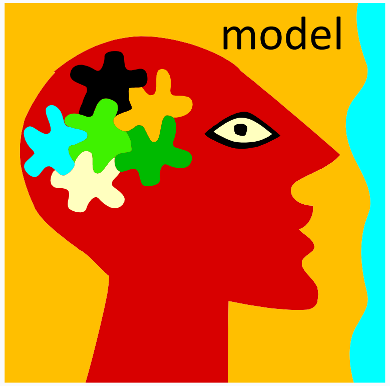
Machine Learning and its Application in Sciences

An introduction

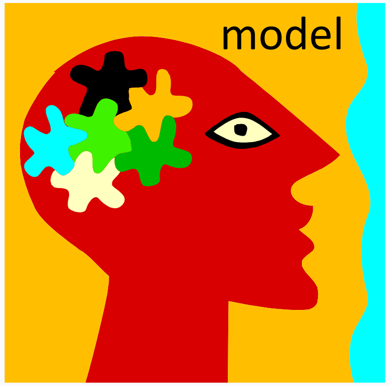
Evolution of (data) sciences: stone ages



observation

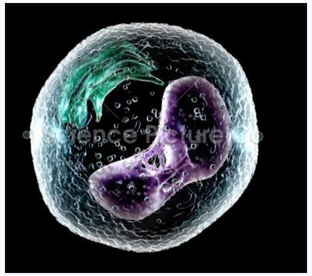
We used only our senes tobserve the word and our mind to make sense of it.

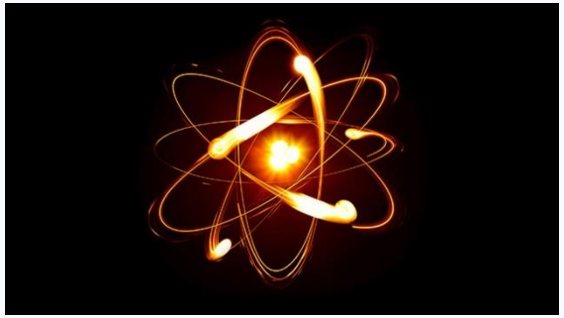
Evolution of (data) sciences: stone ages



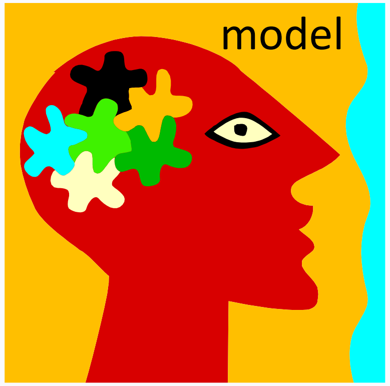
observation

We cannot see cells, atoms or faint far away galaxies. Cannot see UV, Xray, radio, …





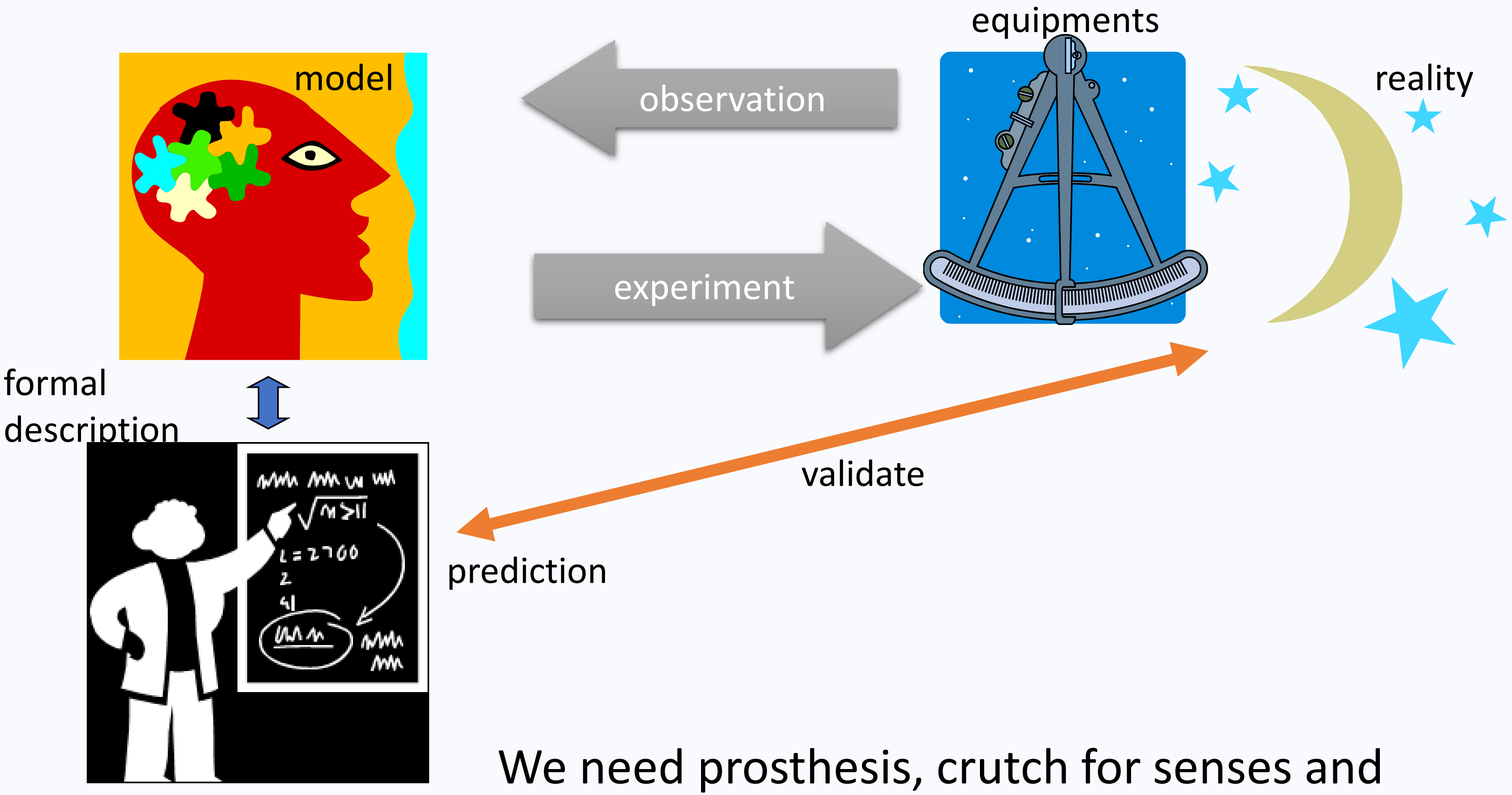
Evolution of (data) sciences: stone ages



observation

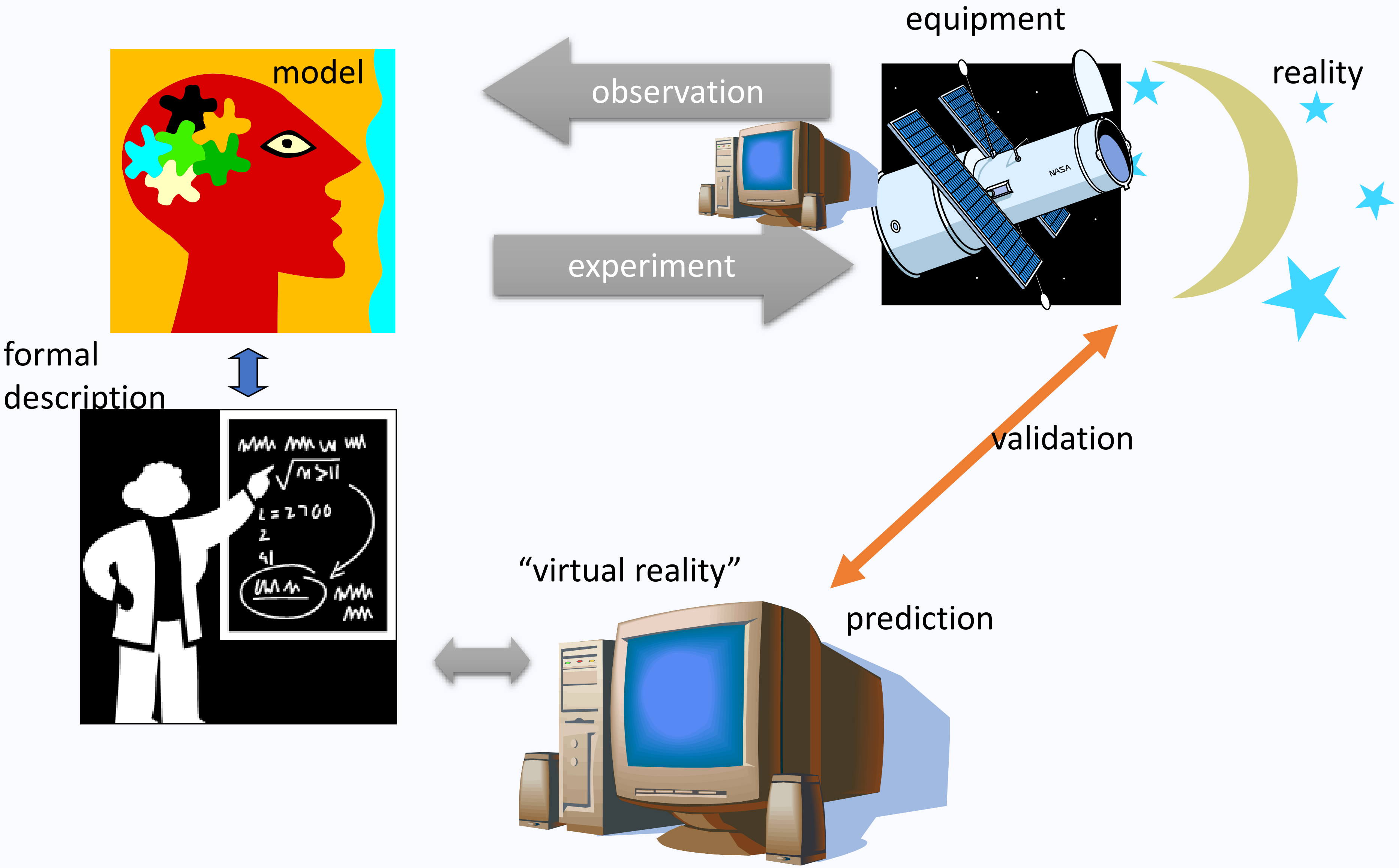
Our senses were „designed” for different purposes: finding food/mate, avoid danger. Though we are proud of it, our brain was not „designed” to do science. It takes 6 years to teach a child to count up to 20, and do elementary operations on numbers. Some of the population never masters much more in math. A simple calculator can do much better.

