

NUS-ISS

Problem Solving Using Pattern Recognition



Deep learning: Before and After

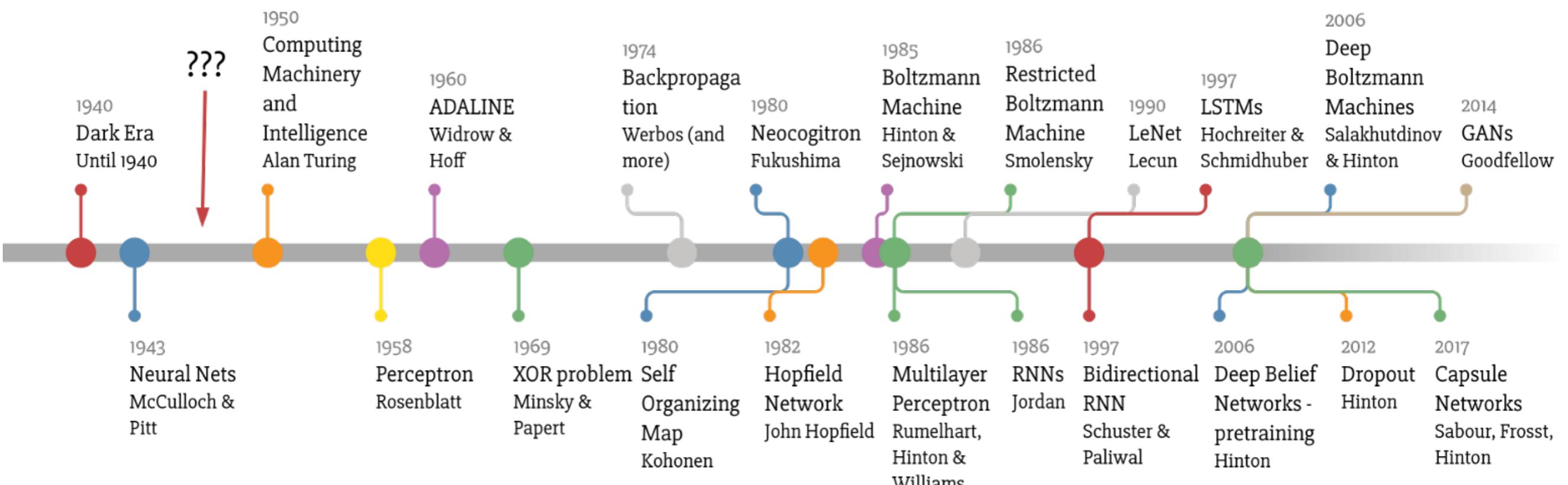
by Dr. Tan Jen Hong

© 2019 National University of Singapore.
All Rights Reserved.

Several slides have been redacted to provide better learning experience. We will upload the complete materials after the class ends.

