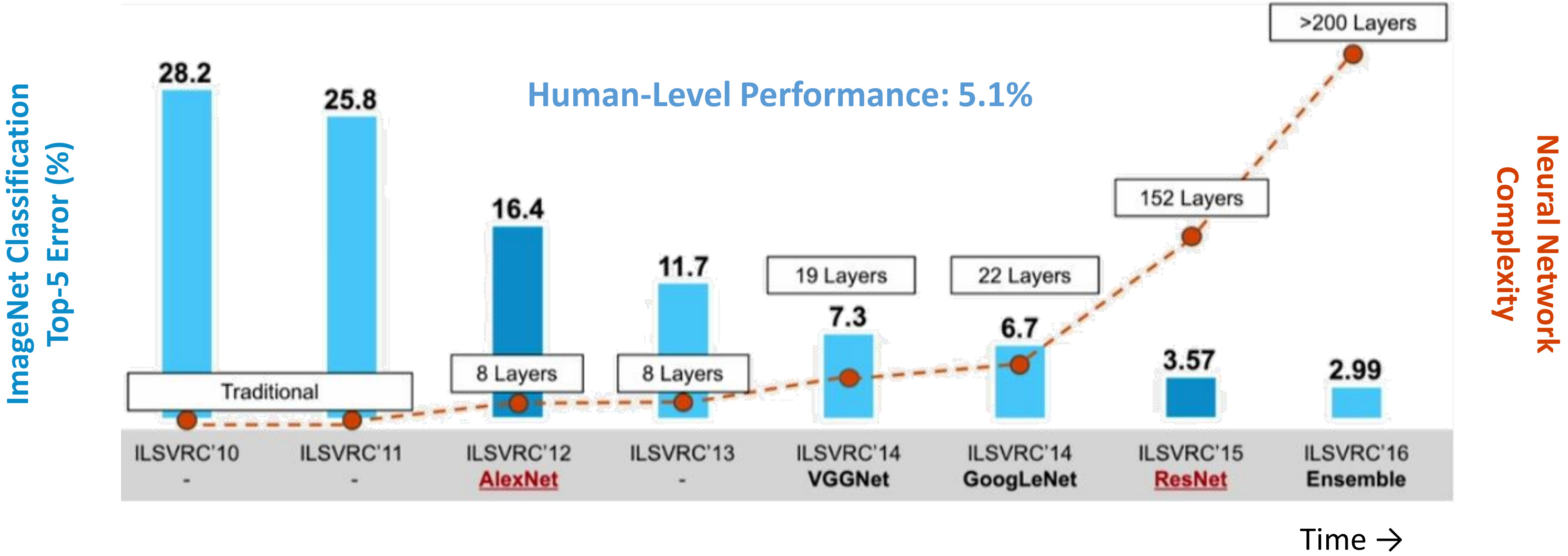
Weakness: General lack of confidence intervals

Source: Simon Haykin (2009)

Machine Surpasses Human-Level Image Classification

- IMAGENET:
  - Image Classification Challenge
  - 1,000 Object Classes
  - 1,431,167 Image



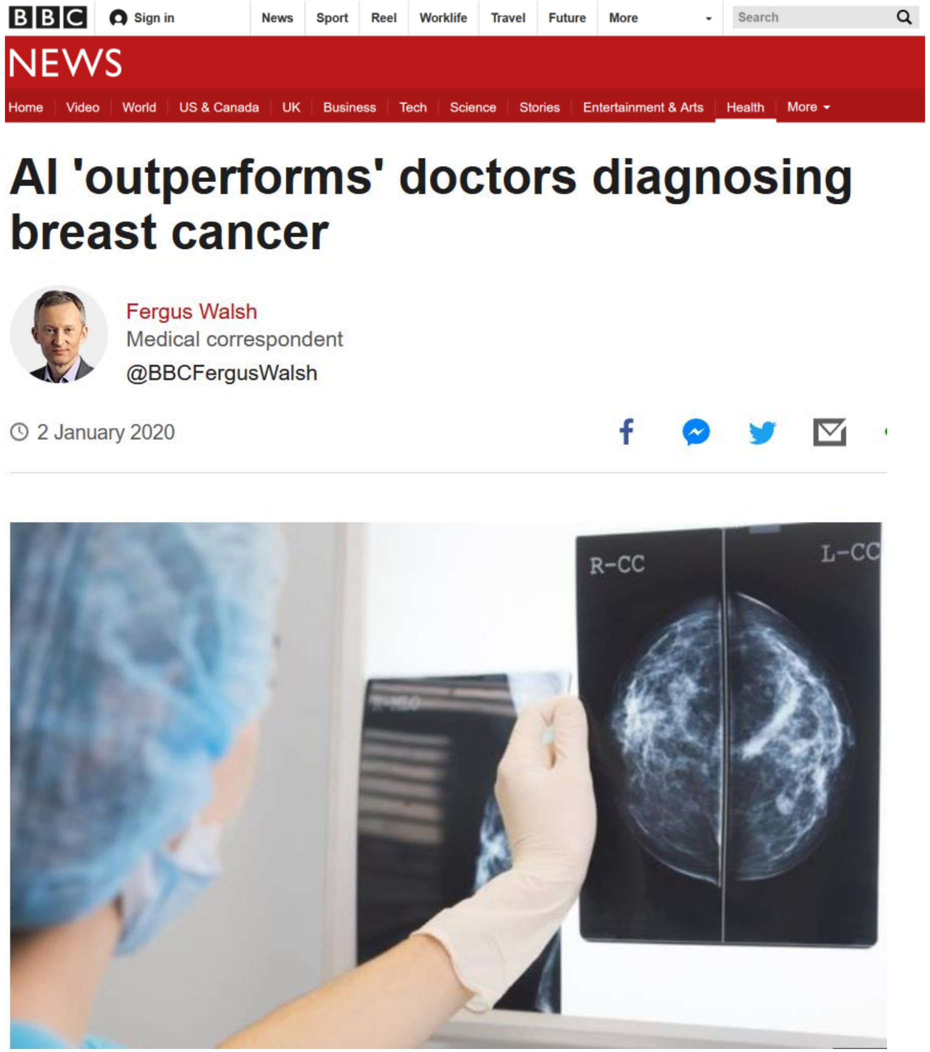


Top-5 error is the fraction of test samples where the correct label does not appear in the top 5 predicted results of the model when results are sorted in decreasing order of confidence.

IMAGENET Source: [Link]

Deep Learning for Image Processing

- AI is more accurate than doctors in diagnosing breast cancer from mammograms, a study in the journal Nature suggests.
- An international team, including researchers from Google Health and Imperial College London, designed and trained a computer model on X-ray images from nearly 29,000 women.
- Algorithm outperformed six radiologists in reading mammograms.
- AI was still as good as two doctors working together.
- Unlike humans, AI is tireless. Experts say it could improve detection.