Evolution of (data) sciences: pre-industrial ages



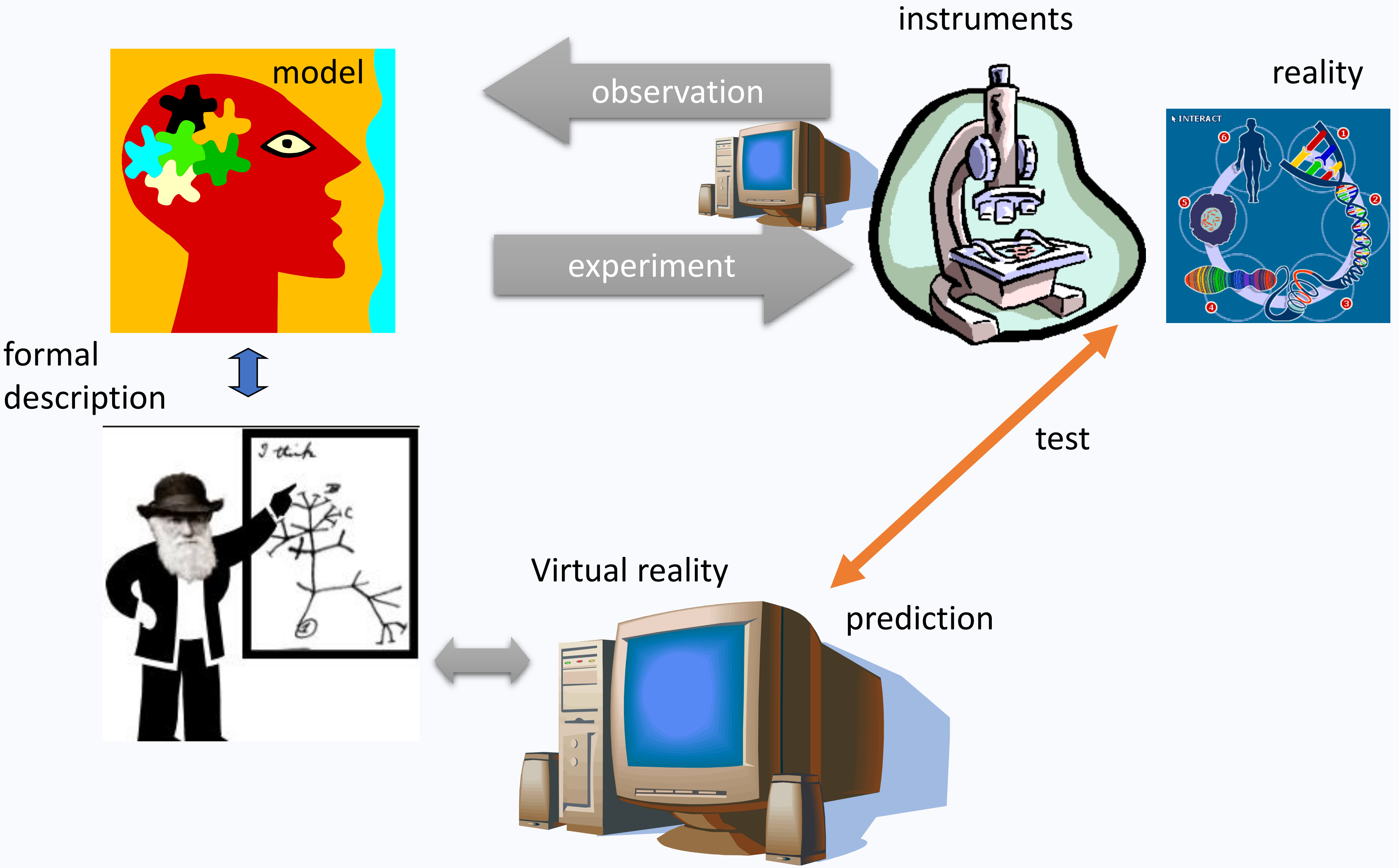
We need prosthesis, crutch for senses and mind!

Evolution of (data) sciences: present



With high throughput equipment and fast computers we have significantly extended our senses and mind.

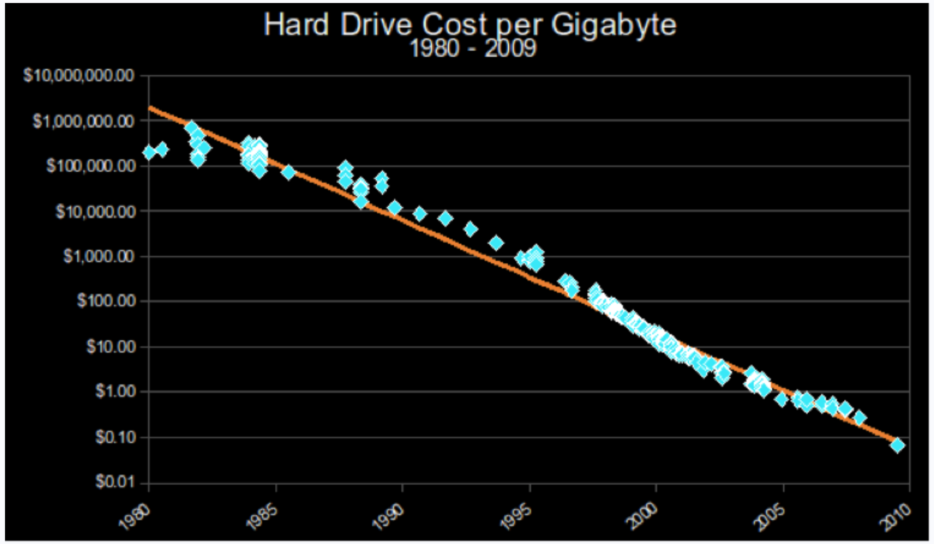
Evolution of (data) sciences: present



With high throughput equipment and fast computers we have significantly extended our senses and mind.

