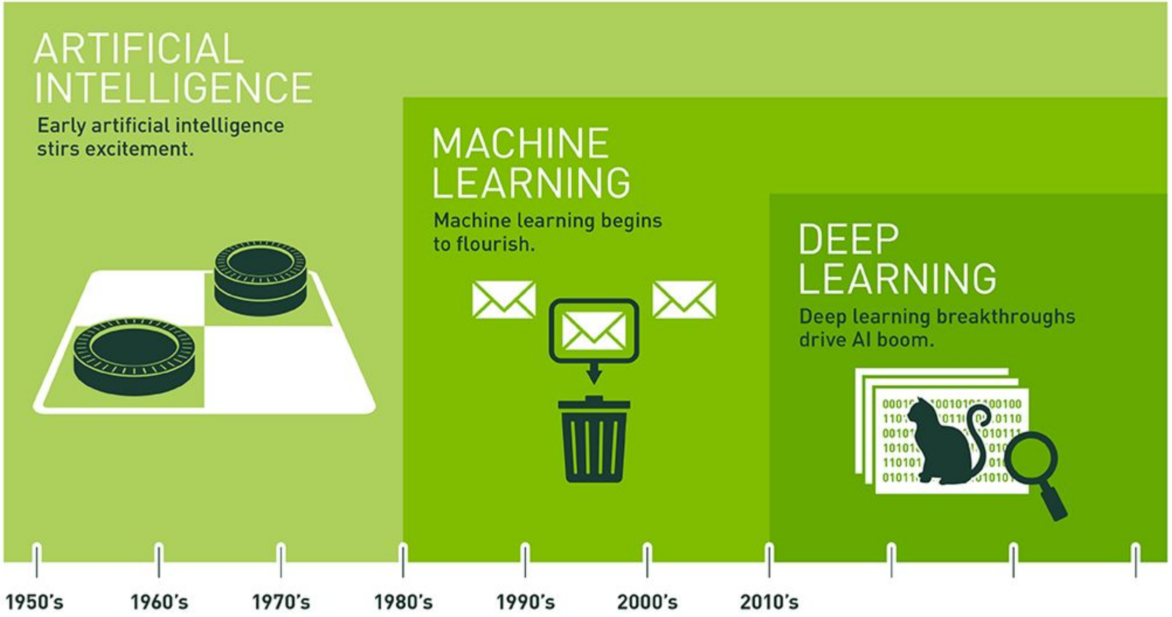
Application of networks in artificial intelligence and deep learning -- essence and tricks

Peter Csermely's course on networks - Zsolt Vassy

“AI is like teenage sex: everyone talks about it, nobody knows how to do it, everyone thinks everyone else is doing it & so claims to do it”

/ Dan Ariely prof. of Psychology and Behavioral Economics at Duke University/

Artificial Intelligence, Machine Learning, Deep Learning



Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

What is Artificial Intelligence?

Ideally, it is a human behavioral system and / or a humanly acting system.

Currently, we can talk about rationally thinking systems, which mostly means machine learning: generating knowledge from experience.

Simplified: **Problem solving in changing circumstances.**

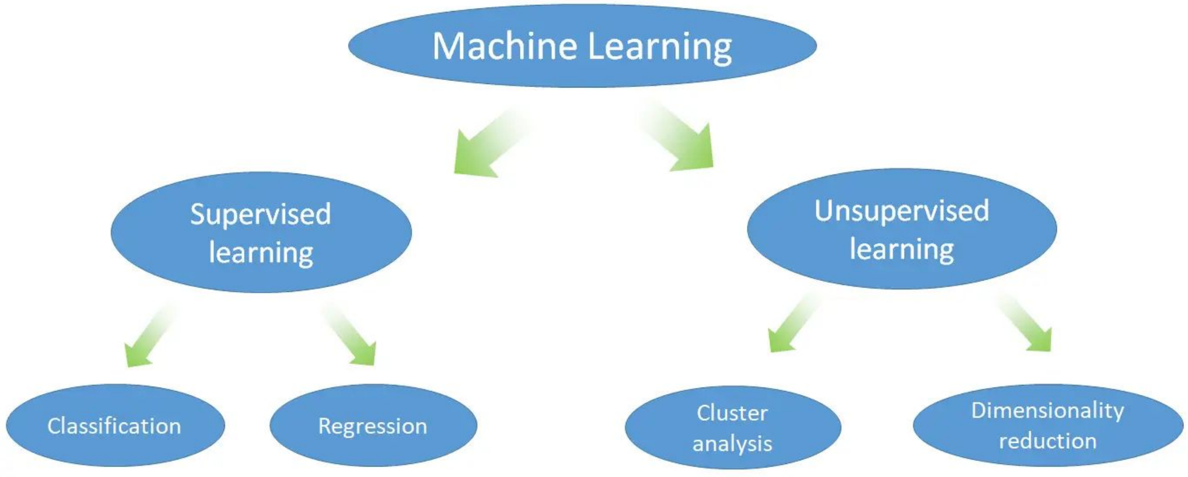
Use of Artificial Intelligence

Algorithms that can learn from and make predictions on data.

If the algorithm system optimizes its parameters such that its performance improves, then it is considered to be learning that task.

There are two types of machine

- **Supervised**
- **Non-Supervised**



Common learning algorithms

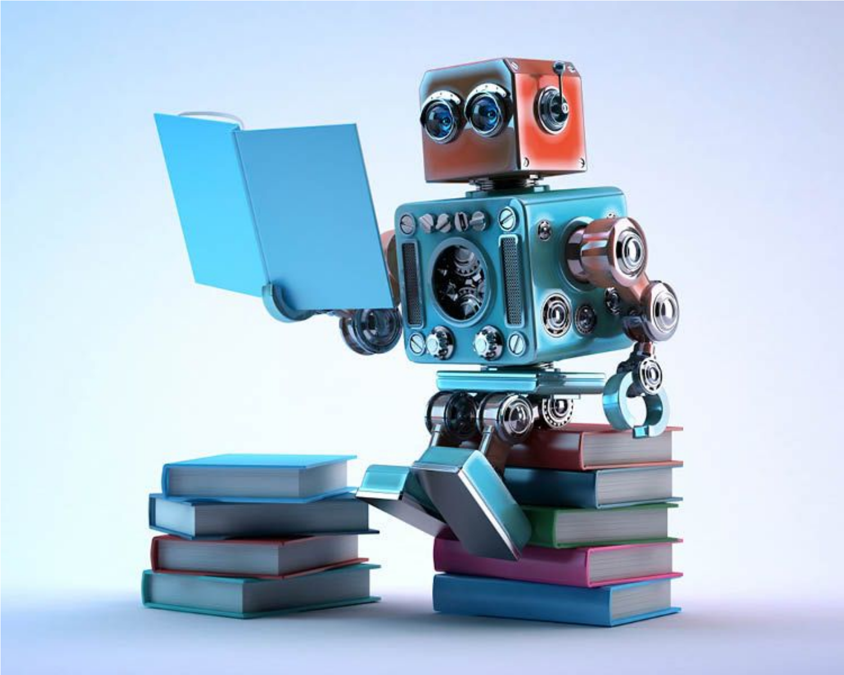
Support vector machine (SVM)

Nearest Neighbors

Decision tree és random forest - **Contains network**

Artificial neural networks - **Contains network**

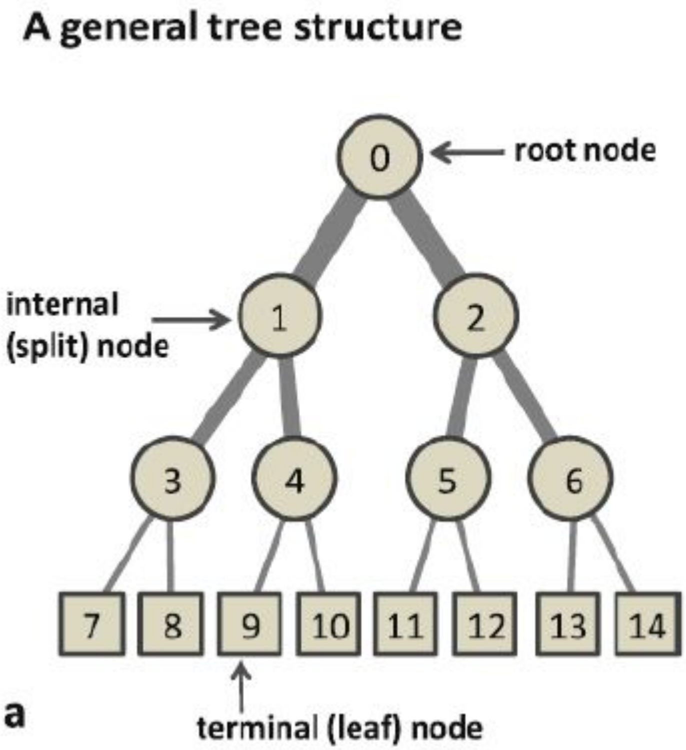
Deep neural network / Deep learning - **Contains network**



Decision trees

A tree is a set of nodes and edges organized in a hierarchical fashion.