BBC Sign in News Sport Reel Worklife Travel Future More Search

NEWS

Home Video World US & Canada UK Business Tech Science Stories Entertainment & Arts Health More

AI 'outperforms' doctors diagnosing breast cancer

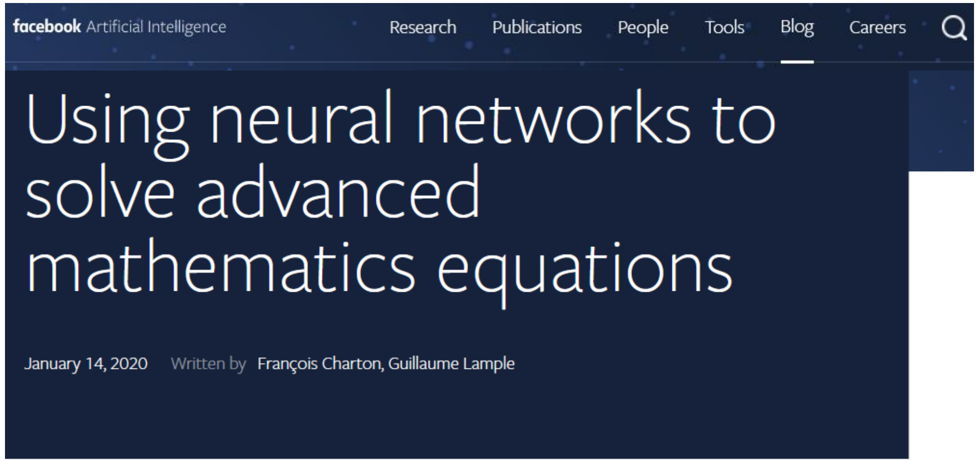
Fergus Walsh
Medical correspondent
@BBCFergusWalsh

2 January 2020

Source: BBC Article | Jan 2020 Nature Article

Deep Learning for Solving Math Problems

- Facebook AI has built the first AI system that can solve advanced mathematics equations using symbolic reasoning.
- By developing a new way to represent complex mathematical expressions as a kind of language and then treating solutions as a translation problem for sequence-to-sequence neural networks, we built a system that outperforms traditional computation systems at solving integration problems and both first- and second-order differential equations.
- Our model currently works on problems with a single variable, and we plan to expand it to multiple-variable equations. This approach could also be applied to other mathematics- and logic-based fields, such as physics, potentially leading to software that assists scientists in a broad range of work.



EQUATION	SOLUTION
$y' = \frac{16x^3 - 42x^2 + 2x}{(-16x^8 + 112x^7 - 204x^6 + 28x^5 - x^4 + 1)^{1/2}}$	$y = \sin^{-1}(4x^4 - 14x^3 + x^2)$
$3xy\cos(x) - \sqrt{9x^2\sin(x)^2 + 1}\,y' + 3y\sin(x) = 0$	$y = c\exp(\sinh^{-1}(3x\sin(x)))$
$4x^4yy'' - 8x^4y'^2 - 8x^3yy' - 3x^3y'' - 8x^2y^2 - 6x^2y' - 3x^2y'' - 9xy' - 3y = 0$	$y = \frac{c_1 + 3x + 3\log(x)}{x(c_2 + 4x)}$

Our model took the equations on the left as input — equations that both Mathematica and Matlab were unable to solve — and was able to find correct solutions (shown on the right) in less than one second.

Source: Facebook AI | Jan 2020 Blog

AI Epidemiologist Sent First Warnings of Wuhan Virus

- On January 9, the World Health Organization notified the public of [a flu-like outbreak in China]
  - A cluster of pneumonia cases had been reported in Wuhan, possibly from vendors' exposure to live animals at the Huanan Seafood Market
  - US CDC had gotten the word out a few days earlier, on January 6
  - But a Canadian health monitoring platform had beaten them both to the punch, sending word of the outbreak to its customers on December 31.
- [BlueDot] uses an AI-driven algorithm that scours foreign-language news reports, animal and plant disease networks, and official proclamations to give clients advance warning to avoid danger zones like Wuhan.

WIRED SIGN IN SUBSCRIBE

ERIC NIILER SCIENCE 01.25.2020 07:00 AM

An AI Epidemiologist Sent the First Warnings of the Wuhan Virus

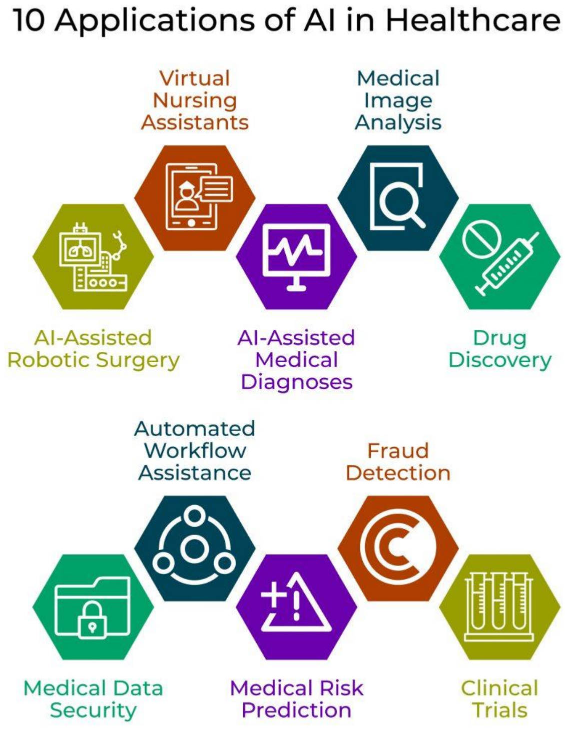
The BlueDot algorithm scours news reports and airline ticketing data to predict the spread of diseases like those linked to the flu outbreak in China.

PHOTOGRAPH: NICOLAS ASFOURI/GETTY IMAGES

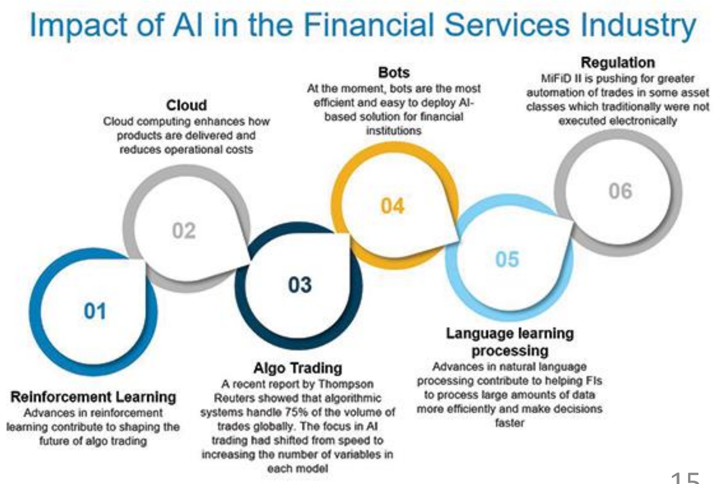
Source: Wired Magazine | Jan 2020 [Article]

Neural Network Applications: *Examples*

Industry	Business Applications Source: MathWorks (2012)
Aerospace	High-performance aircraft autopilot, flight path simulation, aircraft control systems, autopilot enhancements, aircraft component simulation, and aircraft component fault detection
Automotive	Automobile automatic guidance system, and warranty activity analysis
Banking	Check and other document reading and credit application evaluation
Defense	Weapon steering, target tracking, object discrimination, facial recognition, new kinds of sensors, sonar, radar and image signal processing, data compression, feature extraction and noise suppression
Electronics	Code sequence prediction, integrated circuit chip layout, process control, chip failure analysis, machine vision, voice synthesis, and nonlinear modeling
Entertainment	Animation, special effects, and market forecasting
Financial	Real estate appraisal, loan advising, mortgage screening, corporate bond rating, credit-line use analysis, credit card activity tracking, portfolio trading program, corporate financial analysis, and currency price prediction
Industrial	Prediction of industrial processes, such as the output gases of furnaces, replacing complex and costly equipment used for this purpose in the past
Insurance	Policy application evaluation and product optimization
Manufacturing	Manufacturing process control, product design and analysis, process and machine diagnosis, real-time particle identification, visual quality inspection systems, beer testing, welding quality analysis, paper quality prediction, computer-chip quality analysis, analysis of grinding operations, chemical product design analysis, machine maintenance analysis, project bidding, planning and management, and dynamic modeling of chemical process system
Medical	Breast cancer cell analysis, EEG and ECG analysis, prosthesis design, optimization of transplant times, hospital expense reduction, hospital quality improvement, and emergency-room test advisement
Oil and gas	Exploration
Robotics	Trajectory control, forklift robot, manipulator controllers, and vision systems
Securities	Market analysis, automatic bond rating, and stock trading advisory systems
Speech	Speech recognition, speech compression, vowel classification, and text-to-speech synthesis
Telecom	Image and data compression, automated information services, real-time translation of spoken language, and customer payment processing systems
Transportation	Truck brake diagnosis systems, vehicle scheduling, and routing systems