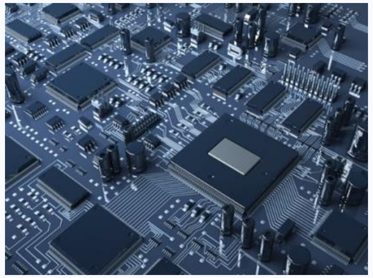
Exponentially cheaper devices – exponentially more data

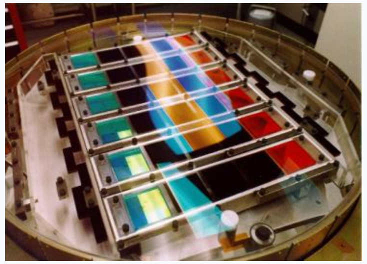
Quantum mechanics → Solid-state physics → Transistor/ Microelectronics



Moore's law

Elektronics → Sensors → Data



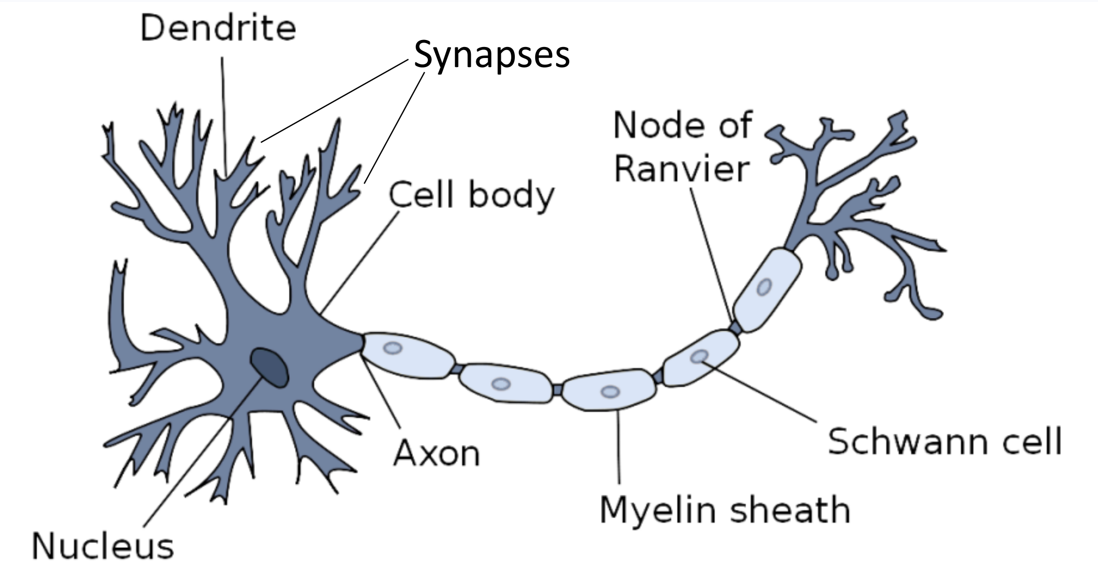


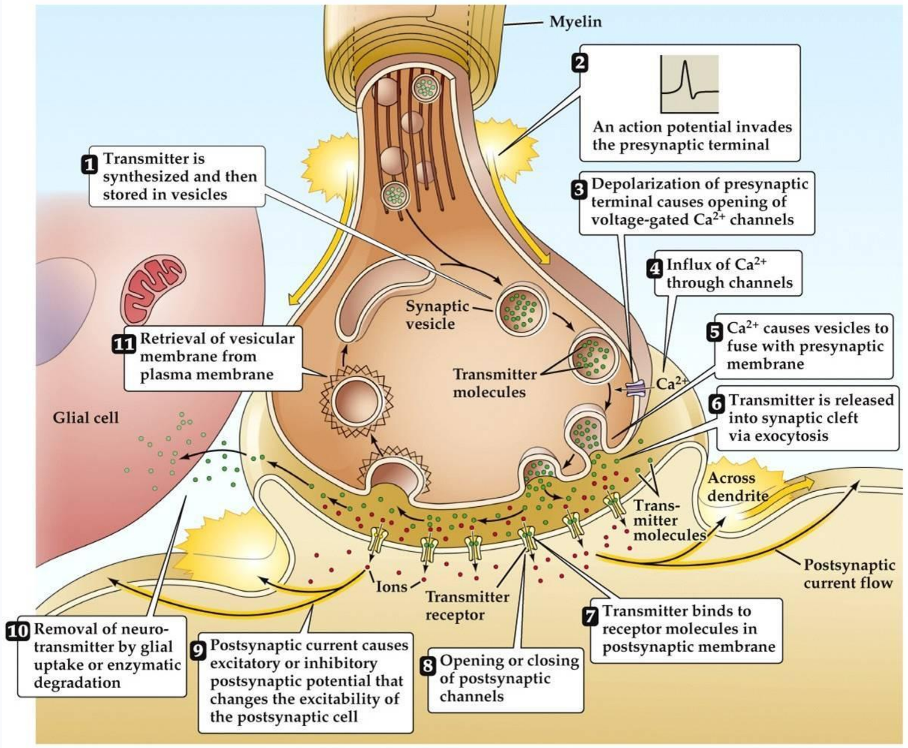
„Biologically inspired" approach

- Idea: mimic the brain's
  - flexibility
  - self-organization
  - self-learning
  - generalization
  - massive parallelism
  - fault-tolerance

	Brain	Computer
No. of processing units	$\approx 10^{11}$	$\approx 10^{9}$
Type of processing units	Neurons	Transistors
Type of calculation	massively parallel	usually serial
Data storage	associative	address-based
Switching time	$\approx 10^{-3}\mathrm{s}$	$\approx 10^{-9}\mathrm{s}$
Possible switching operations	$\approx 10^{13}\frac{1}{s}$	$\approx 10^{18}\frac{1}{s}$
Actual switching operations	$\approx 10^{12}\frac{1}{s}$	$\approx 10^{10}\frac{1}{s}$

Too complex
and not well understood enough
for direct implementation



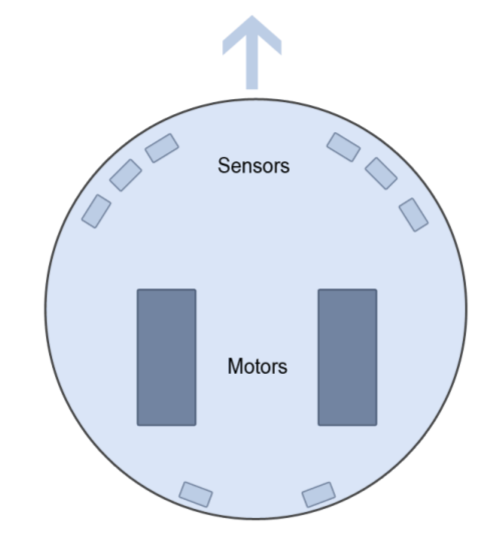


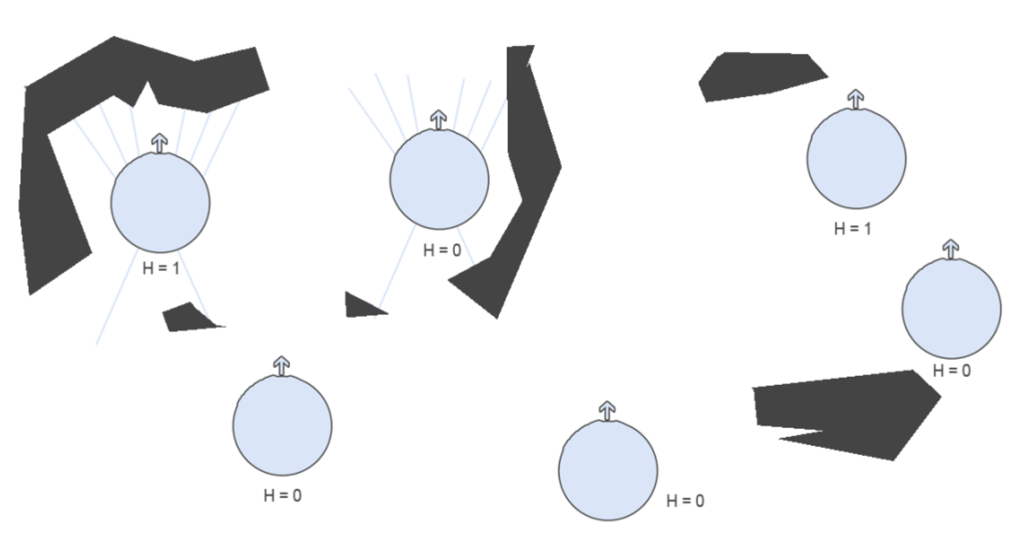
NEUROSCIENCE 5e, **Figure 5.3**

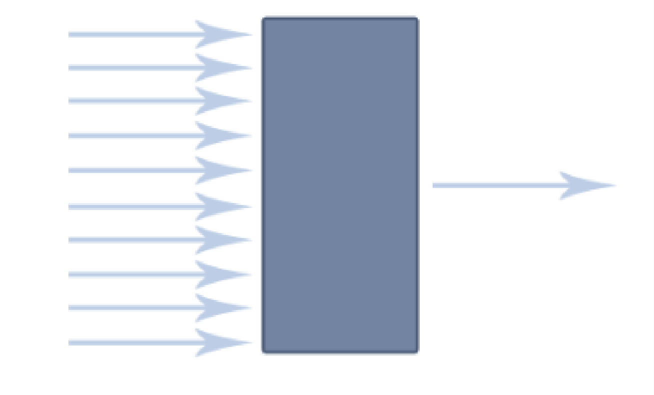
Using figures from: D. Kriesel – A Brief Introduction to Neural Networks (ZETA2-EN): http://www.dkriesel.com/en/science/neural_networks

The black box concept

- Robot example
  - Input: 8 sensors
  - Output: stop or go
- Goal: avoid obstacles
- Learning => regression
- There are several „machine learning” approaches, neural net is just one of them
- Will not cover:
  - SVM, random forest, decision trees …
  - Unsupervised nets