Deep learning: The Before

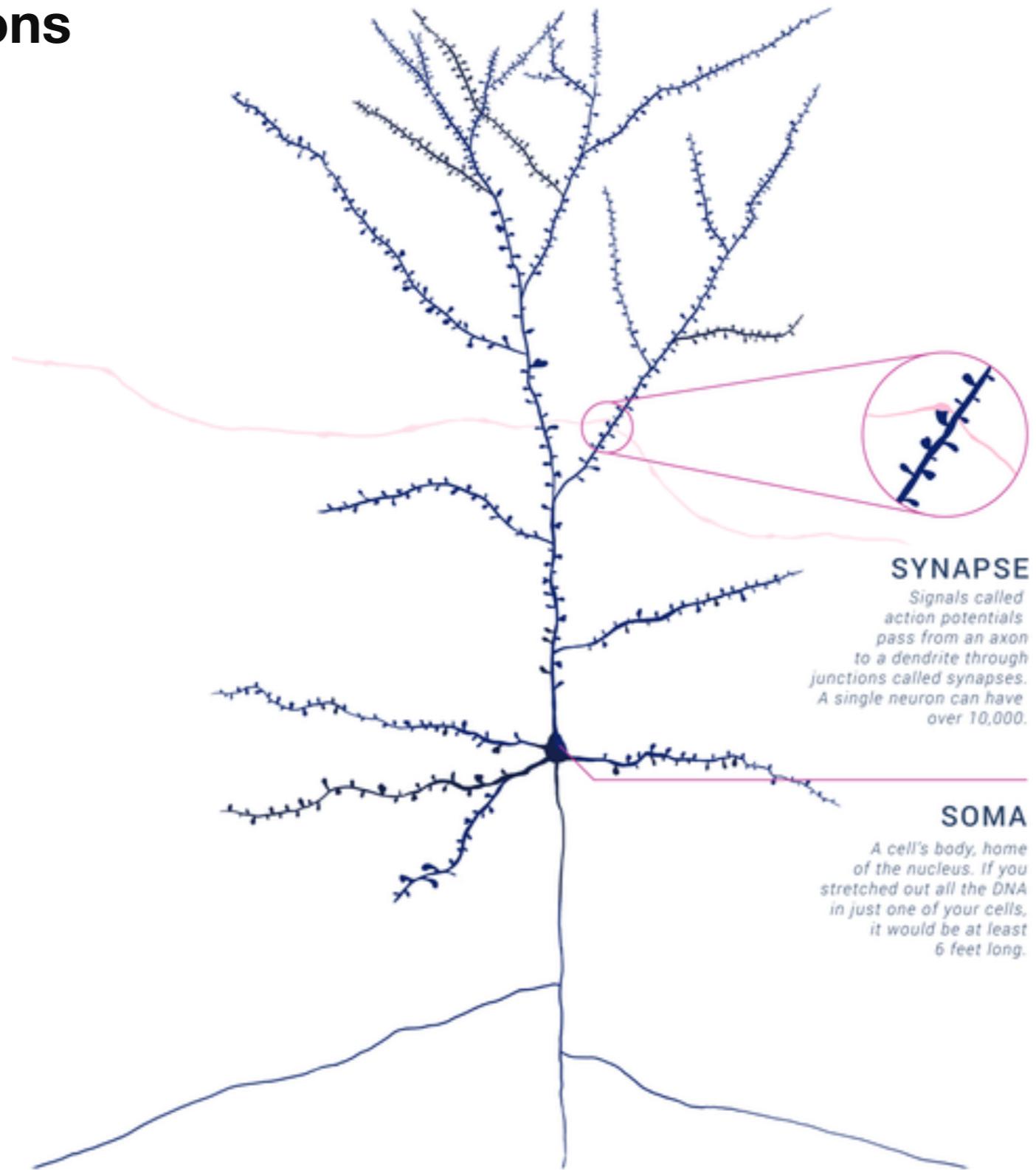
Time line Of deep learning



Source: <https://medium.com/@faviovazquez>

Biological neurons

Simplified illustration

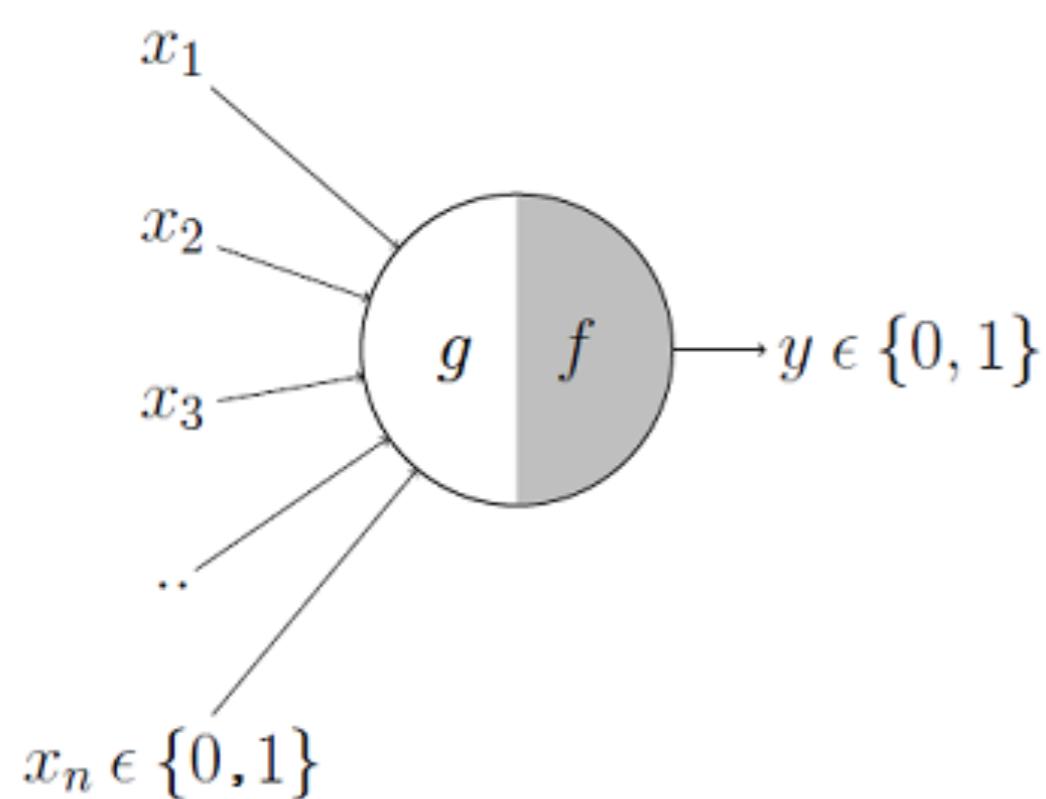


Source: https://en.wikipedia.org/wiki/File:Anatomy_of_a_Neuron_with_Synapse.png

The first artificial neuron

McCulloch-Pitts Neuron

- By Warren McCulloch (neuroscientist) and Walter Pitts (logician) in 1943