Source: IgniteOutsourcing.com



Source: William Benattar | Link

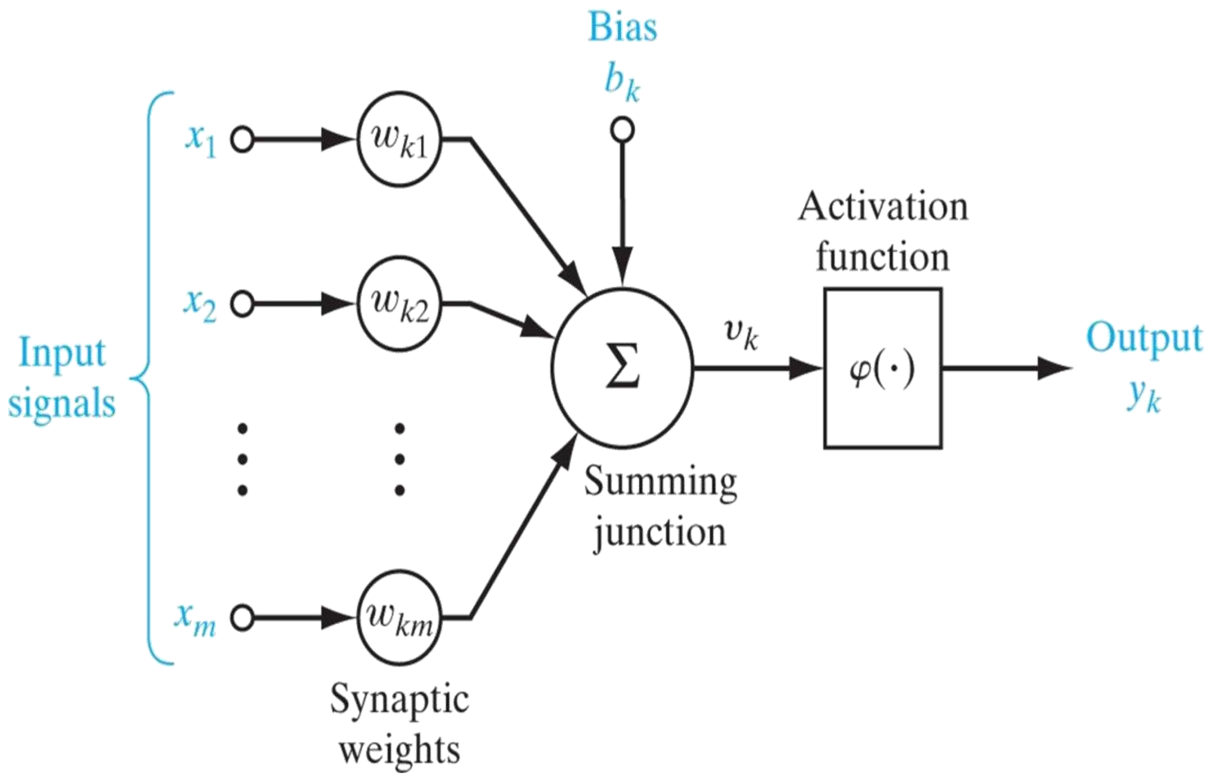
Definition of an ANN (Haykin, 1994)

- A "massively" parallel distributed processor that has a natural property for storing experiential knowledge and making it available for use.
- Employs "massive" interconnection of simple computing cells referred to as "neurons"
- Resembles the brain in two aspects:
  - Knowledge is acquired by the network through a learning process.
  - Interneuron connection strengths known as synaptic weights are used to store the knowledge.
- **Learning Process**: Procedure/algorithm used to modify the synaptic weights of network in an orderly fashion to attain a desired design objective

(Simple) Model of a Neuron

Three basic elements:

1. A set of ***synapses*** or connecting links, each of which is characterized by a *weight* or strength of its own.
   - Weight is +ve if associated synapse is excitatory; it is –ve if synapse is inhibitory.
2. An ***adder*** for summing input signals, weighted by respective weights.
3. An ***activation** function* for limiting the amplitude of output of a neuron and building non-linearity into the network.



Neuron Model

Source: Simon Haykin (2009)

(Simple) Model of a Neuron

Mathematical Terms:

1. Linear Combiner Output:

$$u_k = \sum_{j=1}^{m} w_{kj} x_j$$

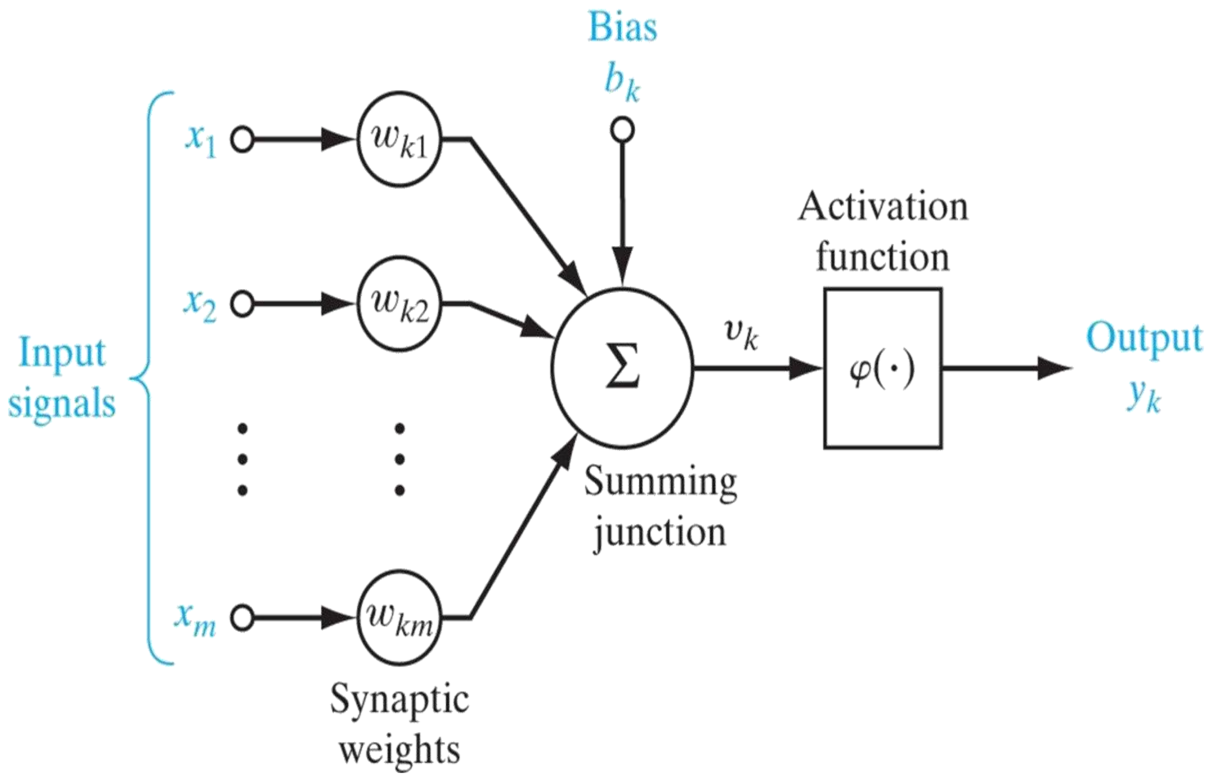
2. Activation Potential:

$$v_k = u_k + b_k \quad \text{or} \quad v_k = \sum_{j=0}^{m} w_{kj} x_j$$

where $x_0 = 1$ and $w_{k0} = b_k$

3. Neuron Output:

$$y_k = \varphi(v_k)$$



Neuron Model

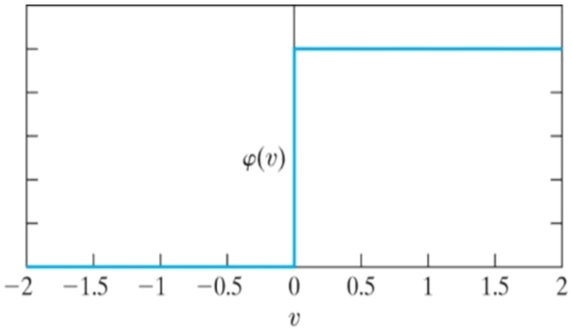
Bias allows the net activation potential to be nonzero even if all the inputs are zeroes.