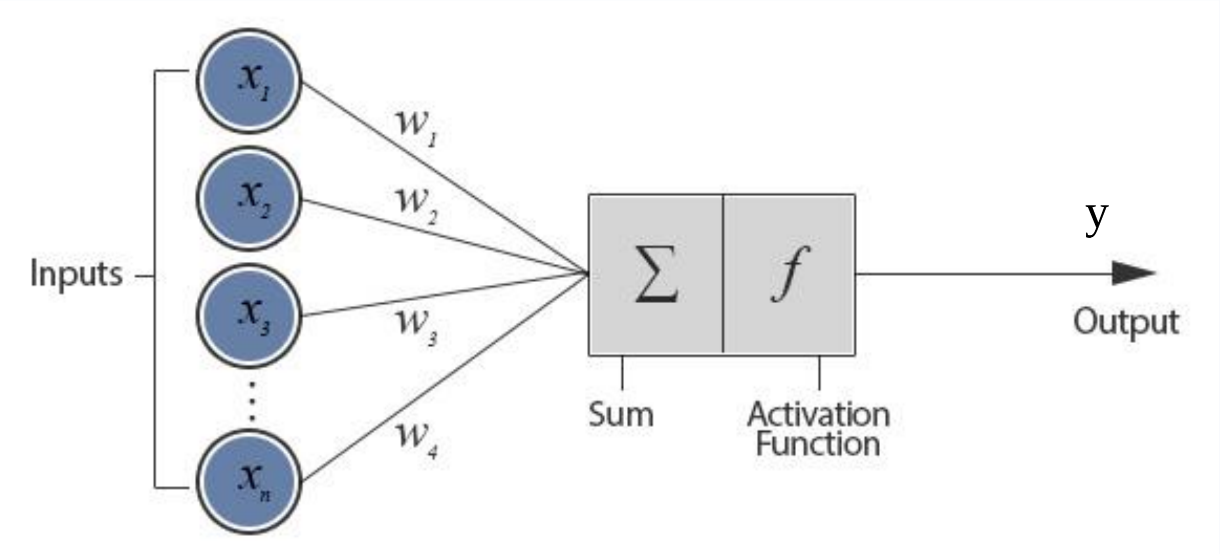




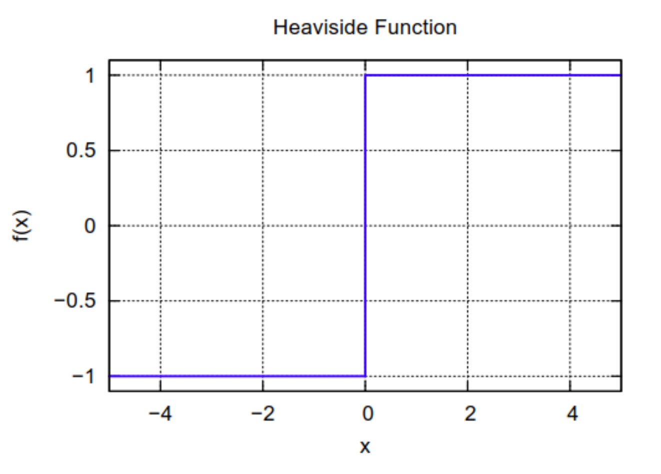


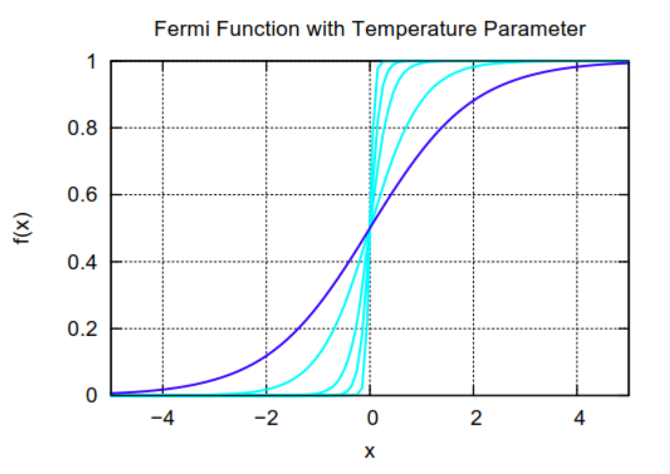
$$f : \mathbb{R}^8 \to \mathbb{B}^1$$

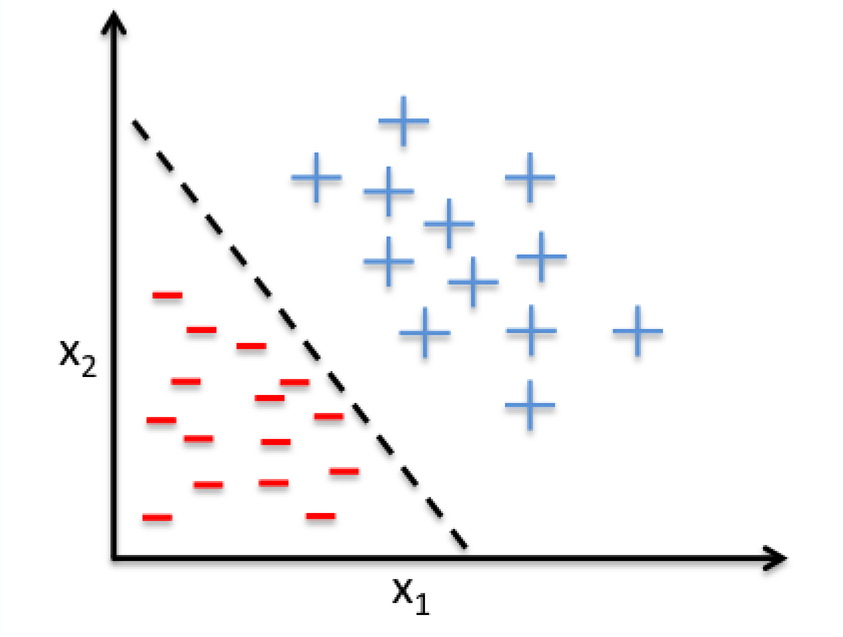
Perceptron: the artificial neuron



$$y = f\left(\sum_i w_i x_i\right)$$







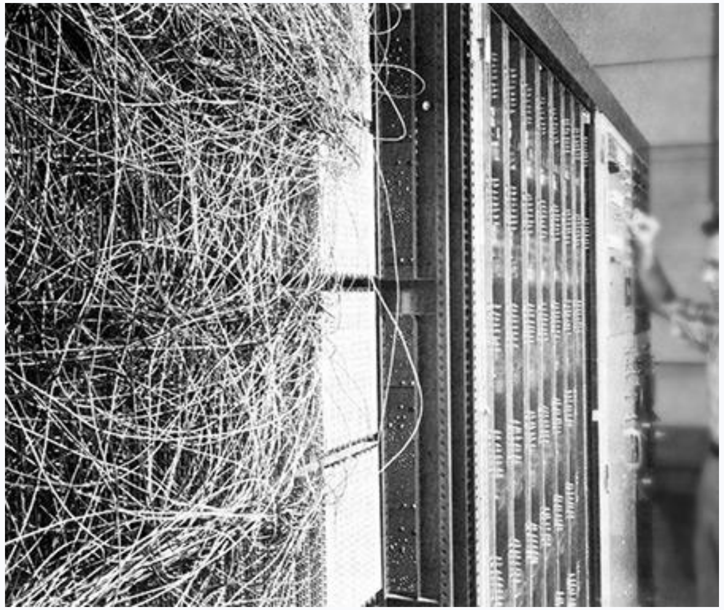
Learning is to find the optimal set of “synaptic weights”

Brief (and gappy) history

- The begining
  - 1943 McCulloch & Pitts: threshold based neurological switches
  - 1949 Hebb: Hebb-rule connections strengthen between simuletnously active neurons
- Golden age
  - 1951 Minsky: adjustable weights
  - 1957 Rosenblatt & Wightman: perceptron
  - 1960 Widrow & Hoff: ADA-LINE, delta-rule (adaptive) – echo filtering in telephones
  - 1969 Minsky & Papert: problems with perceptron