$$g(x_1, x_2, x_3, \dots, x_n) = g(\mathbf{x}) = \sum_{i=1}^n x_i$$

$$\begin{aligned} y &= f(g(\mathbf{x})) = 1 && \text{if } g(\mathbf{x}) \geq \theta \\ &= 0 && \text{if } g(\mathbf{x}) < \theta \end{aligned}$$

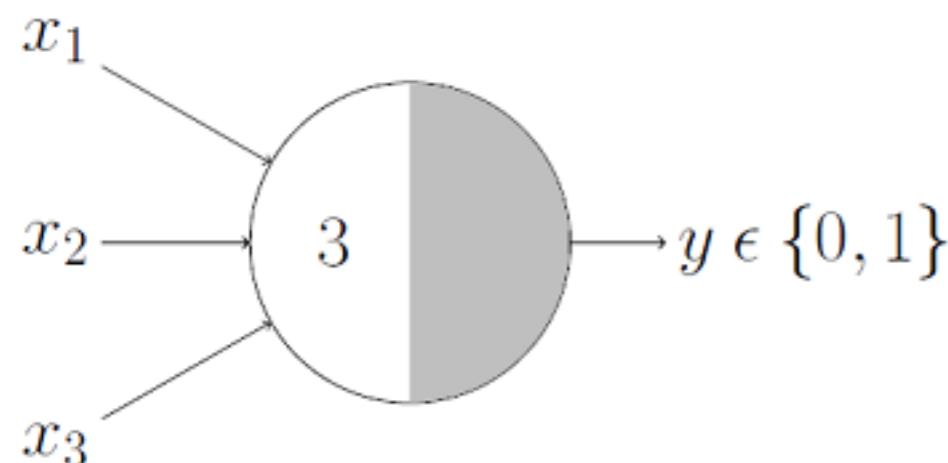
Source: [https://en.wikipedia.org/wiki/
File:Anatomy_of_a_Neuron_with_Synapse.png](https://en.wikipedia.org/wiki/File:Anatomy_of_a_Neuron_with_Synapse.png)