Activation Functions

- Defines neuron output in terms of activity level at its input
- Some popular types: 1) Threshold (Hard-Limiter) Function, 2) Piecewise-Linear Function, 3) Sigmoid Function, 4) Radial Basis Function, 5) Rectilinear Unit, and 6) Softmax Function.

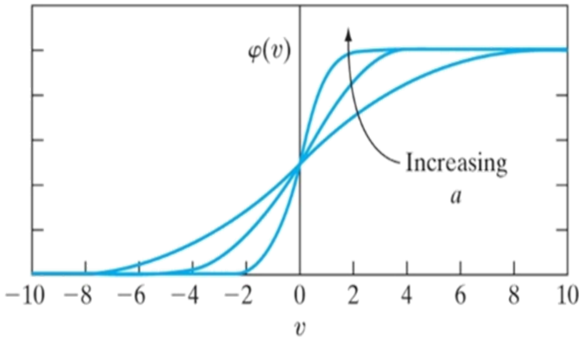
Threshold Functions:
An example:

$$\varphi(\vartheta) = \begin{cases} 1 & if \vartheta \geq 0 \\ 0 & if \vartheta < 0 \end{cases}$$



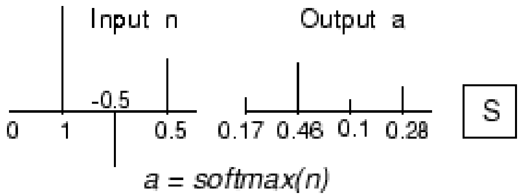
Sigmoid Functions:
Differentiable, strictly increasing, smooth, with asymptotic properties.
Example: Logistic function

$$\varphi(\vartheta) = \frac{b}{1 + e^{-a\vartheta}}$$



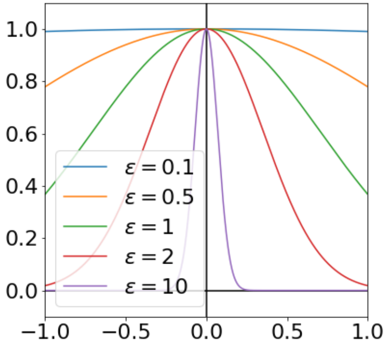
Softmax Function:
Generalization of logistic function that "squashes" a K-dimensional vector of arbitrary real values to a K-dimensional vector of values in the range (0,1) and add to 1:

$$P\left(a_j|\boldsymbol{n}\right) = \frac{e^{n_j}}{\sum_{k=1}^{K} e^{n_k}} \text{ for } j = 1, \ldots, K$$



Activation Functions

Radial Basis Function:
A function that depends only on distance from the origin or some point c, so that $\phi(x, c) = \phi(\|x - c\|)$. Norm is usually Euclidean. Commonly used examples include (writing $r = \|x - x_i\|$):



- Gaussian:
  The first term—that which is used for normalisation of the Gaussian—is missing because every Gaussian has a weight in our sum, thus the normalisation is not necessary.

$$\phi(r) = e^{-(\varepsilon r)^2}$$

- Multiquadric:

$$\phi(r) = \sqrt{1 + (\varepsilon r)^2}$$

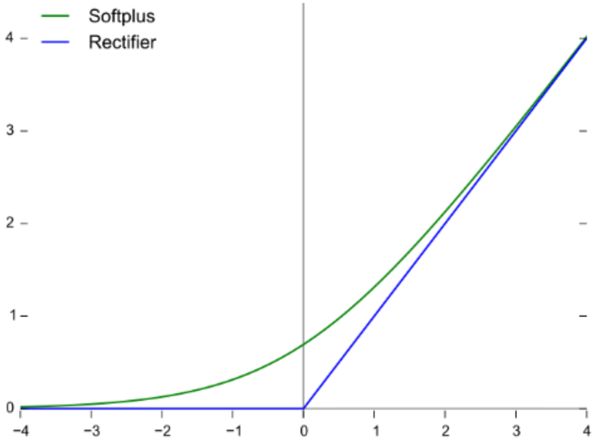
- Inverse quadratic:

$$\phi(r) = \frac{1}{1 + (\varepsilon r)^2}$$

- Inverse multiquadric:

$$\phi(r) = \frac{1}{\sqrt{1 + (\varepsilon r)^2}}$$

Rectilinear Unit (ReLU):
A function that retains the positive part of its argument. There are several variants.

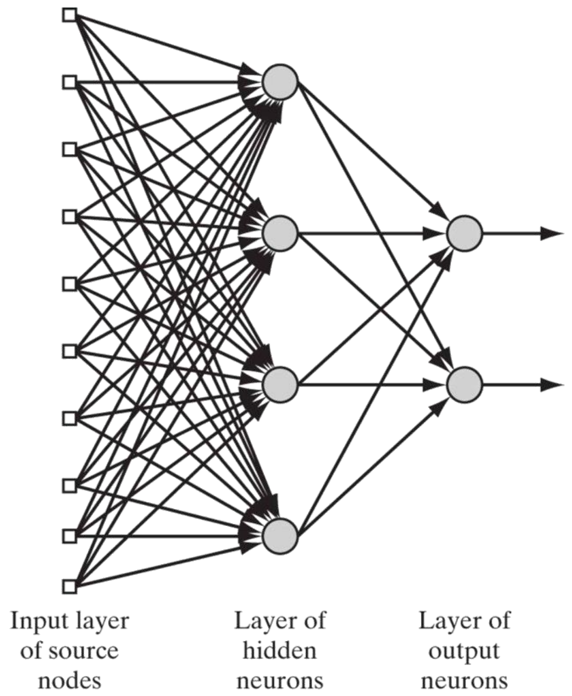


Source: Wikipedia | RBF Link | ReLU Link

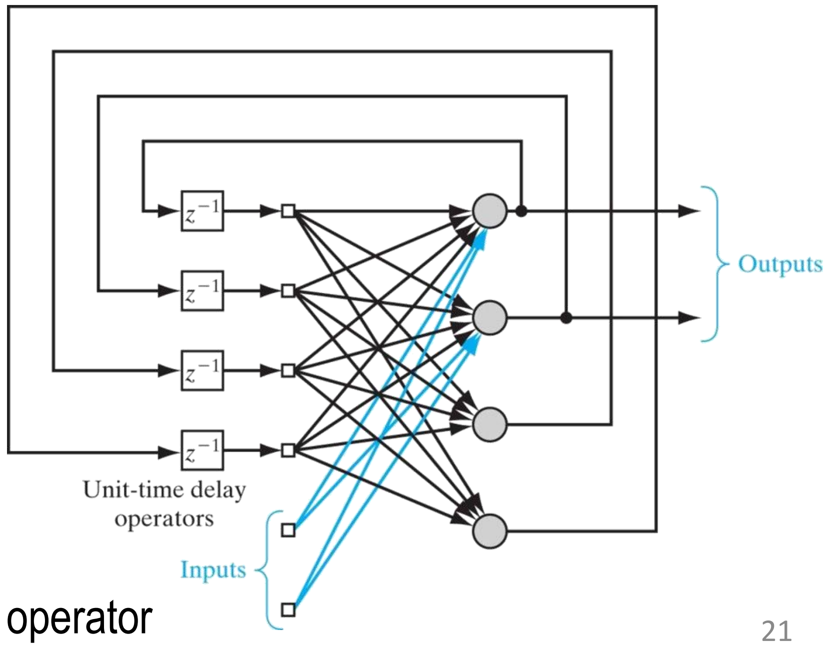
ANN Architecture

- Manner in which neurons of a network are connected
- Two classes of architectures:
- ***Feed-forward Networks***
  - Neurons are organized in layers
    - Input layer, multiple hidden layers, and output layer
  - Information strictly flows forward from input layer to output layer
- ***Recurrent Networks***
  - Has at least one feedback loop
  - Have strong dynamic properties

Applications: Function approximation, pattern recognition, time-series forecasting etc.