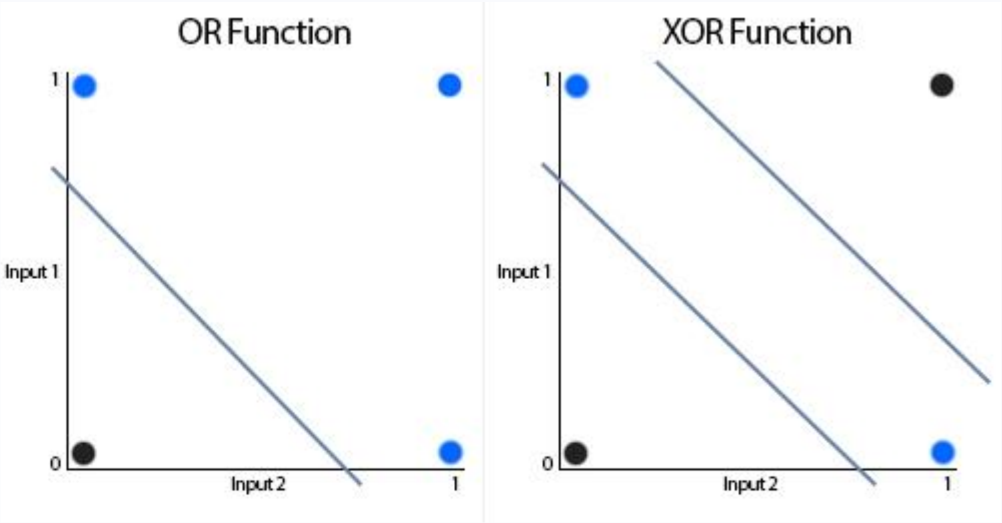
Figure 1.4: Some institutions of the field of neural networks. From left to right: John von Neumann, Donald O. Hebb, Marvin Minsky, Bernard Widrow, Seymour Papert, Teuvo Kohonen, John Hopfield, "in the order of appearance" as far as possible.

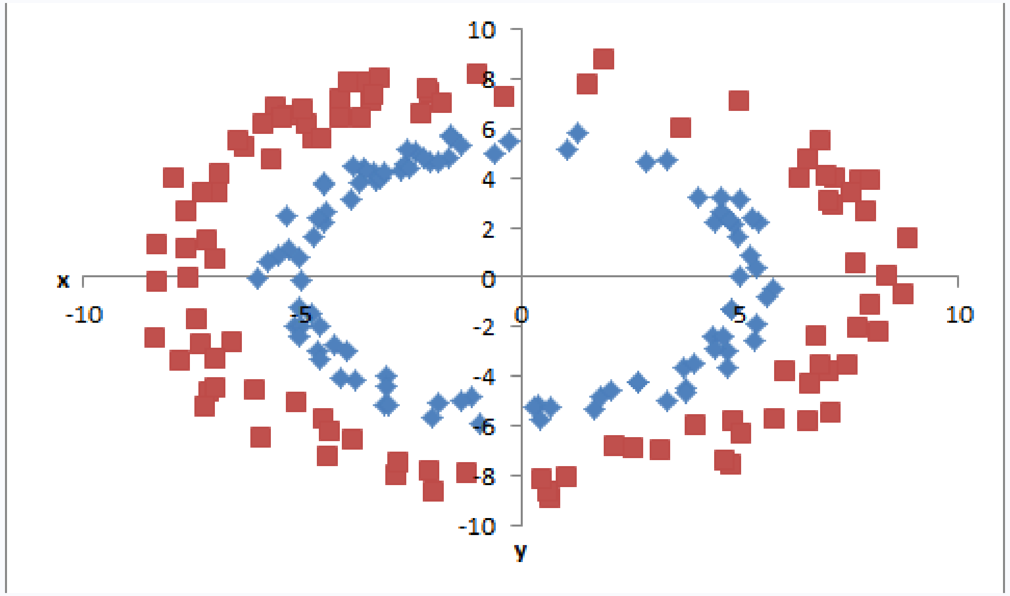
Book: John von Neumann: The computer and the brain, Yale University Press, 1959