The first artificial neuron

McCulloch-Pitts Neuron

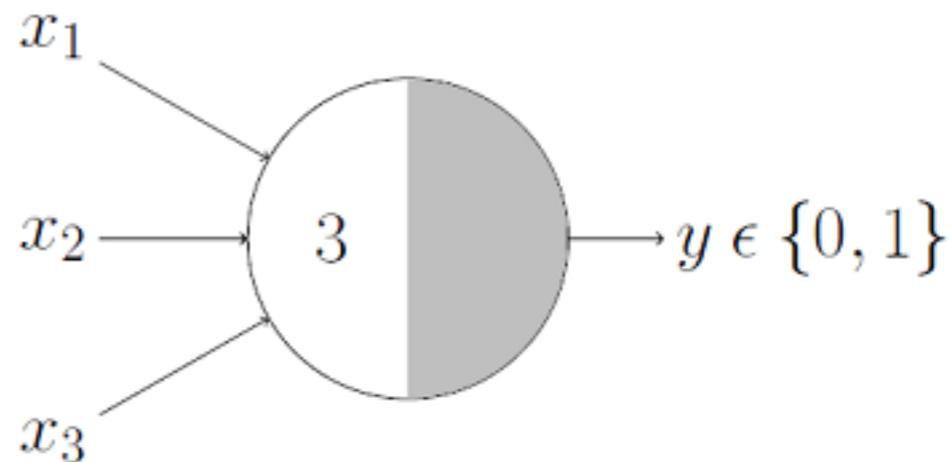
- Can be used to represent a few Boolean functions

AND function



$$y = 1 \quad \text{if} \quad x_1 + x_2 + x_3 \geq 3$$

OR function



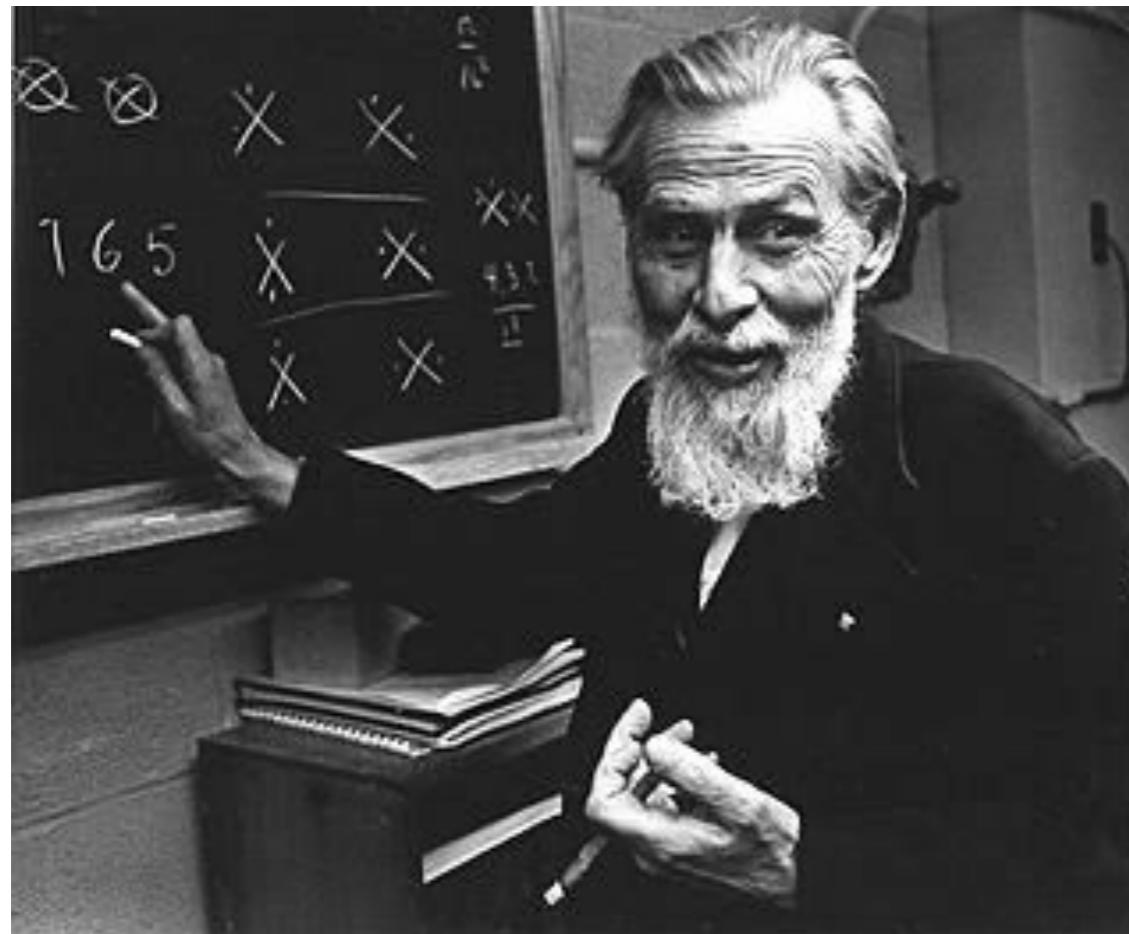
$$y = 1 \quad \text{if} \quad x_1 + x_2 + x_3 \geq 1$$