Applications: Modeling dynamic systems and optimization



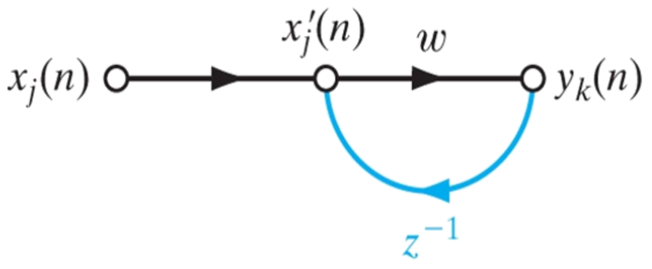
$z^{-1}x(n) = x(n-1)$

z^{-1} denotes a unit-delay operator

Source: Simon Haykin (2009)

Feedback

- Exists when output of an element in system influences in part input applied to that particular element
- Plays a major role in the study of recurrent networks
- Consider the following single-loop feedback system:



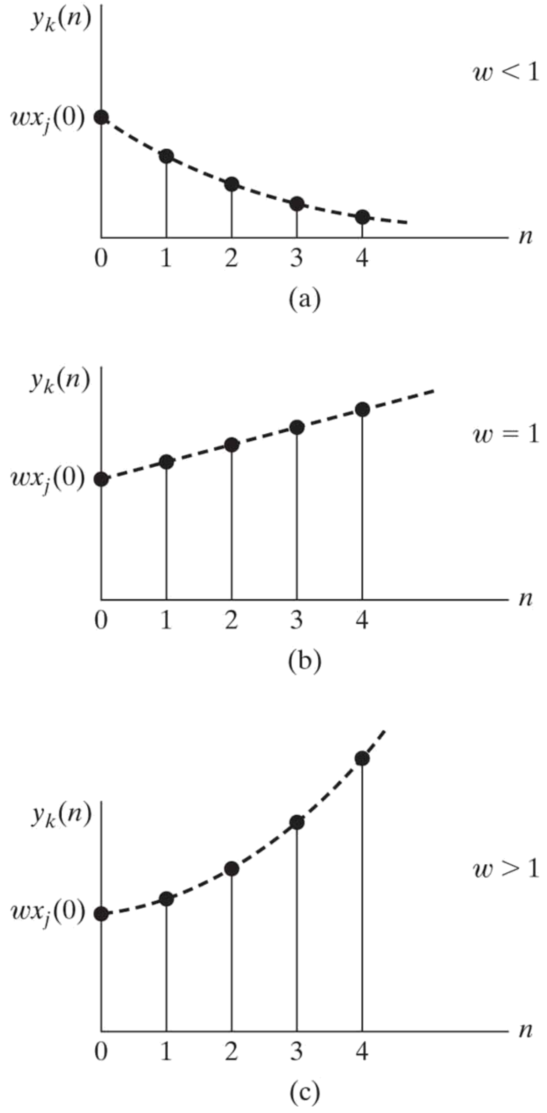
$$y_k(n) = w\left[x_j'(n)\right]$$

$$x_j'(n) = x_j(n) + z^{-1}y_k(n)$$

$$y_k(n) = \frac{w}{1 - wz^{-1}}\left[x_j(n)\right]$$

Binomial Expansion

$$y_k(n) = w\sum_{l=0}^{\infty} w^l z^{-l}\left[x_j(n)\right]$$



Time response for three different values of weight w. (a) Stable with infinite memory. (b) Linear divergence. (c) Exponential divergence.

Source: Simon Haykin (2009)

Example: Fit Simple Regression Model

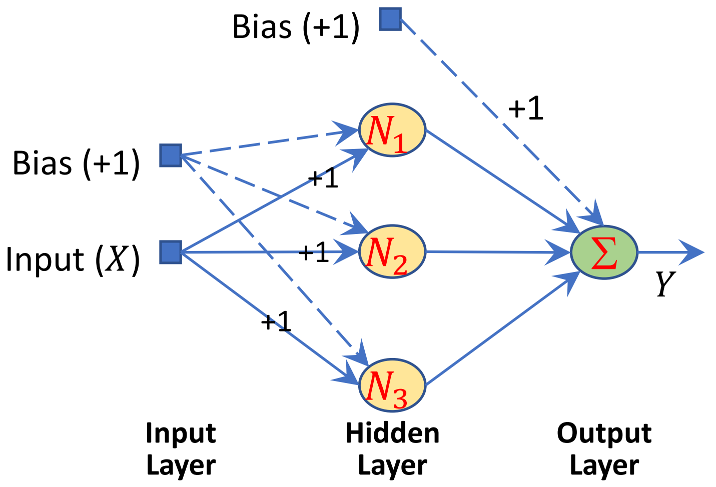
- **Network**: 1 Input & 1 Output
- Hidden Layer Neurons: 3
- Transfer Function:
  - Hidden: Logistic
  - Output: Identity

Logistic Transfer Function:

$$\varphi(\vartheta) = \frac{b}{1 + e^{-a\vartheta}}$$

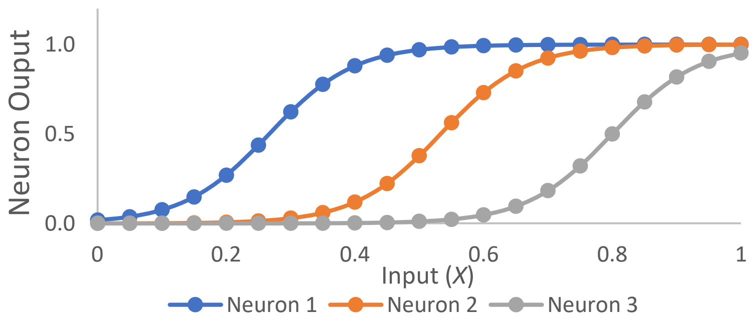
Scale Parameter

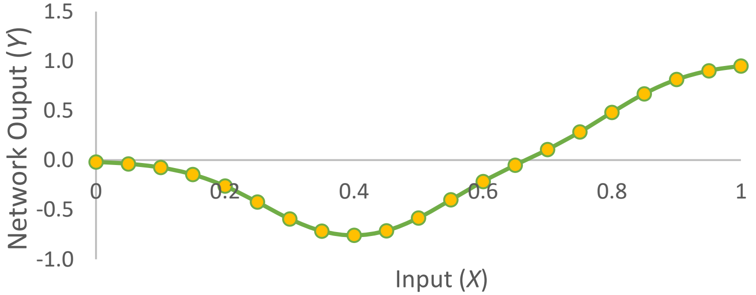
Shape Parameter



Setting: #1

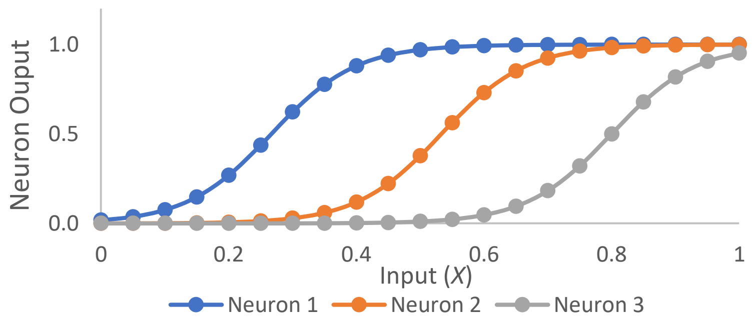
Hidden Neuron	N_1	N_2	N_3
Bias Weight	**4**	**8**	**12**
Shape Parameter	15	15	15
Scale Parameter	1	1	1
Output Weights	**-1**	**1**	**1**