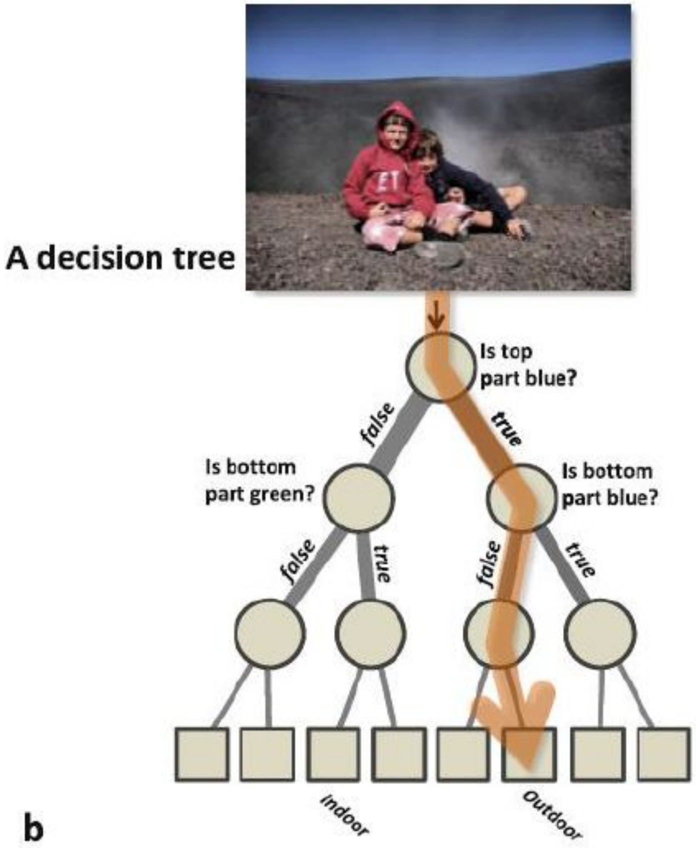
A tree is a graph with no loops.



Decision trees

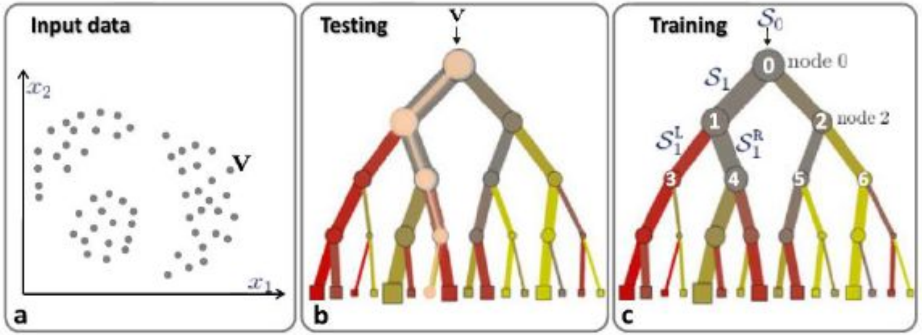
A decision tree is a tree where each internal node stores a split (or test) function to be applied to the incoming data.

Each leaf stores the final answer (predictor). Here we show an illustrative decision tree used to figure out whether a photo represents an indoor or outdoor scene.

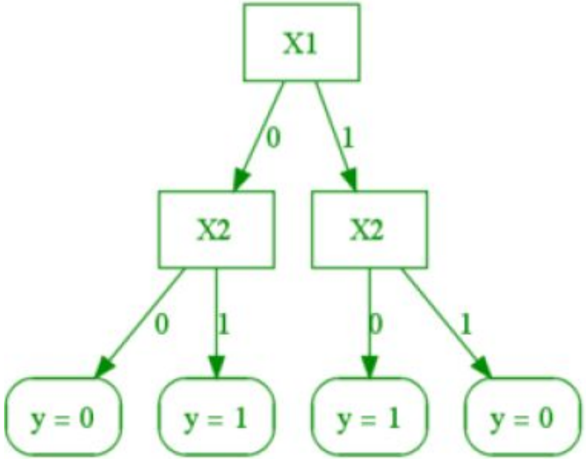


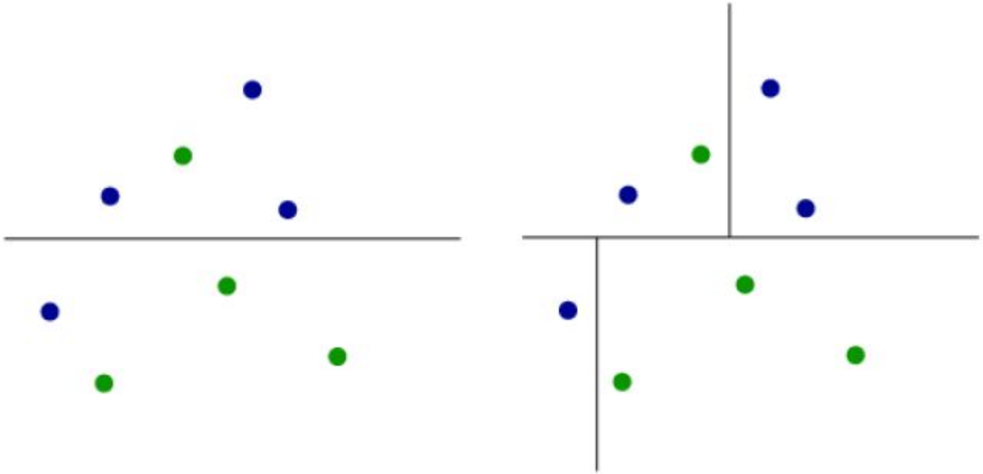
Decision trees

Given input data represented as a collection of points in the d-dimensional space defined by their feature responses (2D in this example), a decision tree is trained by sending the entire training set into the tree and optimizing the parameters of the split nodes so as to optimize a chosen energy function.



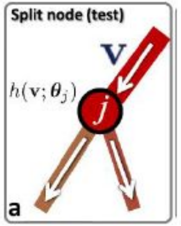
Decision trees - Training example





Binary split function

A test function at a split node j that gives binary outputs. The data point v arriving at the split node is sent to its left or right child node according to the result of the test function.



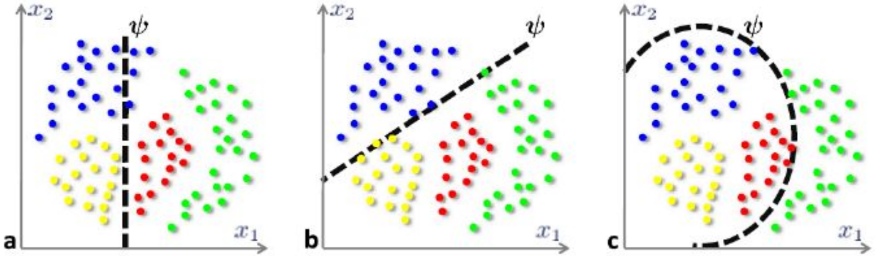
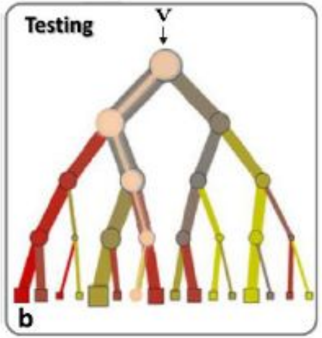


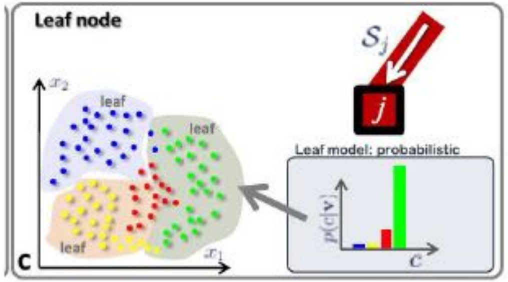
Fig. 3.4 Example weak learners. In this illustration the colors attached to each data point (*circles*) indicate different classes. (**a**) Axis-aligned hyperplane weak learner. (**b**) General oriented hyperplane. (**c**) Quadratic surface (conic in 2D). For ease of visualization here we have $\mathbf{v} = (x_1\ x_2)^\top \in \mathbb{R}^2$ and $\boldsymbol{\phi}(\mathbf{v}) = (x_1\ x_2\ 1)^\top$ in homogeneous coordinates. In general, a data point $\mathbf{v}$ may have a much higher dimensionality and $\boldsymbol{\phi}(\mathbf{v})$ still a dimensionality ≤ 2

Decision trees - Training

In the supervised case, after training, each leaf node remains associated with a subset of (labeled) training data.