Source: <https://towardsdatascience.com/mcculloch-pitts-model-5fdf65ac5dd1>

The first artificial neuron

McCulloch-Pitts Neuron

- Inputs accepts only boolean values
- No learning algorithm



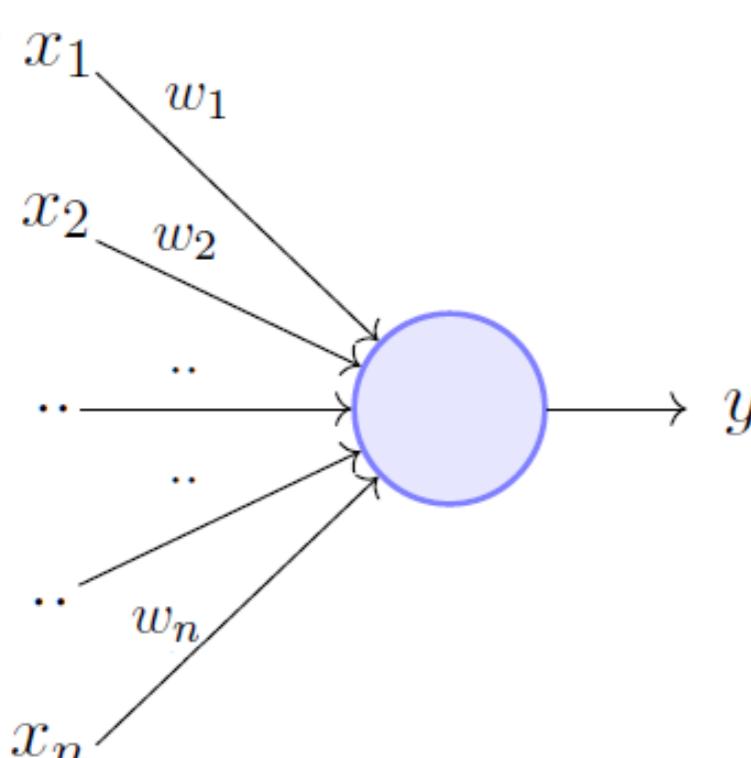
Left: Warren S. McCulloch. Right: Walter H. Pitts Jr.

The improved artificial neuron

Perceptron

- By Frank Rosenblatt, refined by Minsky and Papert

- Support real inputs, not just boolean values



$$y = 1 \quad if \sum_{i=1}^n w_i * x_i \geq \theta$$
$$= 0 \quad if \sum_{i=1}^n w_i * x_i < \theta$$