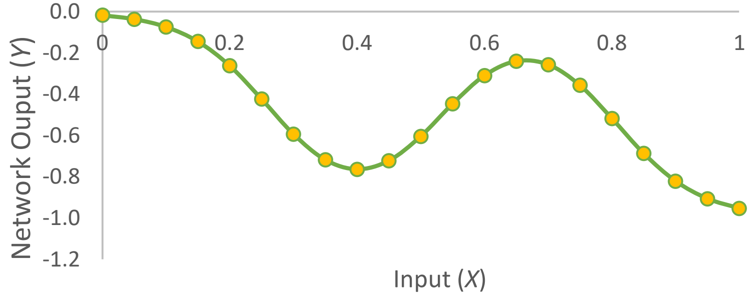




Setting: #2

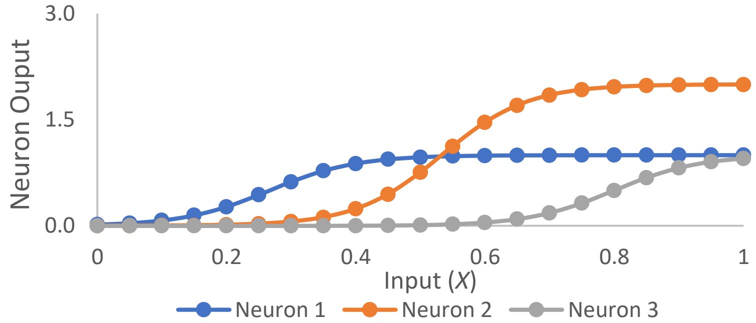
Bias Weight	4	8	12
Shape Parameter	15	15	15
Scale Parameter	1	1	1
Output Weights	-1	1	**-1**

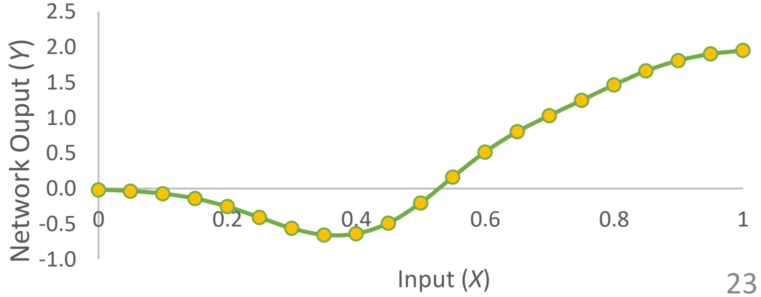




Setting: #3

Bias Weight	4	8	12
Shape Parameter	15	15	15
Scale Parameter	1	**2**	1
Output Weights	-1	1	1





Knowledge Representation in ANNs

- Knowledge can be classified into two kinds of information:
  - Known facts (*prior* information)
  - Observations and measurement (inherently noisy)
- Complicated topic that can only be addressed effectively one application at a time
- **Network Design:** Generally involves these steps:
  - Selection of appropriate architecture
  - Collection of example patterns representative of environment (both +ve and -ve examples)
  - Training network by means of a suitable algorithm
  - Testing network performance for generalization

Source: Simon Haykin (2009)

Knowledge Representation in ANNs …

- Four general rules (pattern classifier metaphor) (Anderson, 1988):
  1. Inputs from "similar" classes should usually produce similar representations inside network
  2. Items from "separate" classes should be given widely different representations in the network
  3. Important features should be allocated a large number of neurons in the network for that feature's representation
  4. *Prior information* and *invariances* should be built into the design of a neural network and not learn them
     - Results in fewer network parameters and a faster/cheaper network
- Image Recognition Example:
  - Prior Information – Two dimensional image, BW/Color, Critical Features, …
  - Invariance Requirements – Invariance to rotation, sharpness, brightness, …

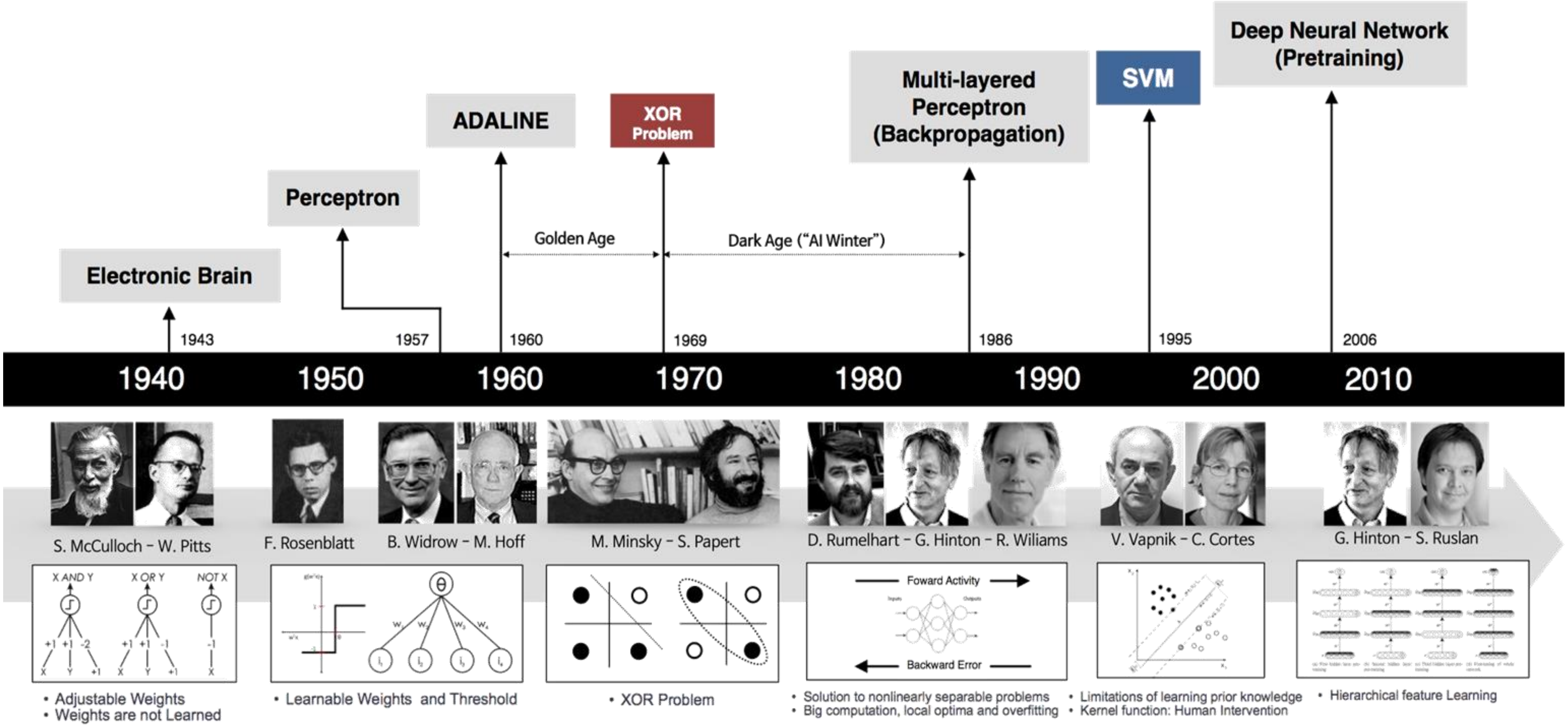
Source: Simon Haykin (2009)