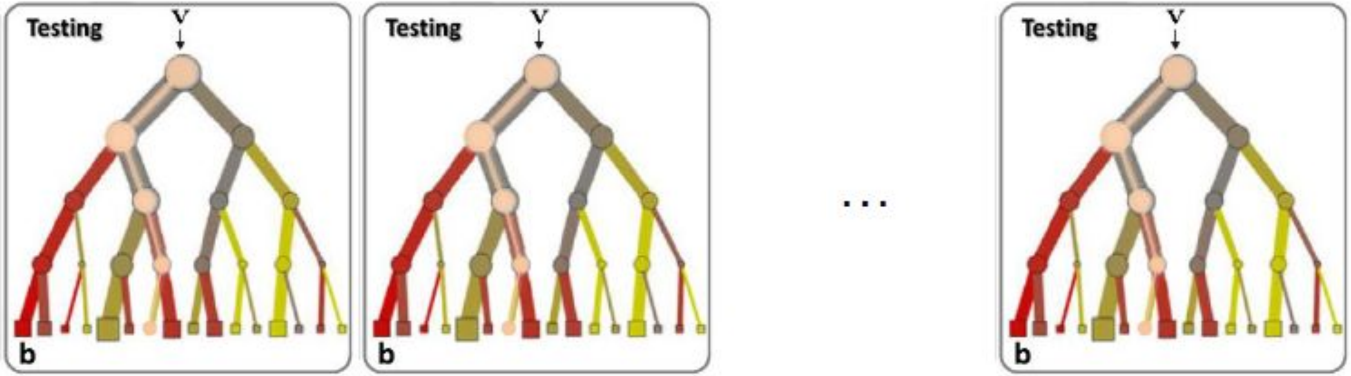
During testing, a previously unseen point traverses the tree until it reaches a leaf.





Random forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks, that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees

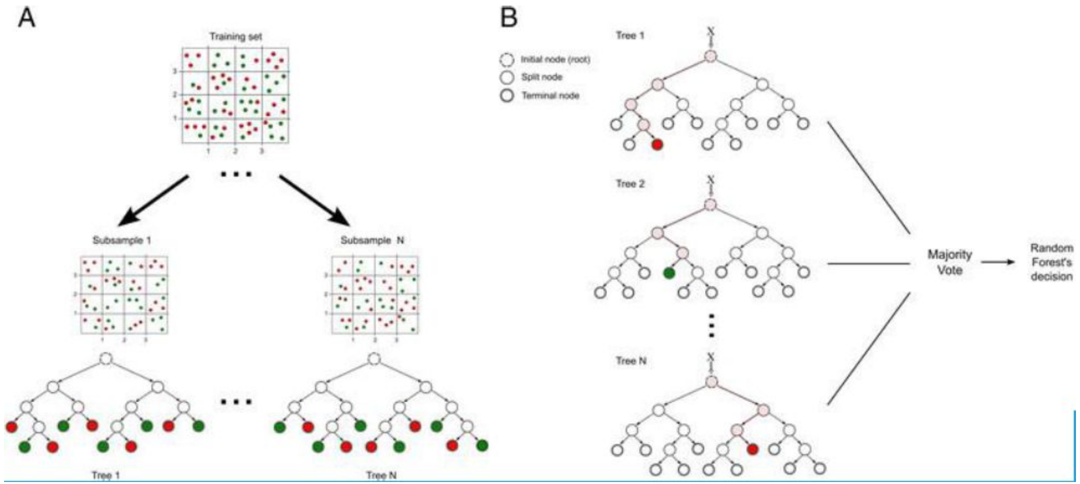


Randomness

Randomness is injected into the trees during the training phase.

Two of the most popular ways of doing so are:

- Random sampling (sample bagging)
- Randomized node optimization (feature bagging)

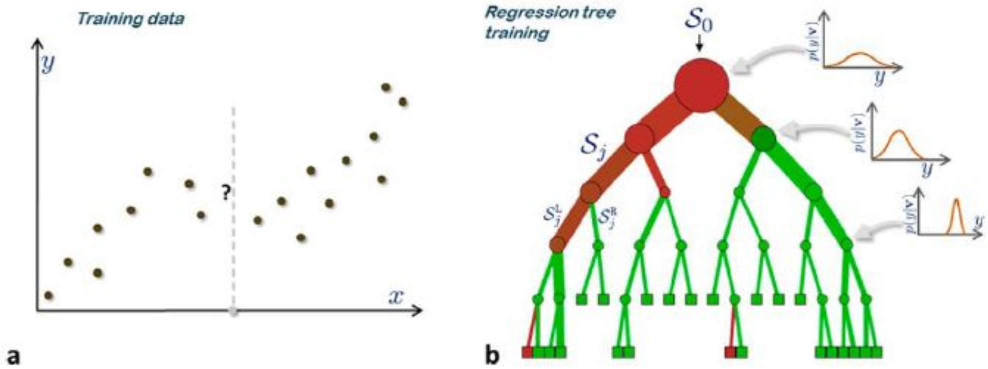


Regression Forest

Regression: training data and tree training.

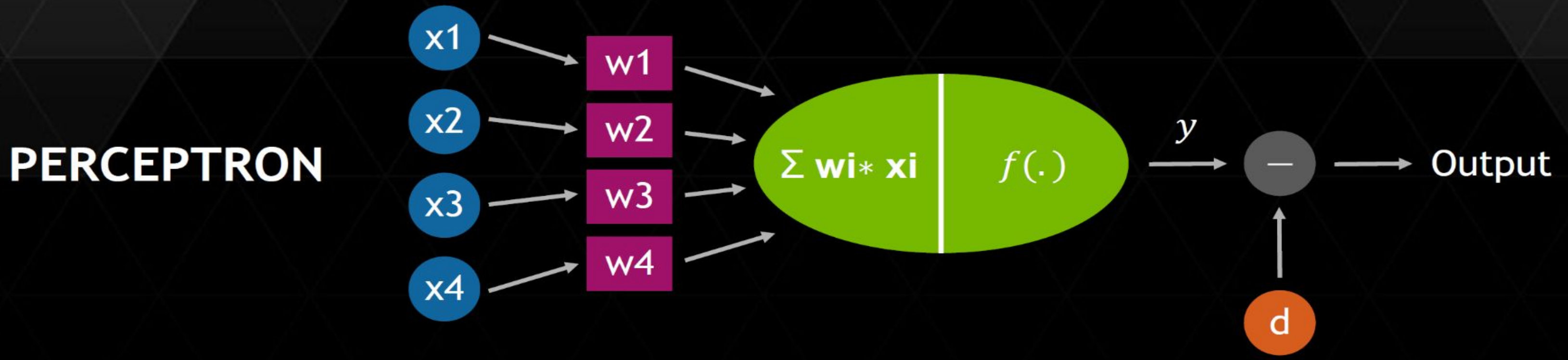
(a)The input feature space is one-dimensional in this example. The associated continuous ground truth label is denoted by their position along the direction of the y axis.

(b) During training a set of labeled training points are used to optimize the parameters of the tree. In a regression tree the entropy of the continuous densities associated with different nodes decreases (their confidence increases) when going from the root towards the leaves.

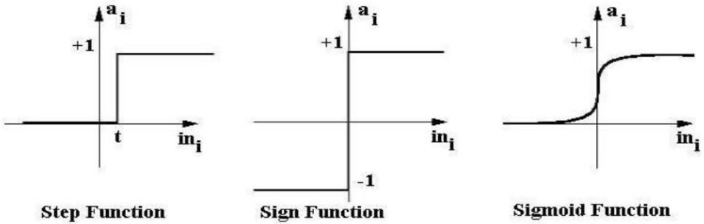


Artificial neural networks - an idea taken from life

THE PERCEPTRON — THE SIMPLEST MODEL



ACTIVATION FUNCTIONS:



LEARNING:

$$y^{(t)} = f\left\{\sum_i w_i^{(t)} x_i^{(t)}\right\}$$

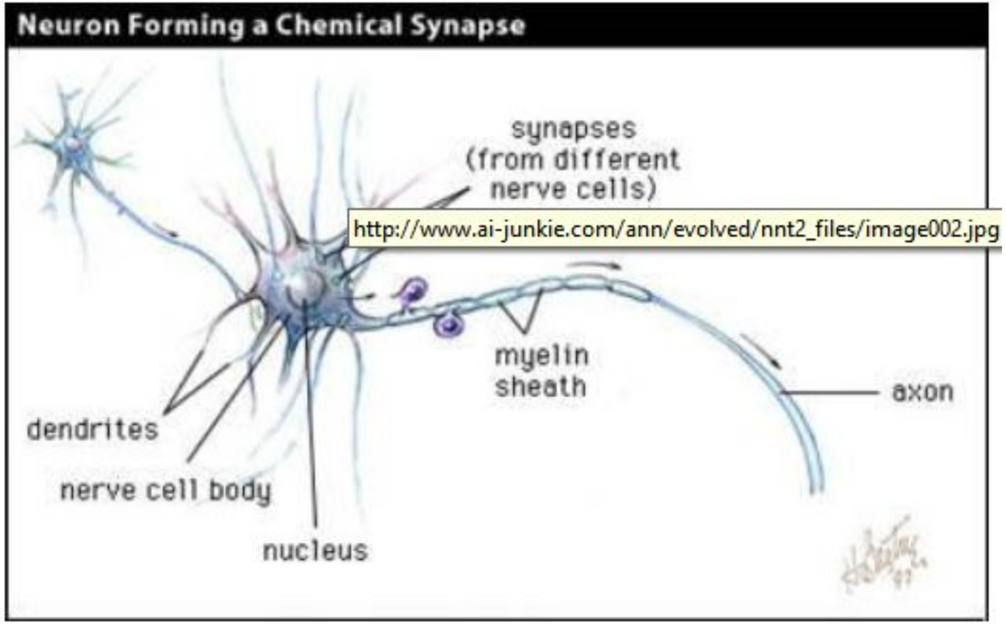
Update

$$\Delta w_i^{(t)} = \varepsilon\left(d^{(t)} - y^{(t)}\right) x_i^{(t)}$$

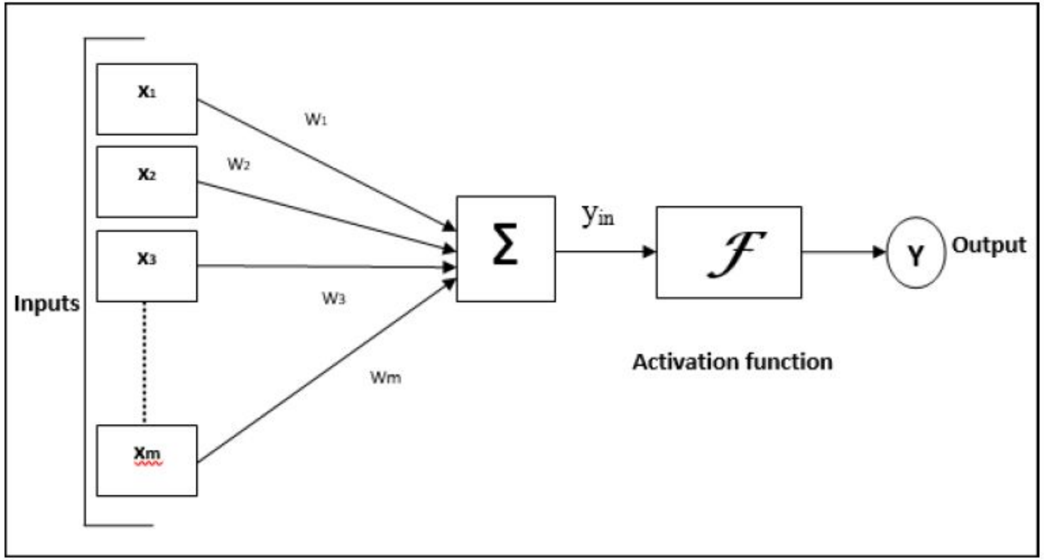
$$w_i^{(t+1)} = w_i^{(t)} + \Delta w_i^{(t)}$$

Slide Credit: Geoff Hinton

Real neuron


Neuron Forming a Chemical Synapse
synapses
(from different
nerve cells)
http://www.ai-junkie.com/ann/evolved/nnt2_files/image002.jpg
myelin
sheath
axon
dendrites
nerve cell body
nucleus

Artificial neuron



$$y_{in} = x_1.w_1 + x_2.w_2 + x_3.w_3 \ldots x_m.w_m$$

i.e., Net input $y_{in} = \sum_i^m x_i.w_i$

The activation function

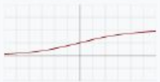
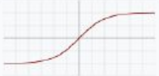
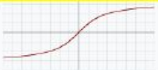
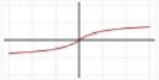
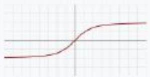
The activation function of a node defines the output of that node given an input or set of inputs.

In ANN, they are often nonlinear that allows ANN to compute non-trivial problems. They are often continuously differentiable (easier for training).

Commonly used activation functions include::

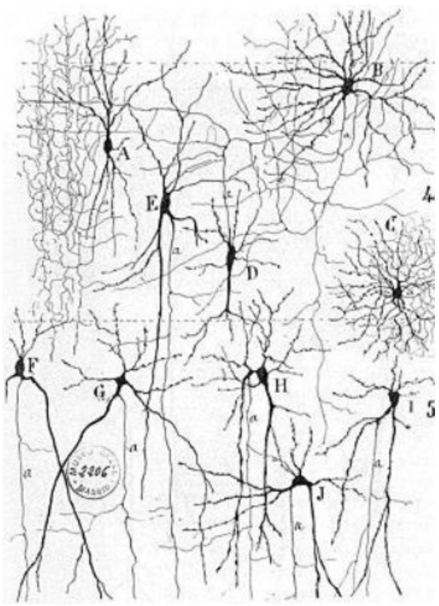
- Sigmoid
- Tanh

The activation function

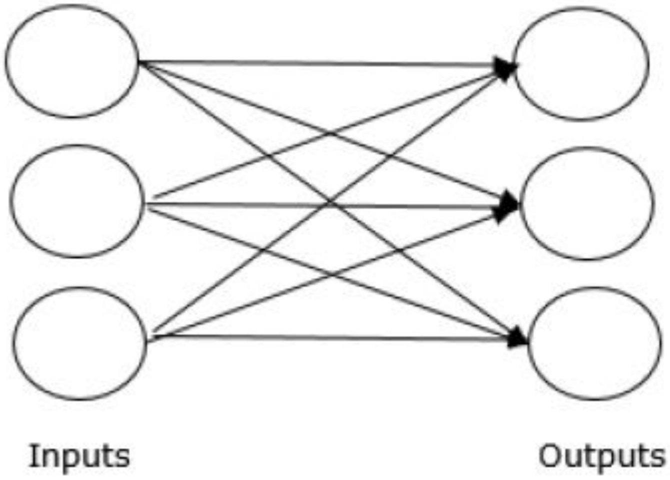
Name	Plot	Equation	Derivative (with respect to *x*)
Identity		$f(x) = x$	$f'(x) = 1$
Binary step		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x \neq 0 \\ ? & \text{for } x = 0 \end{cases}$
Logistic (a.k.a. Sigmoid or Soft step)		$f(x) = \sigma(x) = \frac{1}{1+e^{-x}}$ [1]	$f'(x) = f(x)(1 - f(x))$
TanH		$f(x) = \tanh(x) = \frac{(e^{x} - e^{-x})}{(e^{x} + e^{-x})}$	$f'(x) = 1 - f(x)^2$
ArcTan		$f(x) = \tan^{-1}(x)$	$f'(x) = \frac{1}{x^2+1}$
Softsign [7][8]		$f(x) = \frac{x}{1+\lvert x\rvert}$	$f'(x) = \frac{1}{(1+\lvert x\rvert)^2}$
Inverse square root unit (ISRU)[9]		$f(x) = \frac{x}{\sqrt{1+\alpha x^2}}$	$f'(x) = \left(\frac{1}{\sqrt{1+\alpha x^2}}\right)^3$
Rectified linear unit (ReLU)[10]		$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$	$f'(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$