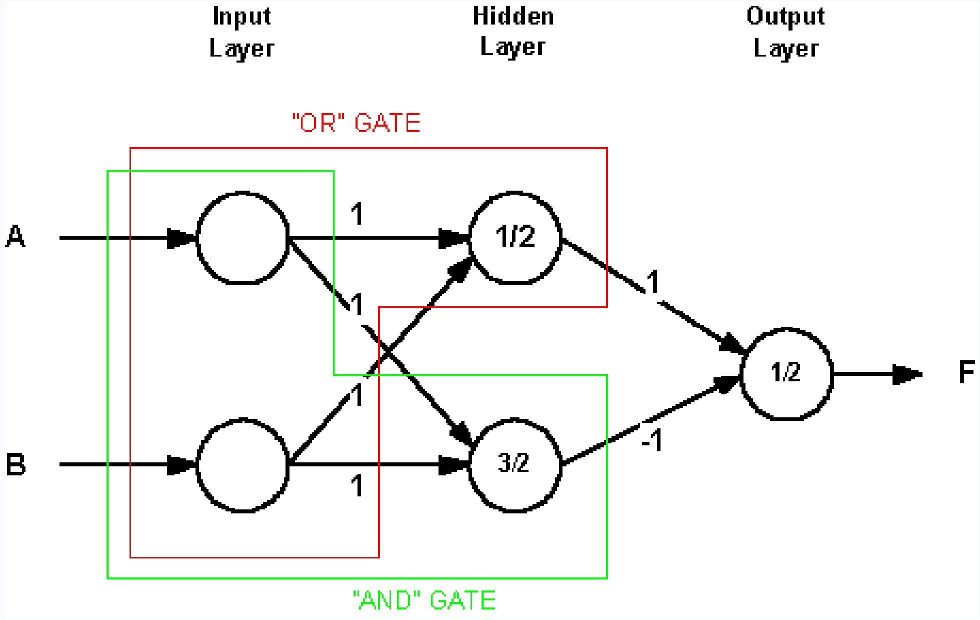


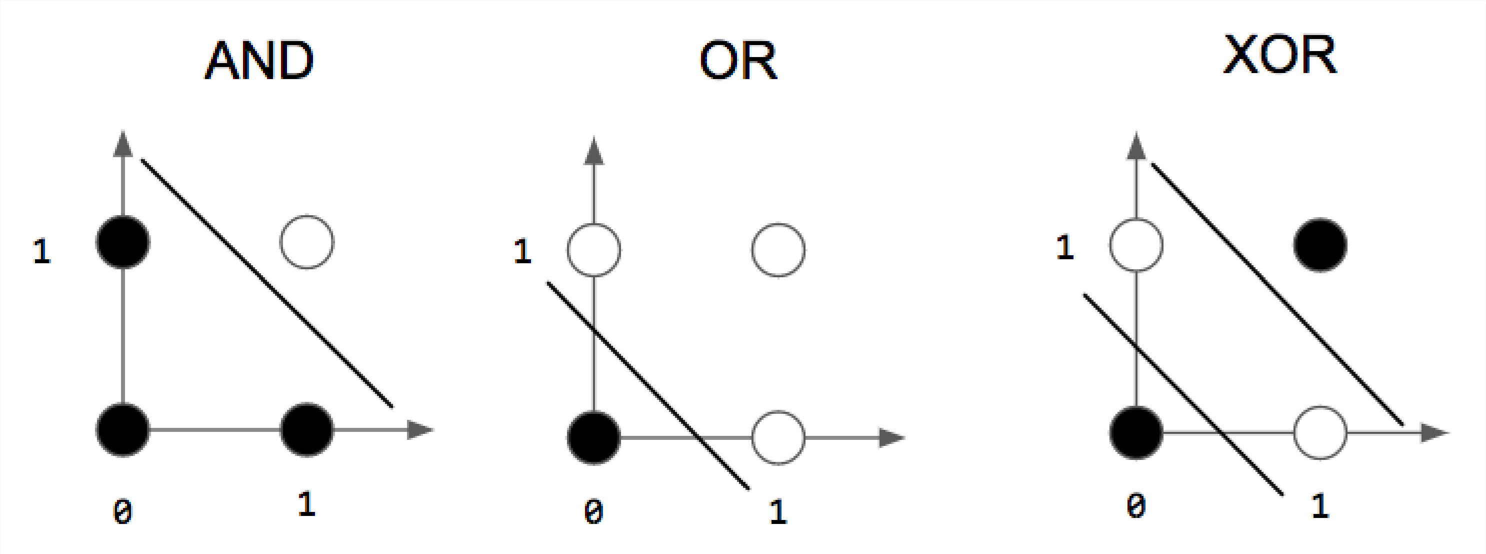
Winter is coming …

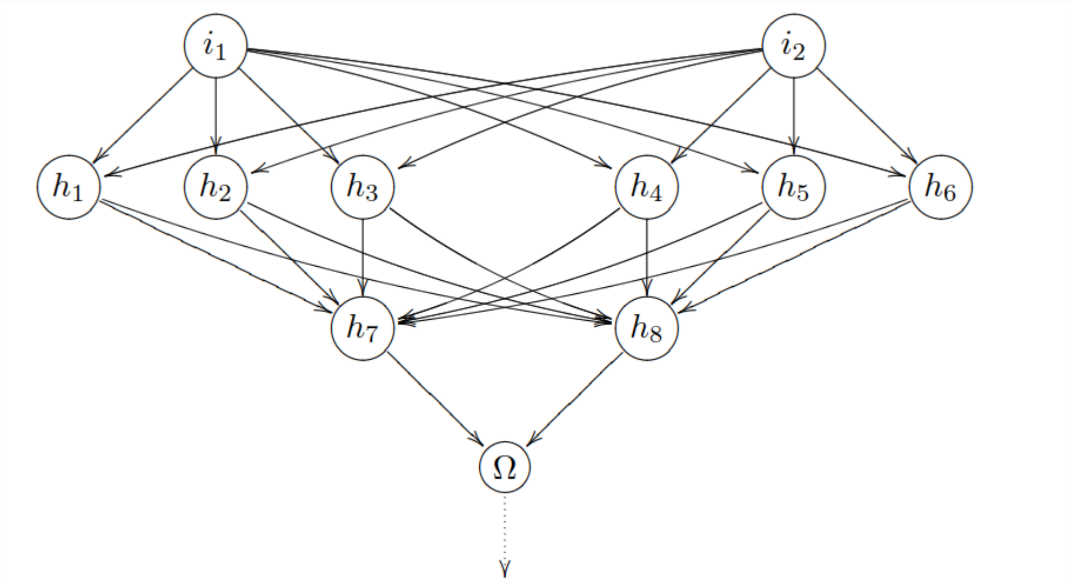
- The XOR problem
- Not all classification problems are linear, especially the interesting ones
- Minsky & Papert 1969 -> the "AI winter"
- Other approaches: expert systems
- Some "connectionist Eskimos"
    - 1976 Grossberg & Carpenter: Adaptive Resonance Theory (ART)
    - 1982 Kohonen: Self-organizing feature maps (unsupervised)

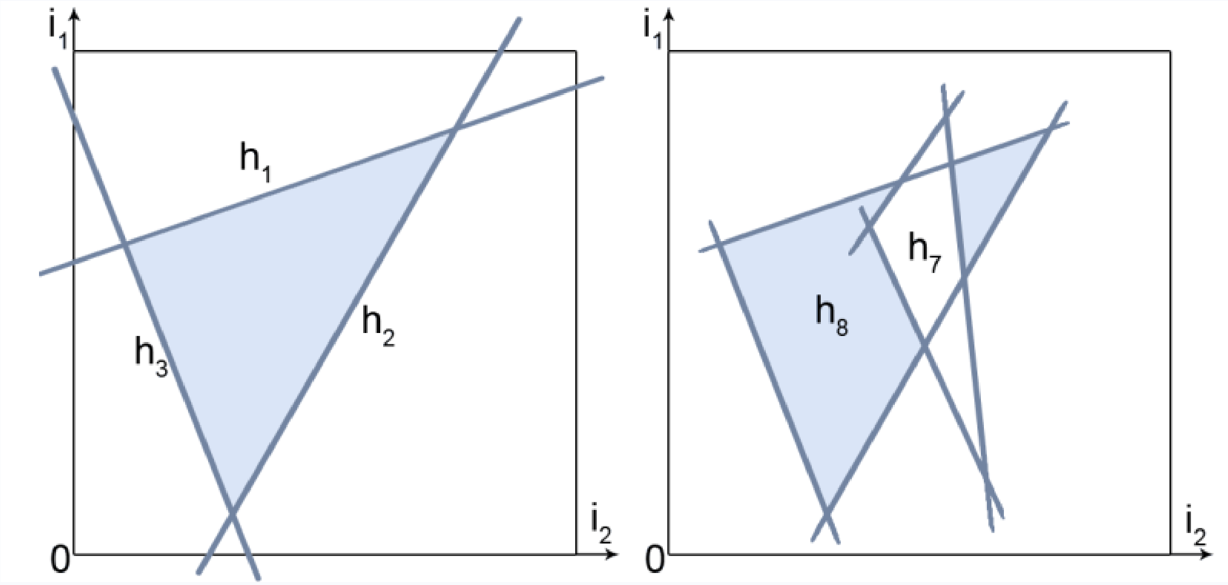




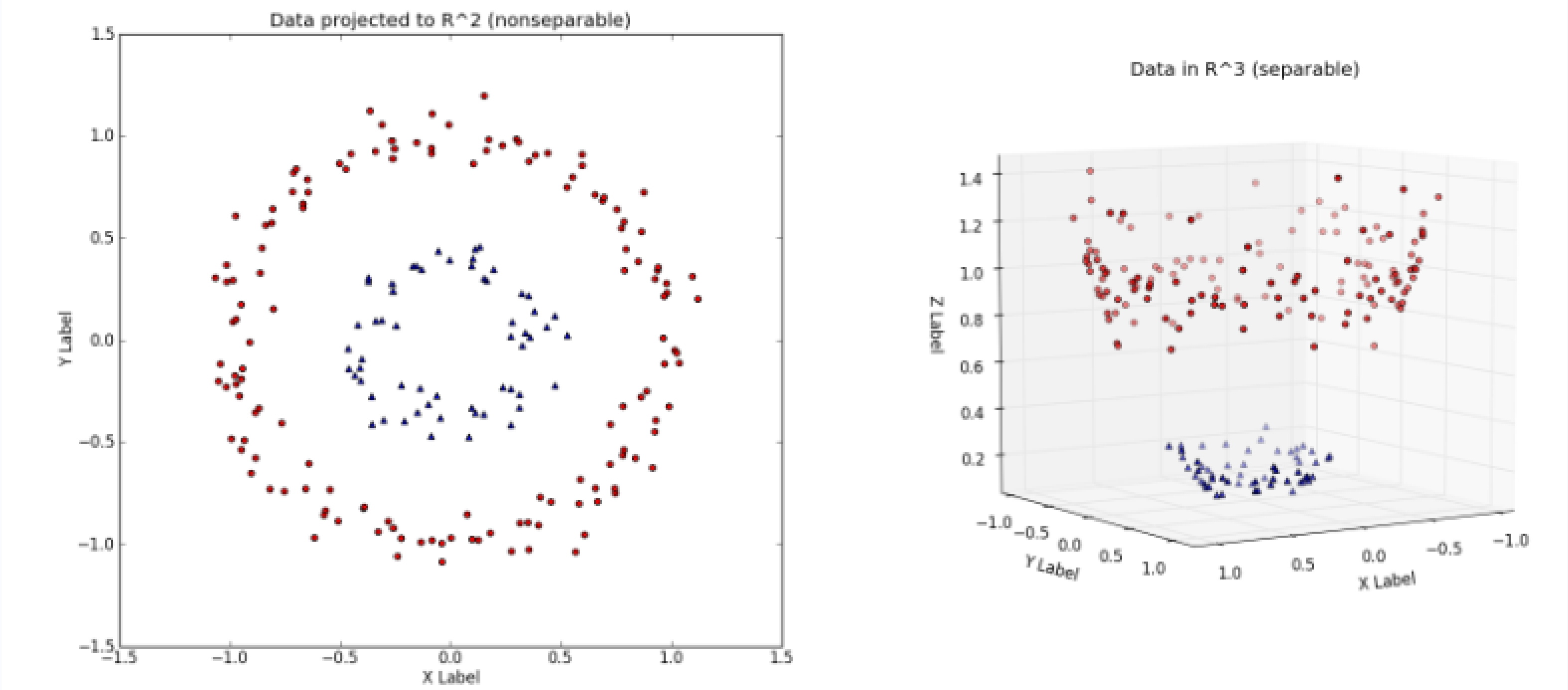

Input
Layer
Hidden
Layer
Output
Layer
"OR" GATE
A
B
1
1
1/2
1
-1
1
1/2
F
1
3/2
-1
1
"AND" GATE