Rewriting the above,

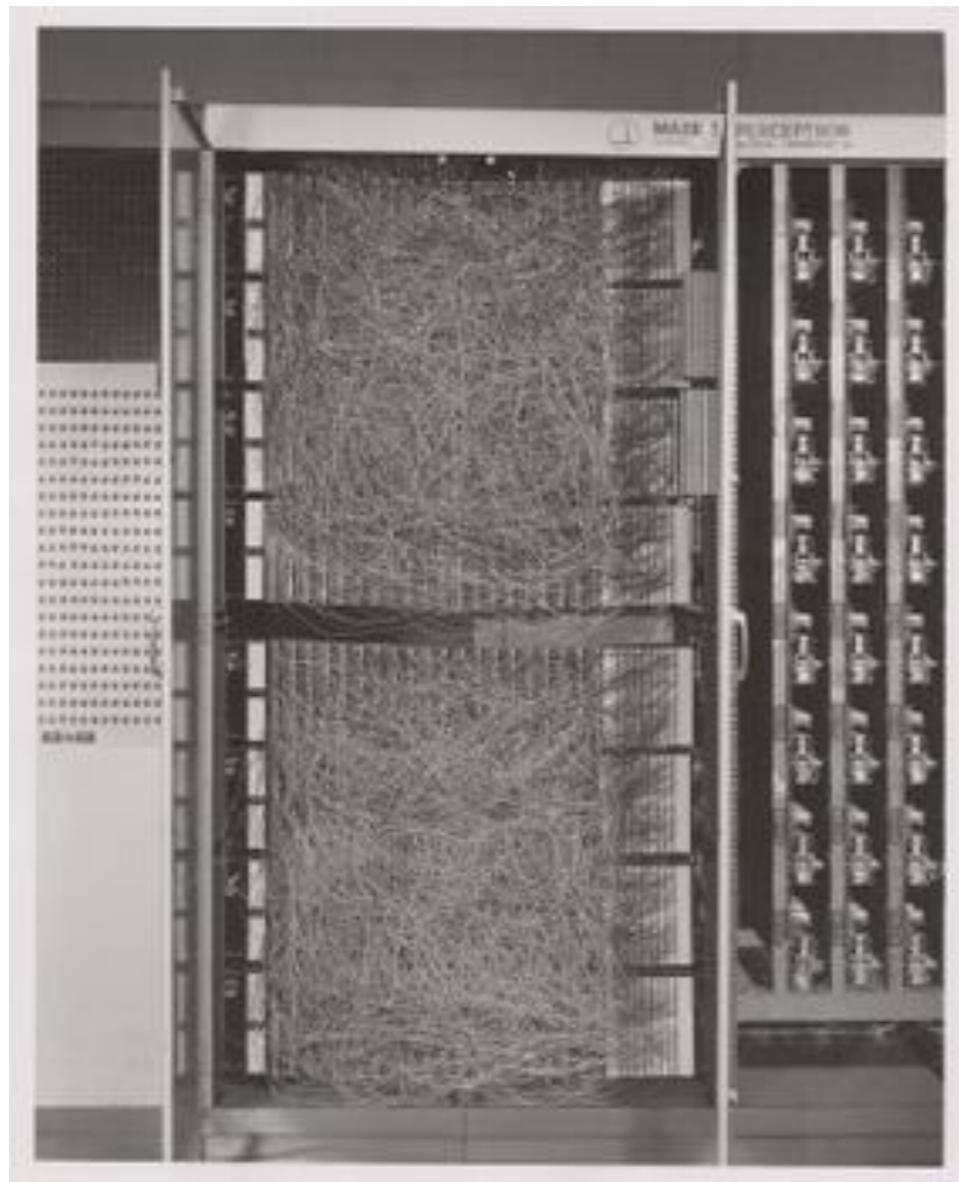
$$y = 1 \quad if \sum_{i=1}^n w_i * x_i - \theta \geq 0$$
$$= 0 \quad if \sum_{i=1}^n w_i * x_i - \theta < 0$$

Source: <https://towardsdatascience.com/perceptron-the-artificial-neuron-4d8c70d5cc8d>

The improved artificial neuron

Perceptron

- Rosenblatt's achievement: artificial neurons could actually learn from data



- He came up a supervised learning algorithm!
- He implemented Perceptron in custom hardware, which can learn to classify simple shapes correctly with 20x20 pixel-like inputs

Source: 'Mark I Perceptron at the Cornell Aeronautical Laboratory', hardware implementation of the first Perceptron
(Source: Wikipedia / Cornell Library)