Network architecture

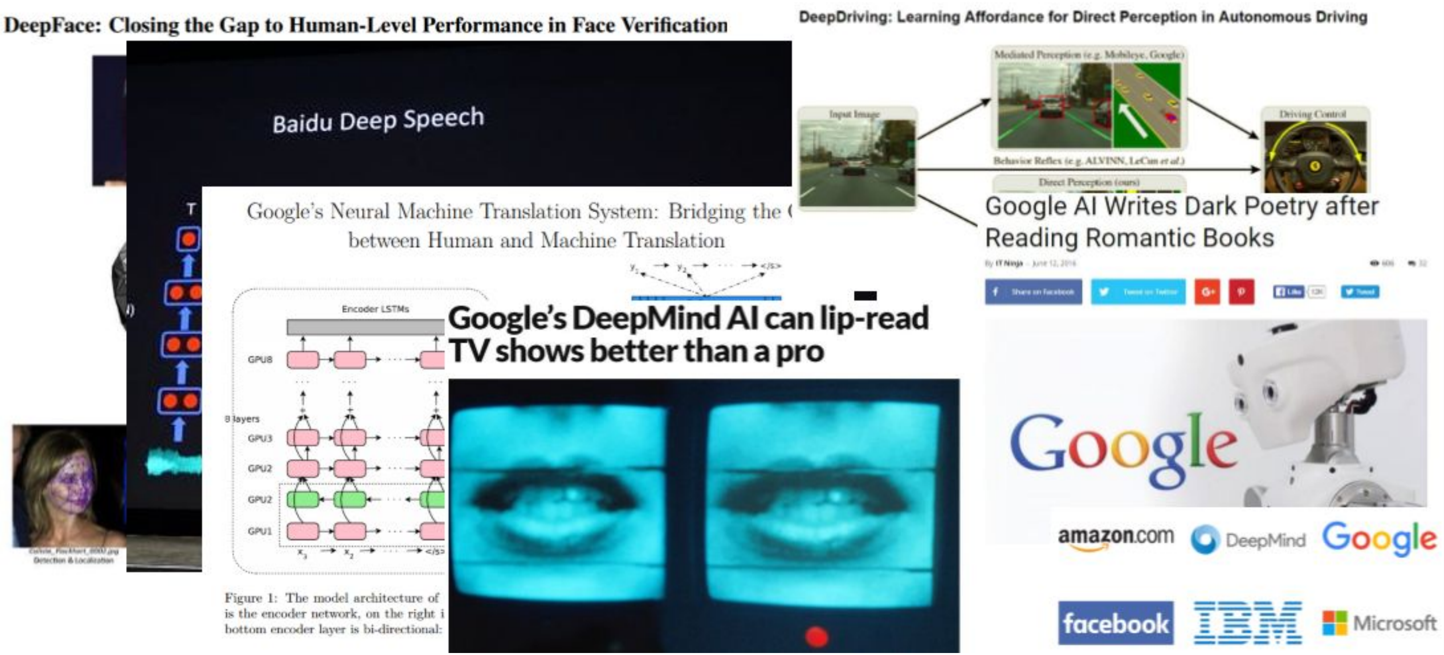
Biological neural network:



Artificial neural network:

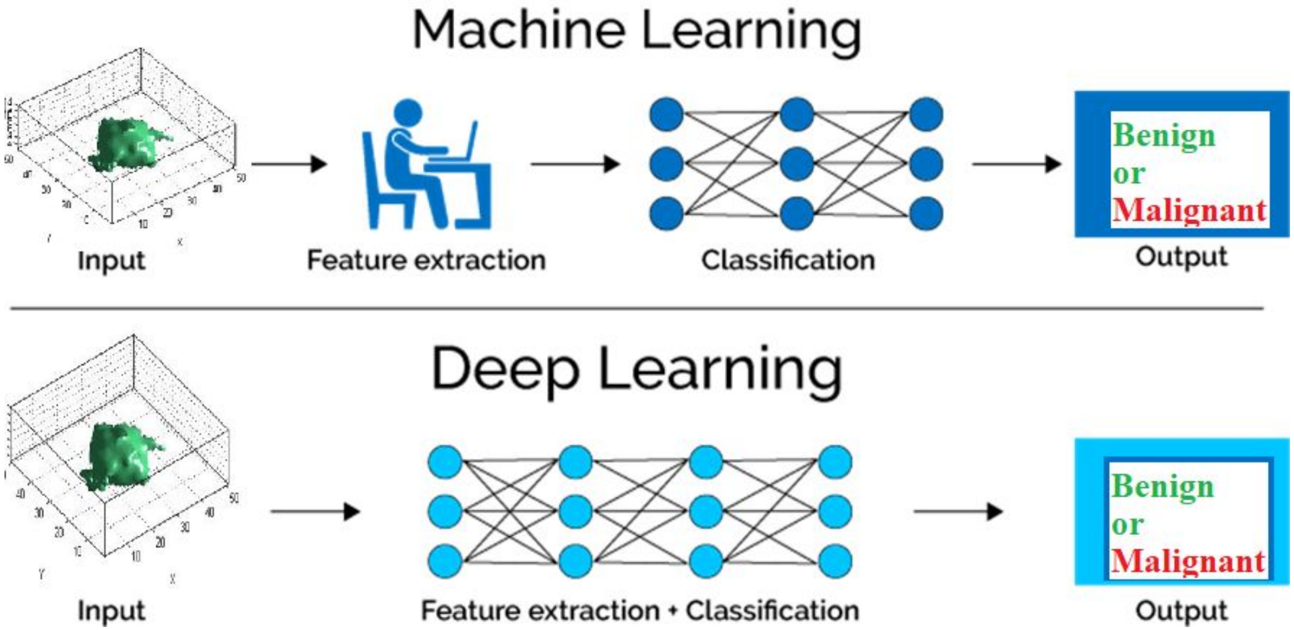


Artificial neural networks and deep learning

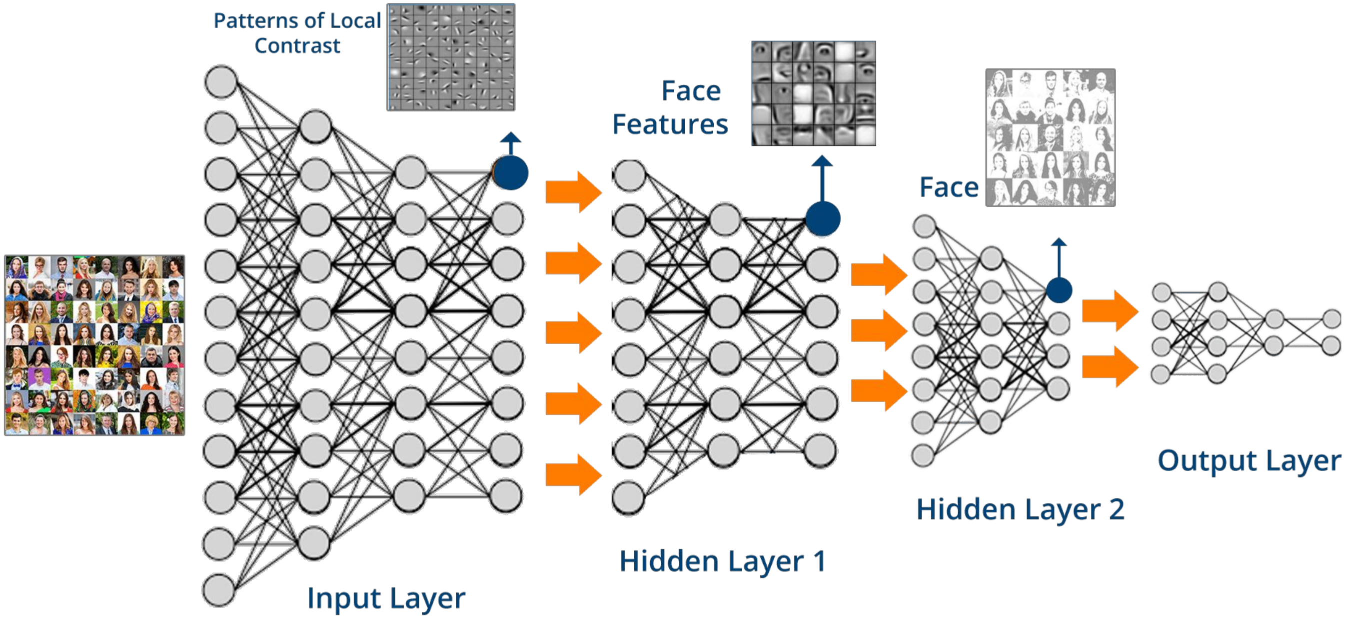


DeepLearning: Data and GPU

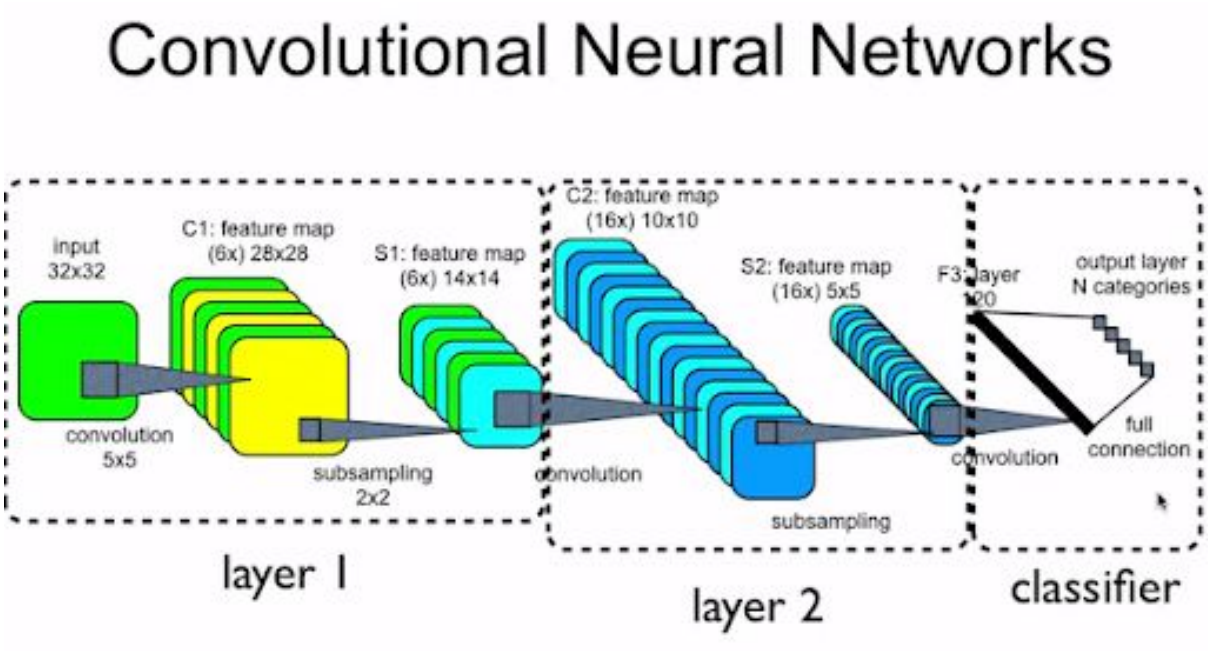
AI: Feature extraction + neural network



Deep learning can learn everything

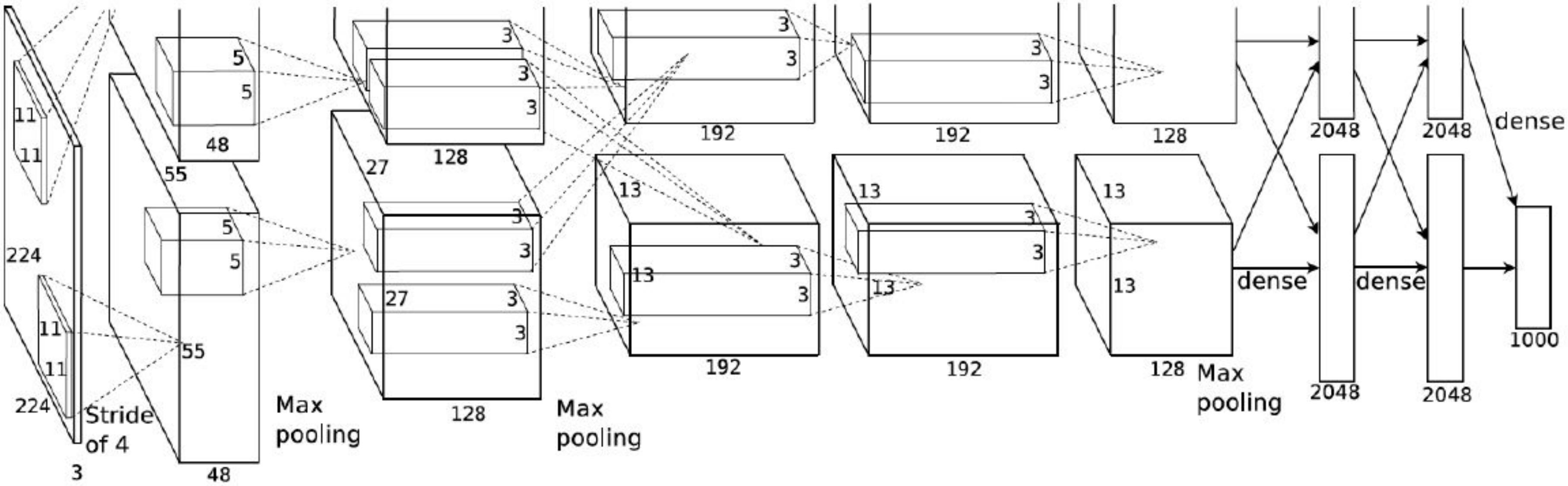


LeNet:



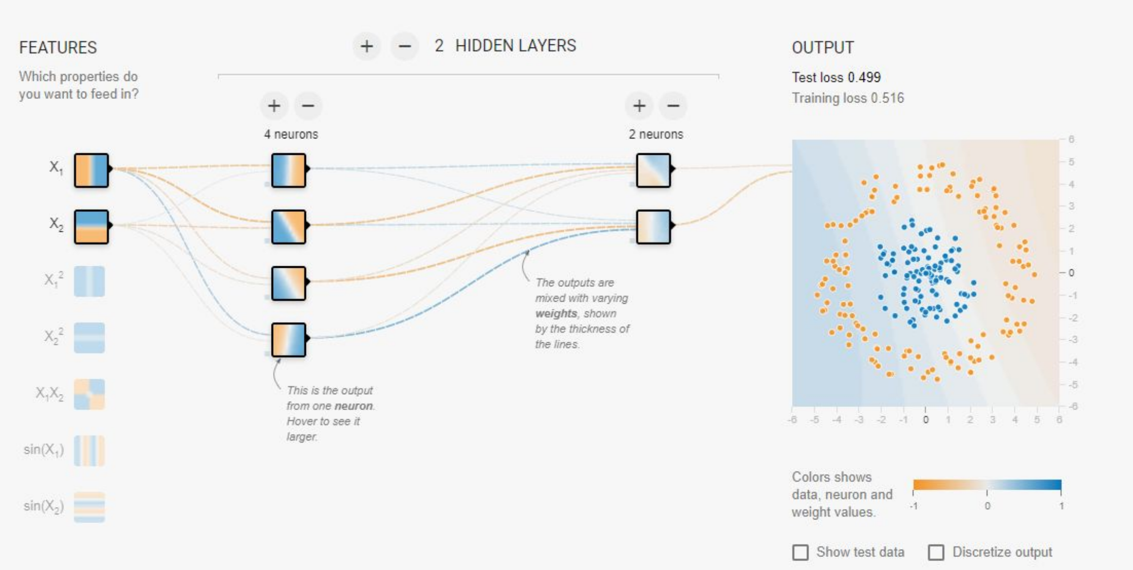
Convolutional Neural Networks
input
32x32
C1: feature map
(6x) 28x28
S1: feature map
(6x) 14x14
C2: feature map
(16x) 10x10
S2: feature map
(16x) 5x5
F3: layer
120
output layer
N categories
convolution
5x5
subsampling
2x2
convolution
subsampling
convolution
full
connection
layer 1
layer 2
classifier

Alexnet:



11
11
224
11
11
224
Stride of 4
3
5
5
48
55
5
5
55
48
Max pooling
3
3
3
27
128
3
3
27
3
3
128
Max pooling
3
3
192
13
13
3
3
192
3
3
192
13
3
3
13
192
128
13
13
128
Max pooling
dense
2048
2048
dense
2048
2048
dense
1000

AI Network examples: https://playground.tensorflow.org



AI in Healthcare

Demonstration through examples:

- Melanoma Detection
- Voice based diagnostics

There is a rule for it!

ABCDE rule

A - Asymmetry

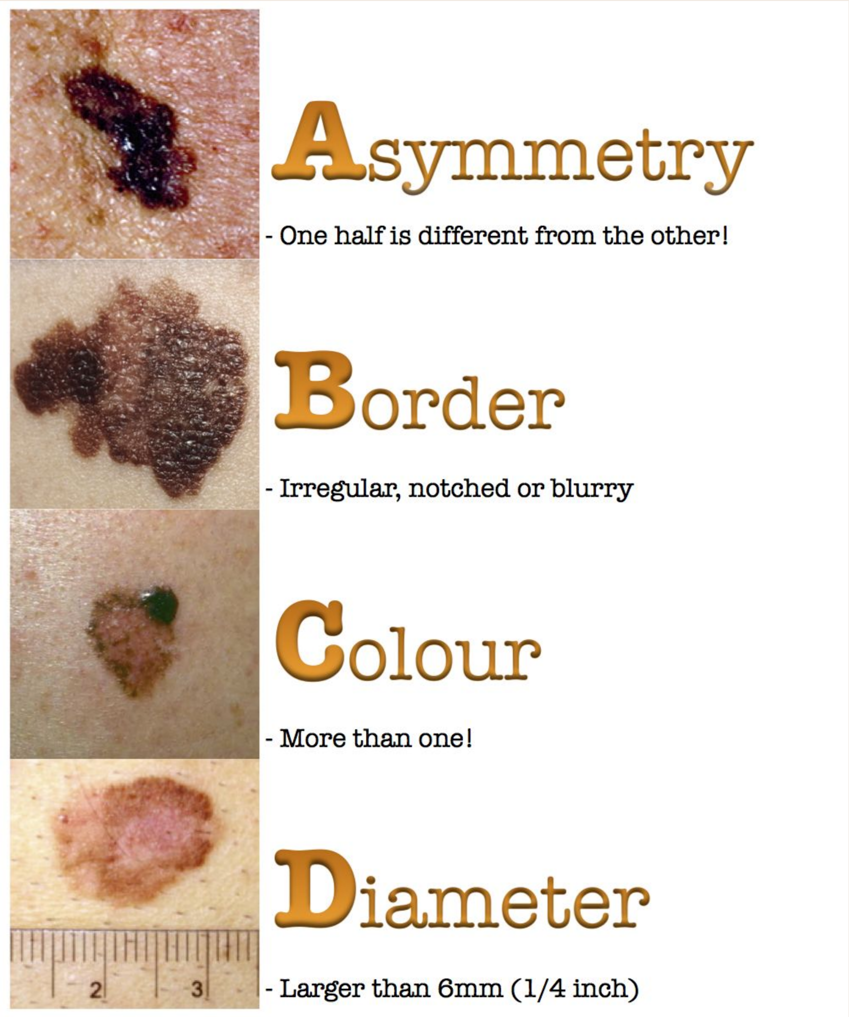
B - Border

C - Colour

D - Diameter

(E - Elevation)