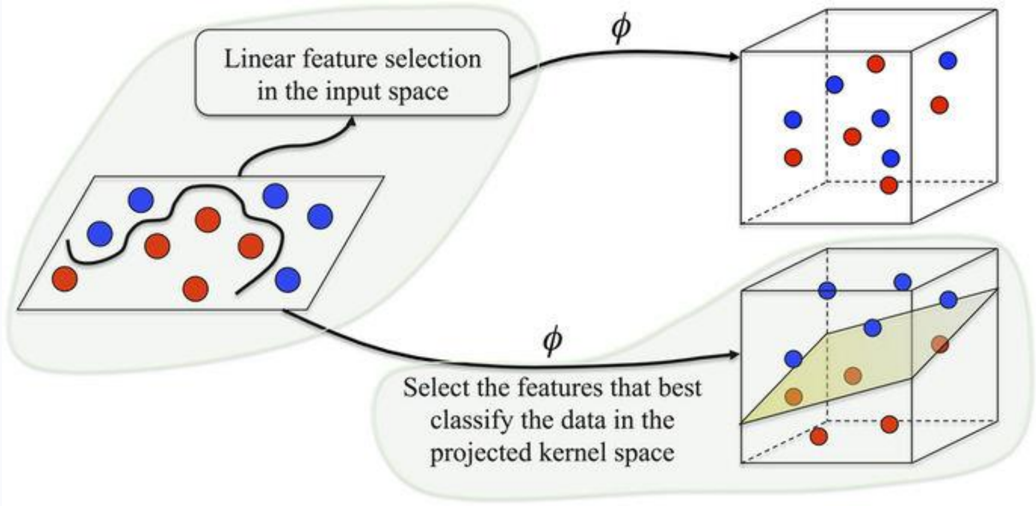

AND
OR
XOR
1
0
1
1
0
1
1
0
1

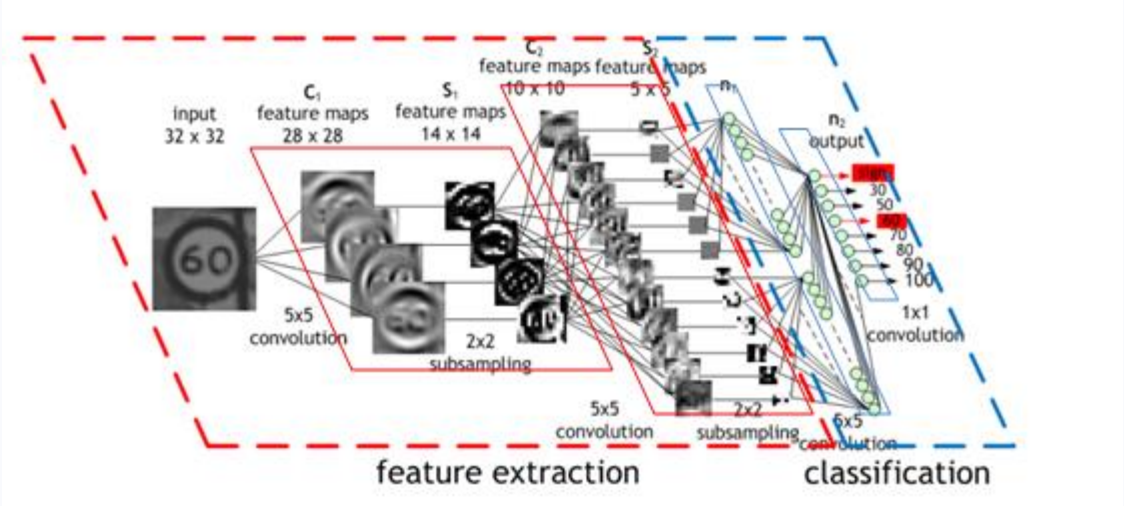

i_1
i_2
h_1
h_2
h_3
h_4
h_5
h_6
h_7
h_8
Ω


i_1
h_1
h_2
h_3
0
i_2
i_1
h_7
h_8
0
i_2

- Transforming the input representation, adding new dimensions may help





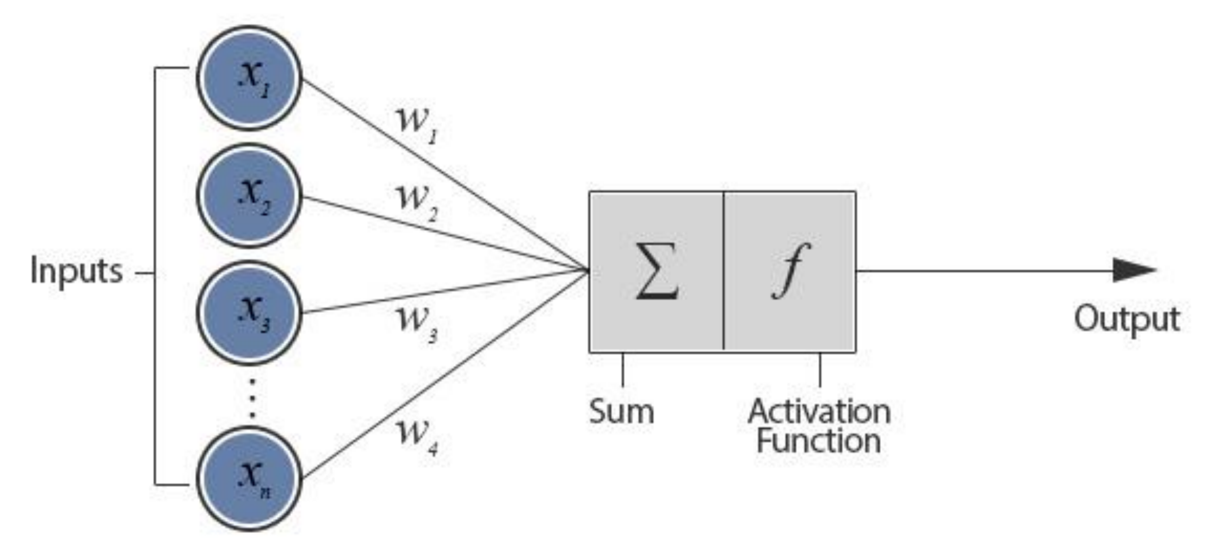


Brief (and gappy) history

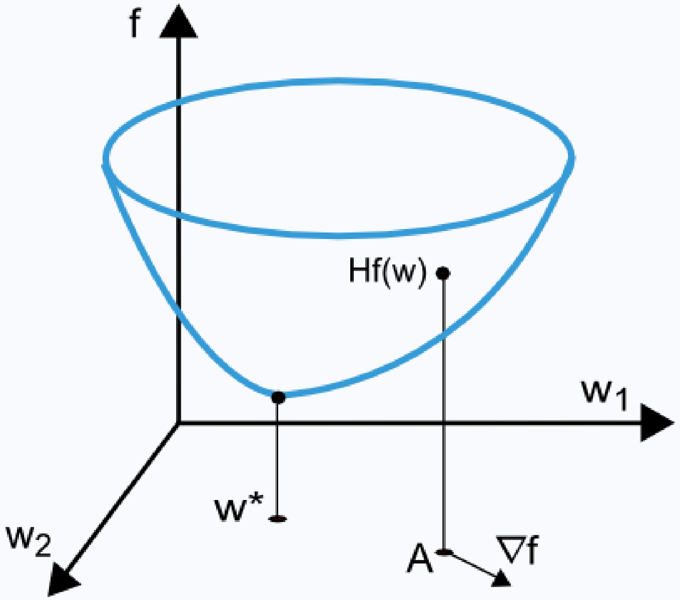
- The renaissance
  - 1982 Hopfield: Hopfield networks, associative memory "spin glass", 1985 solving TSP
  - 1986 Parallel Distributed Processing group (Rumelhart, Hinton, Williams): backpropagation of errors
- Mostly the same architecture, with various improvements : deep learning

Figure 1.4: Some institutions of the field of neural networks. From left to right: John von Neumann, Donald O. Hebb, Marvin Minsky, Bernard Widrow, Seymour Papert, Teuvo Kohonen, John Hopfield, "in the order of appearance" as far as possible.