Knowledge Representation in ANNs …

- In general, prior information can be built into ANNs in two ways:
  1. Through network architecture restrictions
  2. Through synaptic weight constraints
- In general, invariances can be built into the ANNs in three ways:
  1. Through structure (architecture restrictions and synaptic weight constraints)
  2. Through training
  3. Through invariant feature space

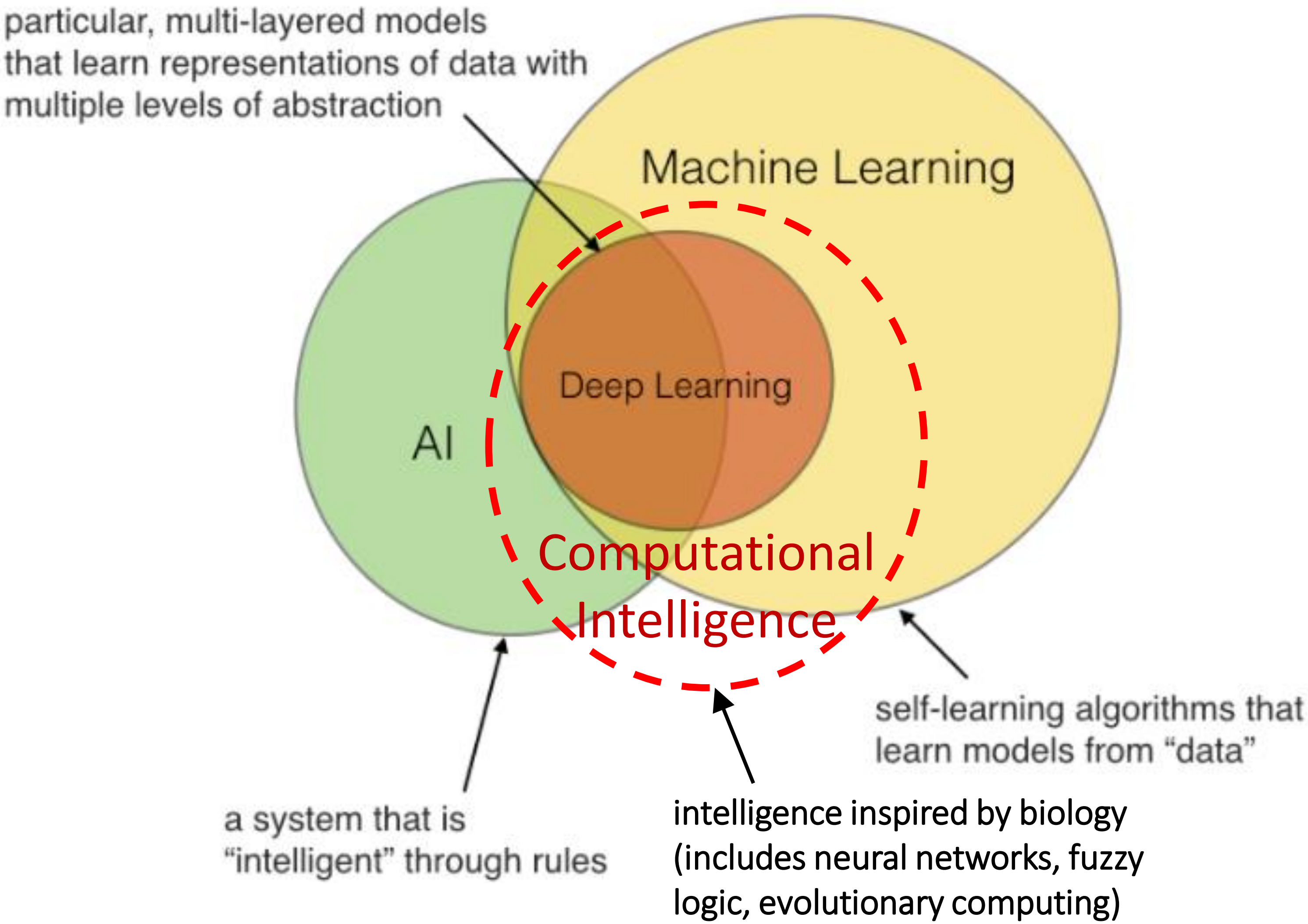
Source: Simon Haykin (2009)

Neural Networks Timeline



SOURCE: Unknown

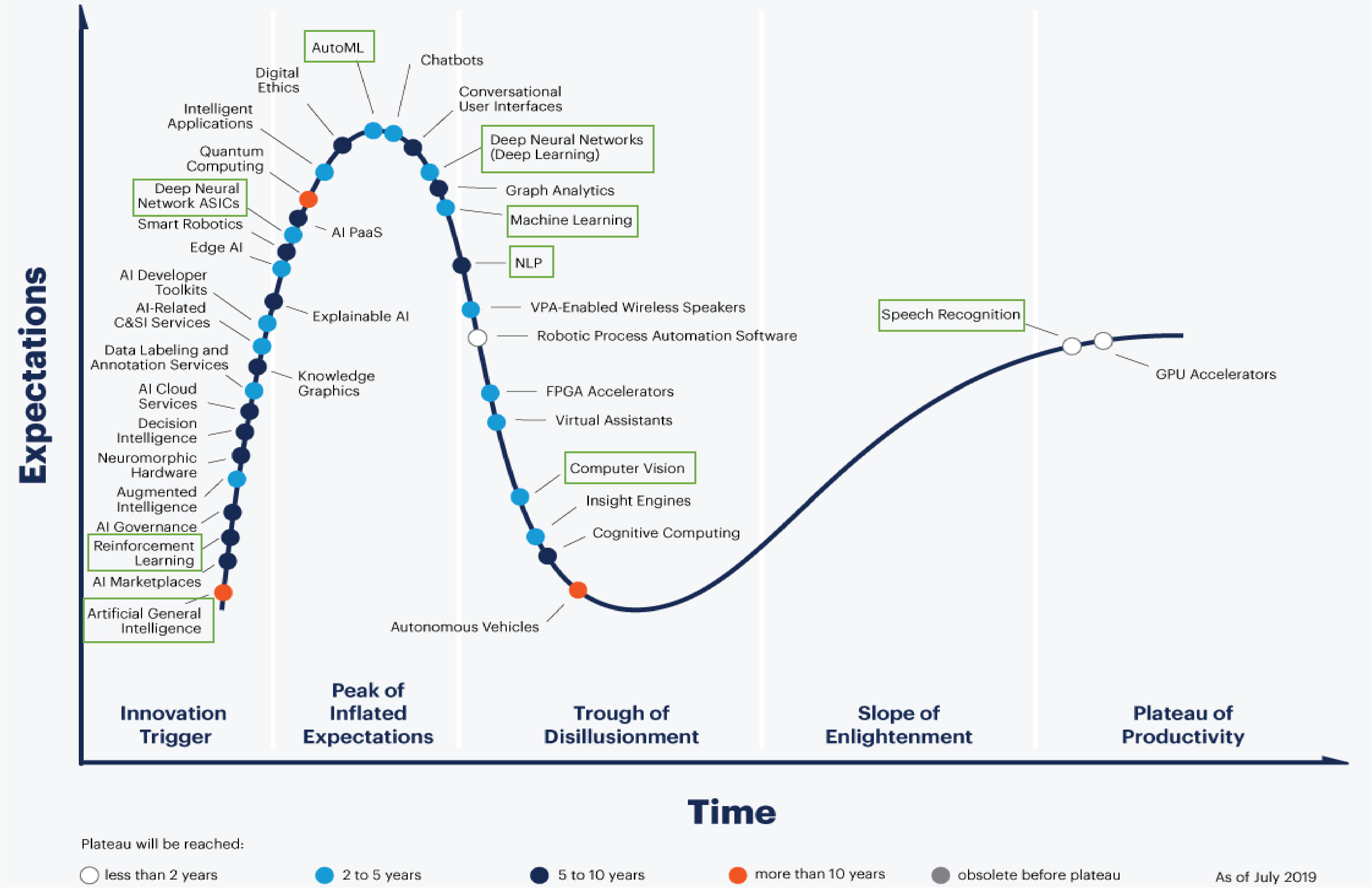
AI vs Machine Learning vs CI vs Deep Learning