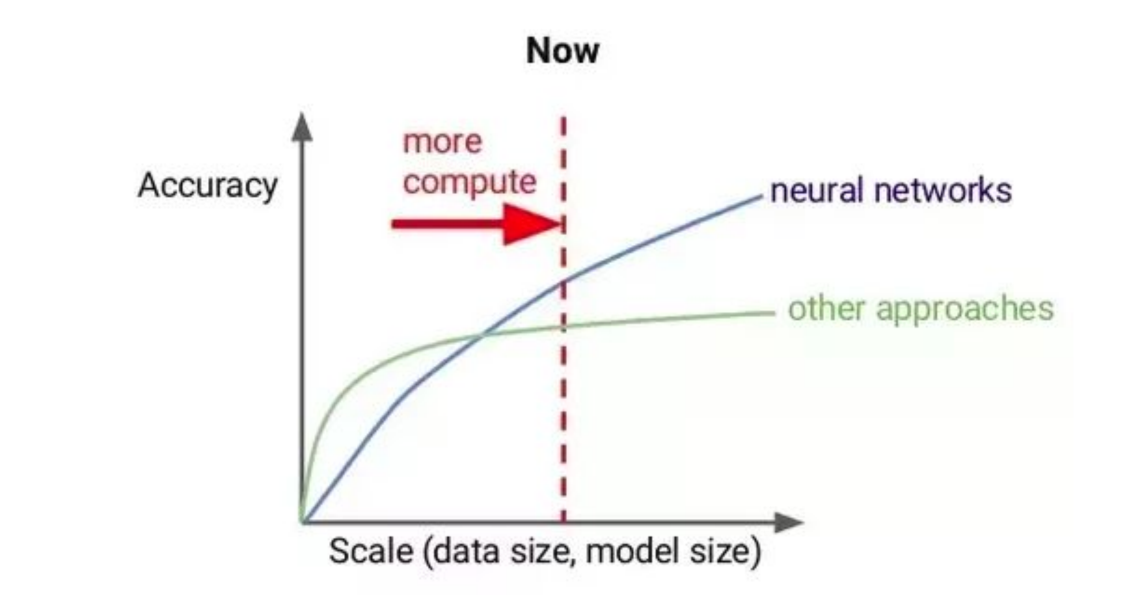
Human efficiency above 90%



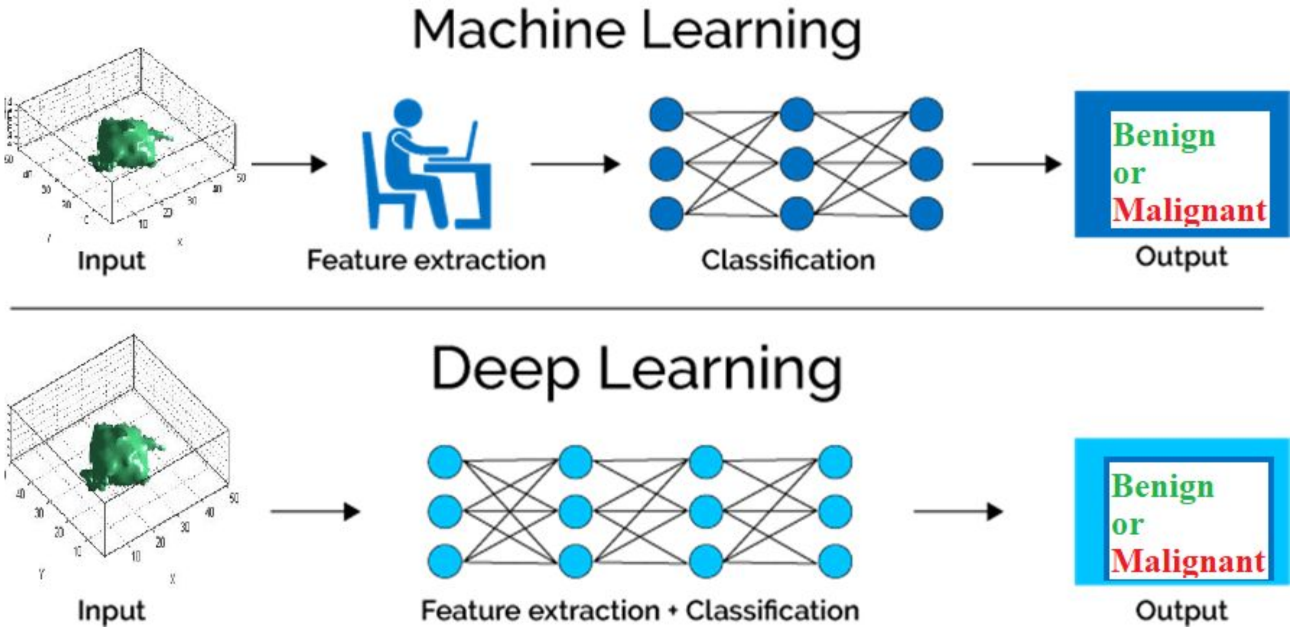
The era of deep learning

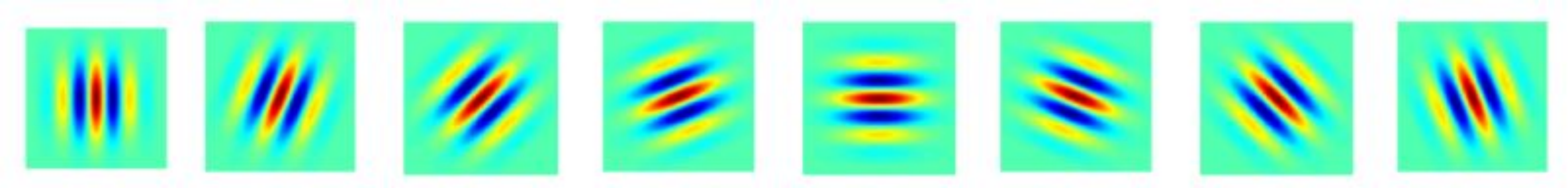
Systems using purely deep learning performed better than traditional procedures.

Recently, old techniques have begun to be combined with deep learning (ensemble learning) and these systems are the most effective.



AI: Feature extraction + artificial neural network





Traditional procedures (physics, it)

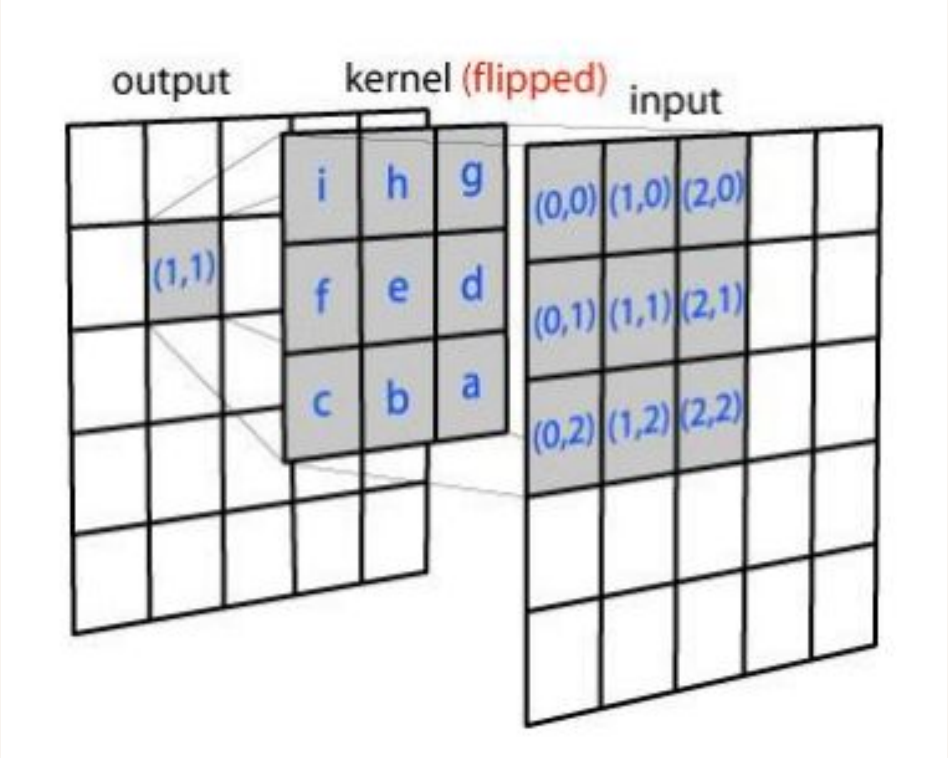
- Edge search:
  - Covers the ABD rule (Asymmetry, Border, Diameter)
  - Procedure: Gaussian derivatives
- Color search:
  - HSV (Hue, Saturation, Value)
  - DCD (Dominant Color Detection)
- Pattern recognition:
  - Gabor-filtering

Revolutionary deep learning

Deep learning has really achieved new, orders of magnitude better results.