Learning as energy minimization



$$y = f\left(\sum_i w_i x_i\right)$$

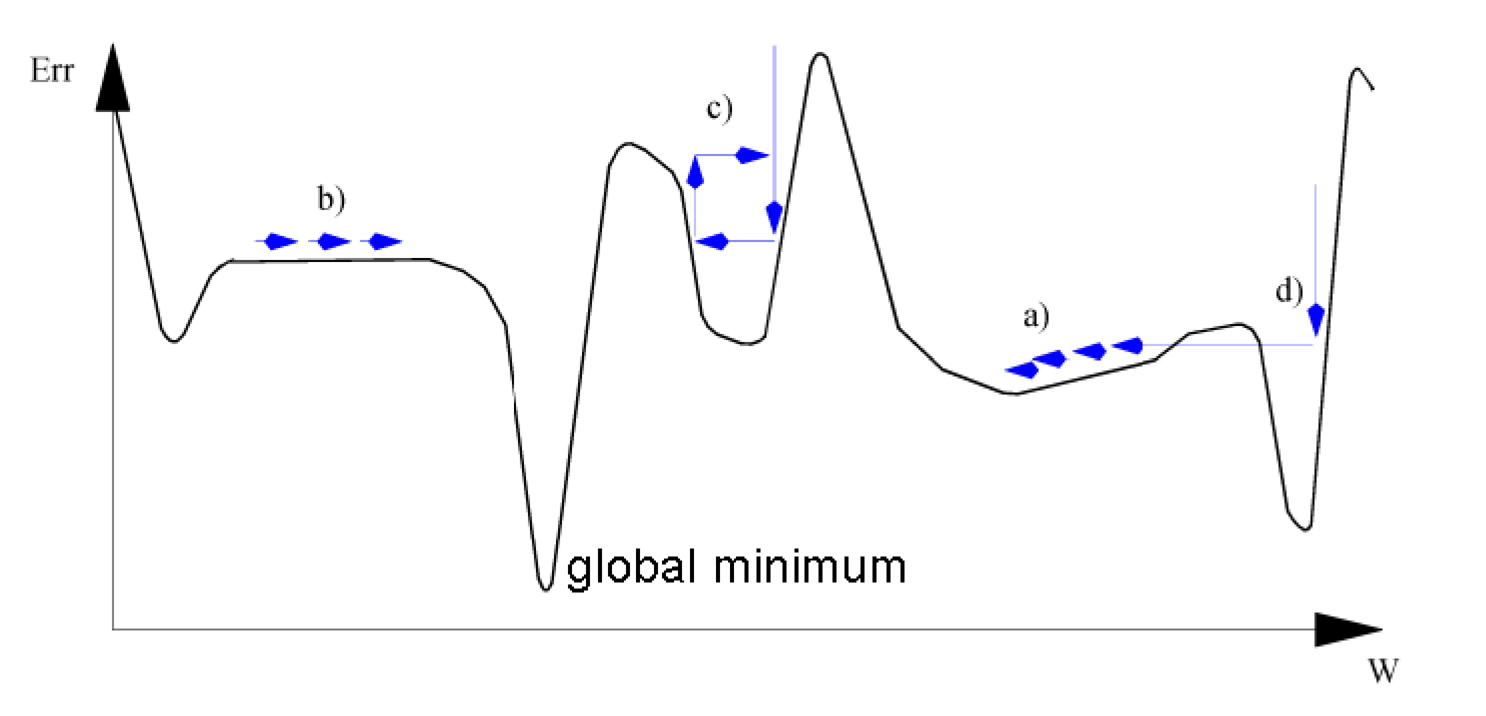


$$E(w_{ij}) = \sum_i \left(f_i(w_{ij}, x_j) - y_i\right)^2$$

$$\mathrm{argmin}_w \sum_i \left(f_i(w_{ij}, x_j) - y_i\right)^2$$

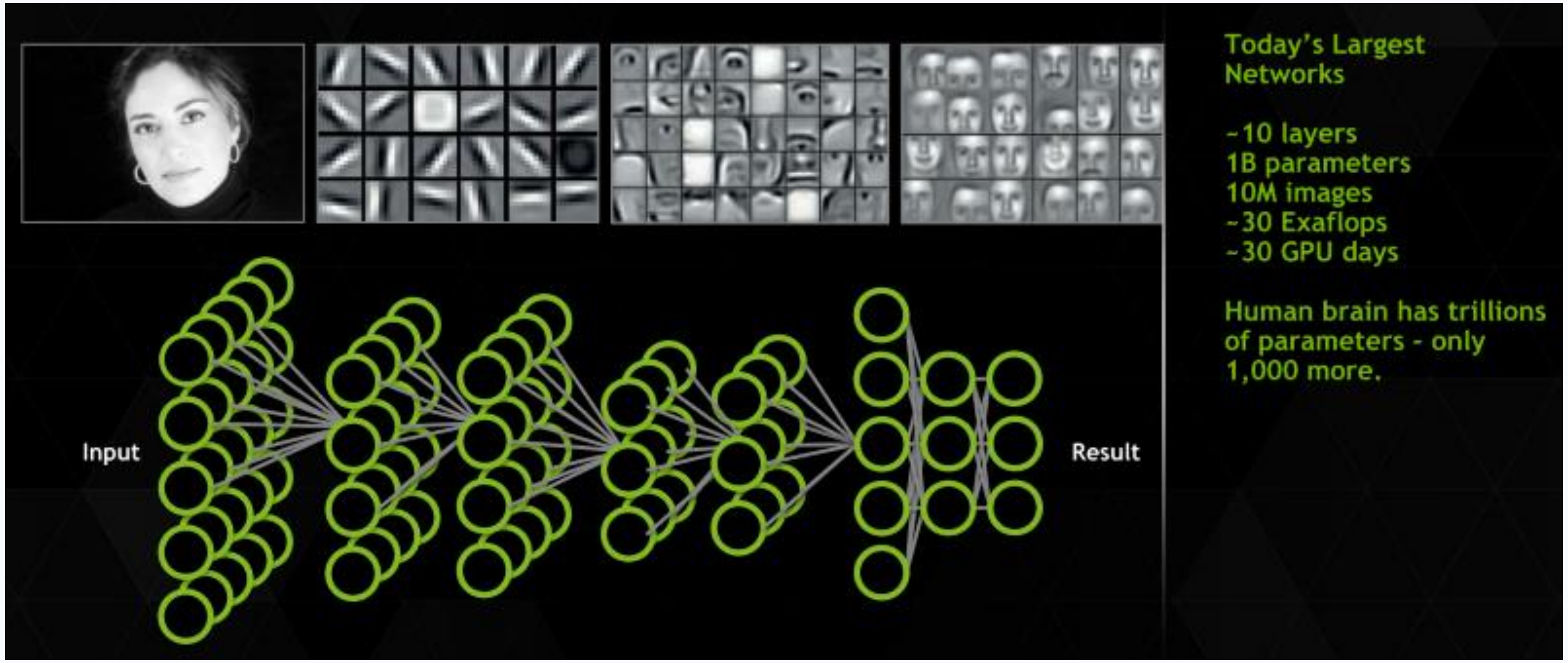
Complex nonlinear optimization problem

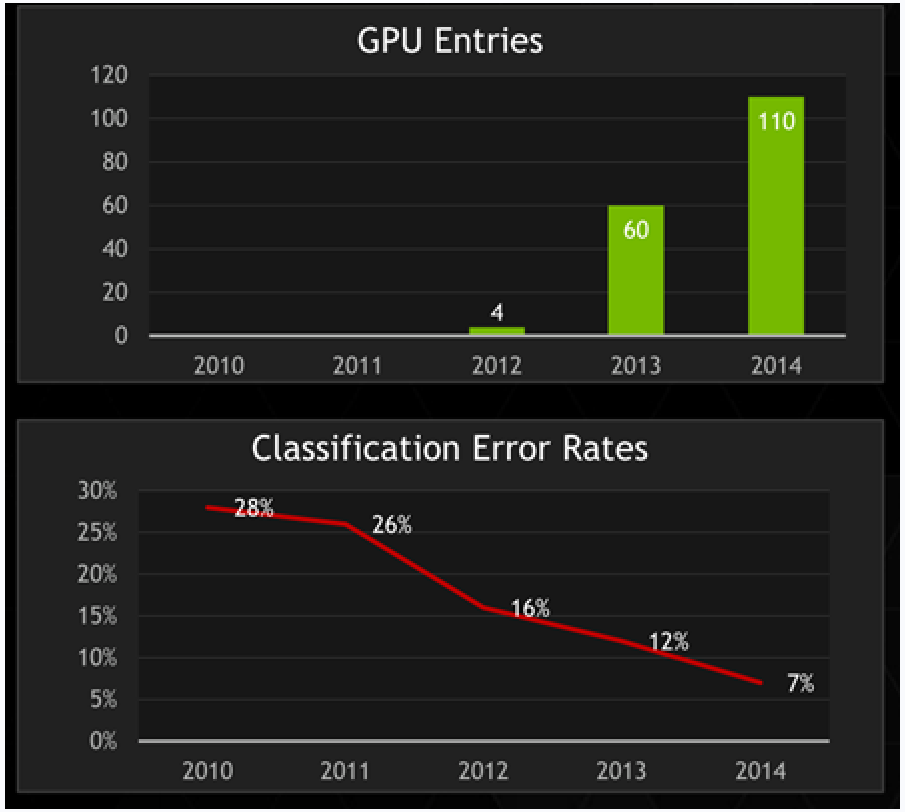
- Concepts from statistical physics
- Boltzmann machines
- Simulated annealing



Deep learning – why now

- Moore's law
  - Computational power: GPUs, TPU, millions of synaptic weights
  - Training sets: millions of images/documents/videos
- Some algorithmic developments (faster processors -> faster development)





Deep learning in sciences –projects at the department

- Astronomy
  - Photometric redshift estimation from images
  - Inverse problem: physical parameters of galaxies from spectral features
- Medical sciences
  - Detecting cancer in radiology images / CT
  - Analysis pathology images
- Quantum chemistry
  - Predicting (bio)chemical properties from wave function
- Genetics
  - Detecting mutations in DNA sequencing
  - Predicting immune response (MHC binding)
  - Predicting antibiotic resistance
  - Predicting transcription factor binding
- Time series analysis (smart watch traces), social networks, …

Skills

deeplea17em	Deep learning és gépi tanulás a tudományokban	2	S/O	2017/18/2	1	Előadás/ Normál	48/0/50	K:16:00-18:00

- Linear algebra
- Statistics
- Programming
- IT tools, large code, large systems, collaboration

- Not traditional lecture
- Interactive, project based class with homeworks
- Online materials

http://qati.me/dl-class.html

- Dezső Ribli, Attila Bagoly, Bálint Pataki, István Csabai

deeplea17m