AI: Artificial intelligence
CI: Computational intelligence

Source: J. Chu, Understanding AI, Machine Learning, and Deep Learning | [LINK]

Gartner's Hype Cycle for AI (2019)



Source: Gartner Hype Cycle for Artificial Intelligence, 2019 | [LINK]

Neural Network Hardware: *Examples*

THE FASTEST PATH TO DEEP LEARNING

Fast-track your initiative with a solution that works right out of the box, so you can gain insights in hours instead of weeks or months. NVIDIA® DGX-1™ is the integrated software and hardware system that supports your commitment to AI research with an optimized combination of compute power, software and deep learning performance.

https://developer.nvidia.com/deep-learning

DISCOVER THE BENEFITS

TURN DATA INTO KNOWLEDGE

NVIDIA DGX-1 lets you quickly uncover patterns in large data sets leading to new knowledge and insights in hours or minutes.

DGX-1 Infographic

BHGE customer story

EFFORTLESS PRODUCTIVITY

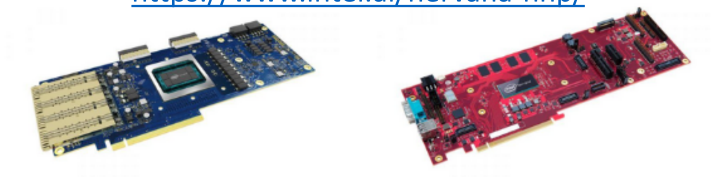
Simplified workflows and improved team collaboration make you more productive immediately. DGX-1 is designed to save you from the typical deep learning setup costs, which can be hundreds of thousands of dollars in software engineering hours, and months of delays for the open source software to stabilize.

INTEL® NERVANA™ NEURAL NETWORK PROCESSORS

See the design philosophy and research behind the Intel® Nervana™ Neural Network Processors, designed from the ground up for deep learning training and inference at massive scale.

INTEL® NERVANA™ NNP ARCHITECTURE REVEALED AT HOT CHIPS 2019

https://www.intel.ai/nervana-nnp/

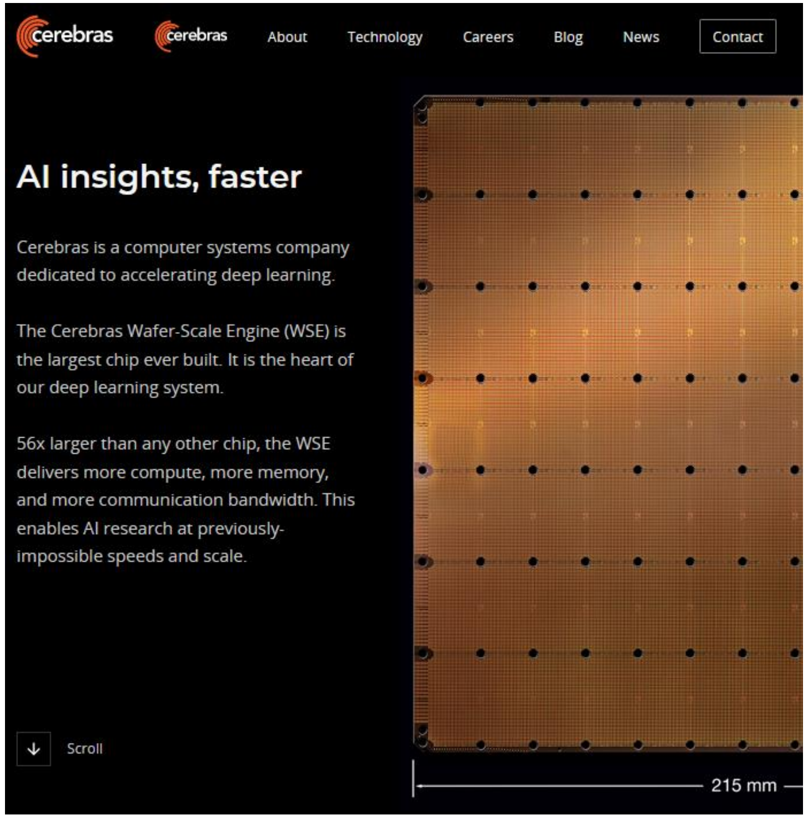


INTEL® NERVANA™ NEURAL NETWORK PROCESSORS FOR TRAINING (NNP-T)

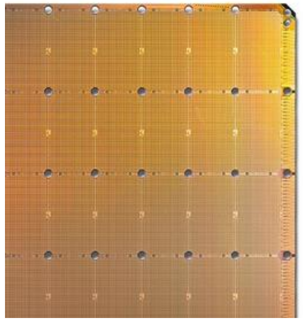
To quickly process vast, sparse, or complex data for large models within a power budget, AI hardware must deliver a critical balance of compute, communication, and memory. The Intel® Nervana™ Neural Network Processor for Training (Intel® Nervana™ NNP-T) does just that. With an all-new architecture that maximizes the re-use of on-die data, the NNP-T was purpose-built to train complex deep learning models at massive scale, and simplify distributed training with out-of-the-box scale-out support.

INTEL® NERVANA™ NEURAL NETWORK PROCESSORS FOR INFERENCE (NNP-I)

Enterprise-scale AI deployments are significantly increasing the volume of inference cycles, while demanding ever-stricter latency requirements. The Intel® Nervana™ Neural Network Processor for Inference (Intel® Nervana™ NNP-I) was built for this intensive, near-real-time, high-volume compute. By combining a CPU core and purpose-built AI inferencing engine, NNP-I delivers the novel hardware architecture that emerging, increasingly complex use cases demand, turning customer data into knowledge with an incredibly efficient, mutli-modal inferencing solution.



https://www.cerebras.net/



The WSE is the largest chip ever built

56x the size of the largest GPU

Challenges with Deep Learning & AI

Jerome Pesenti, VP of AI @ Facebook

Source: Wired Interview with J Pesenti, Dec 2019 | LINK

- IBM Watson called out that this is a commercial market and that there are applications. It was remarkable. But there was too much overhyping. I don't think that served IBM very well.
- Deep learning and current AI has a lot of limitations. We are very very far from human intelligence. It can propagate human biases, it's not easy to explain, it doesn't have common sense, it's more on the level of pattern matching than robust semantic understanding.
  - But we're making progress and the field is still progressing pretty fast.
- AI "reproducibility crisis," or the difficulty of recreating groundbreaking research. It's something that Facebook AI is very passionate about. When people do things that are not reproducible, it creates a lot of challenges. It's a lot of lost investment.
- OpenAI noted that compute power required for advanced AI is doubling every 3.5 months.
  - When you scale deep learning, it tends to behave better and to be able to solve a broader task in a better way. But clearly the rate of progress is not sustainable. Nobody can afford that.
  - We really need to think in terms of optimization, in terms of cost benefit, and we really need to look at how we get most out of the compute we have. This is the world we are going into.

Good Article: Gary Marcus, An Epidemic of AI Misinformation, The Gradient, 30 Nov 2019 | Link

On-Demand Web Based Access to Grid Computing

Example:
Jupyter Notebooks for Python & R with GPUs

- Visit OnDemand [Website](https://ondemand.grid.wayne.edu)
- Select Interactive Apps
- Specify Compute Requirements
- Launch
- Session Ready in Minutes
- Upload necessary code/data
- Compute away!