However, these results should always be accepted with reservations:

- Black box procedure may contain errors
- The machine procedure drastically reduces false positive diagnoses

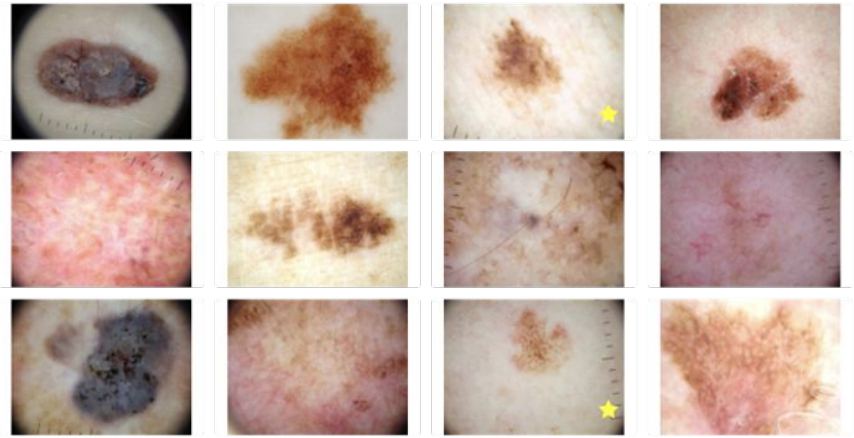


Deep learning - black box

Nature paper in 2017, cited over 1000 times

Deep neural network matching the accuracy of 21 board-certified dermatologists

BIAS: It was more likely to detect malignant cancer if there was a **ruler** in image.



Problems with deep learning

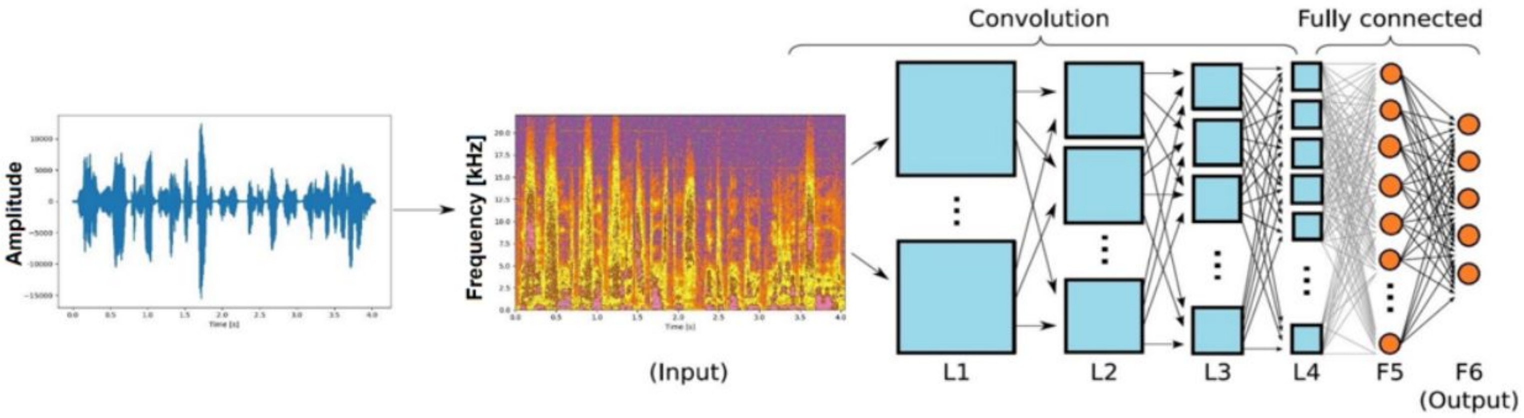
- It needs a lot of data and a lot of resources to process it (long learning time)
- He learns what he sees, he does not apply logic
- Black box, you don't see what he's doing
- Easy to cheat

Voice based diagnostics

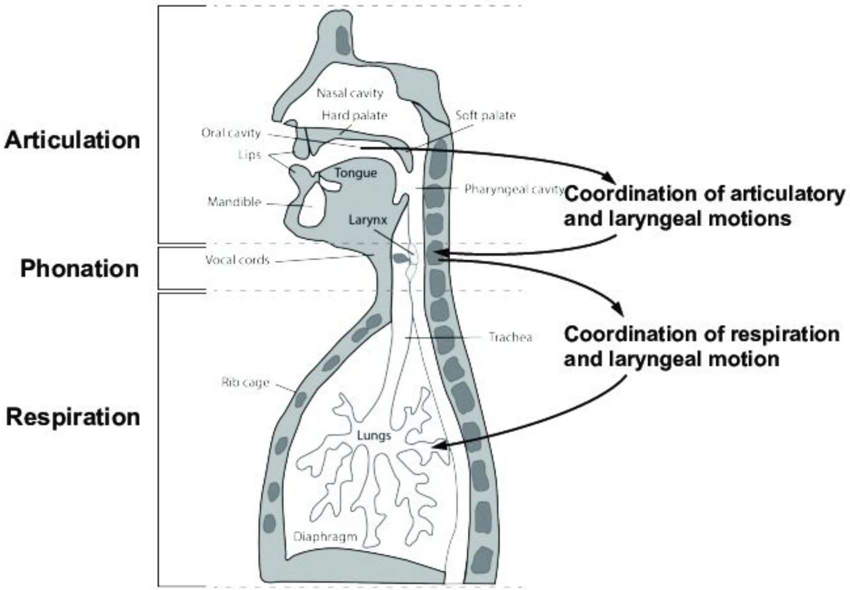
Science background: Emotion detection from sound

Previously used features in sound are waveform and spectrogram.

Features used to study abnormal sound disturbances: time, frequency, disturbances, noise, spectral structural properties



COVID-19 Voice based diagnostics



COVID-19 also uniquely modifies sound and cough.

Characteristics of speech altered as a result of the disease:

- acoustic waveform amplitude (influenced by respiratory coordination)
- speech rate and cepstrum peak (affects laryngeal movement)
- formant mid frequencies (affected by laryngeal movement and articulation)

Other diagnostic options and biomarkers

Post-traumatic stress diagnosis from voice.

Diagnosis of depression from voice and behavioral analysis.

Relationships between depression and eye movement: degree of eye opening, prolonged blinking (perhaps due to avoidance of eye contact).

Diagnosis of Alzheimer's disease based on acoustic disorder, semantic impairment, syntactic impairment, and information degradation. / Joint application of MI and linguistic analysis required /

Results

Diagnosis of depression based on eye movement with 70-75% probability / Eye movement analysis for depression detection, "IEEE International Conference on Image Processing, Melbourne, 2013 /

Diagnosis of Alzheimer’s disease averaged 7 years earlier and with an efficiency of more than 70%. / Linguistic markers predict onset of Alzheimer's disease ”The Lancet VOLUME 28, 2020 /

Post-traumatic stress disorder from speech with 89% efficiency / Speech - based markers for posttraumatic stress disorder in US veterans. Depress Anxiety. 2019 /

Pros and Cons

- Plenty of diagnostic options: with the rapid development of the area, there are opportunities for early diagnosis and prevention, even from home, implemented with home devices.
- Non-invasive, low-cost procedure: it can become massive through its low cost, making health care more cost-effective in the long run.
- With continuous observation of a person, changes can be detected.

- Black box operation: it is not possible to know exactly what the system has learned, in case of sound samples other parameters (age, body size, fatigue) may be misleading
- Few samples for teaching / analysis: 100-300 databases for a technology that requires a lot of data, reproducibility may be in doubt
- Privacy: the patient may be identifiable, sensitive medical information may become available

Summary

Learning systems provide strong support for diagnostics

Doctors don’t have to be afraid of their jobs

Physician - artificial intelligence collaboration can be fruitful

Thank You!

Zsolt Vassy
LinkGroup