The first AI winter

XOR affair

- Marvin Minsky, founder of MIT AI lab, and Seymour Papert, director of the lab in 1969 published a book named 'Perceptrons'



Source: <https://amethix.com/2018/06/ai-winter-is-coming/>

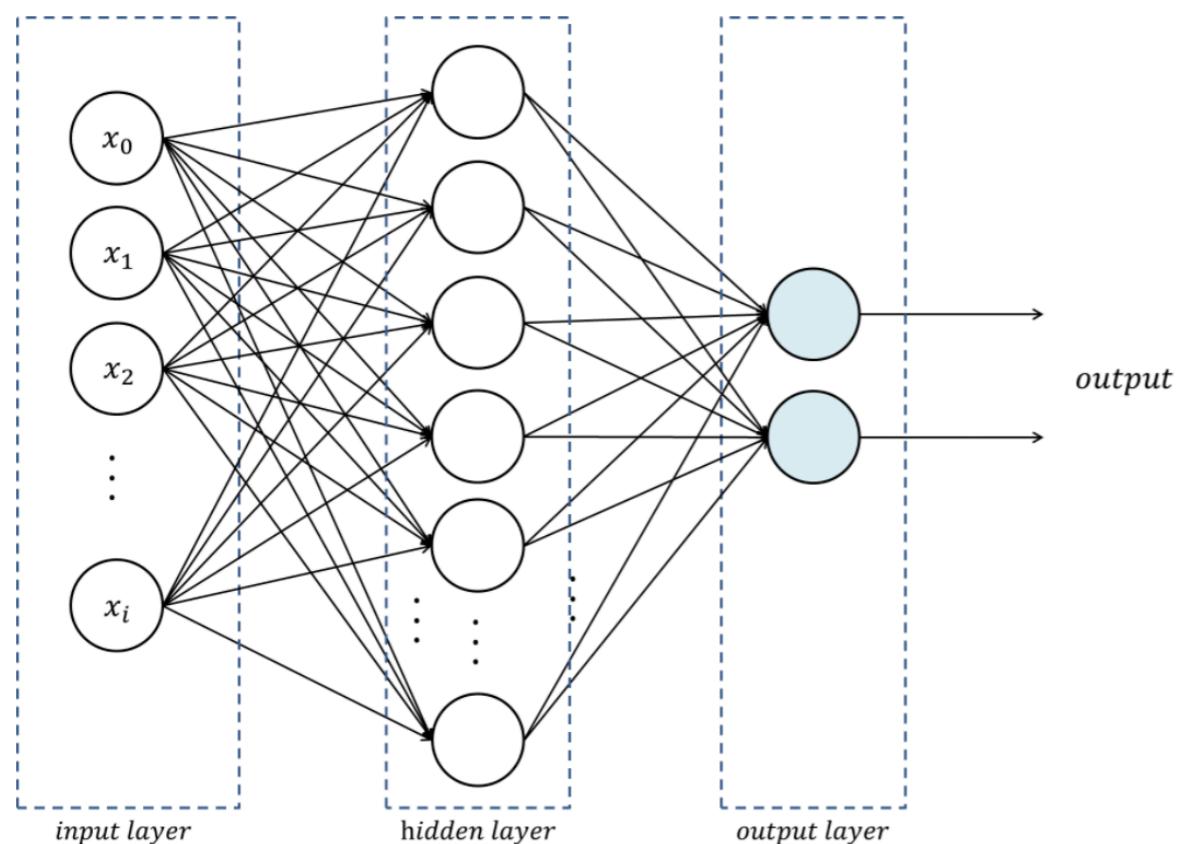
- They showed that a single perceptron cannot do XOR
- Multiple layers of Perceptron can do XOR, but the proposed learning algorithm does not work for that!
- Here comes the winter ...

Where is the learning algorithm?

The thaw of AI winter

- Multilayer layers of perceptron should work, but need learning algorithm

- Between 60s and 80s, several researchers separately derived the solution, but few people knew



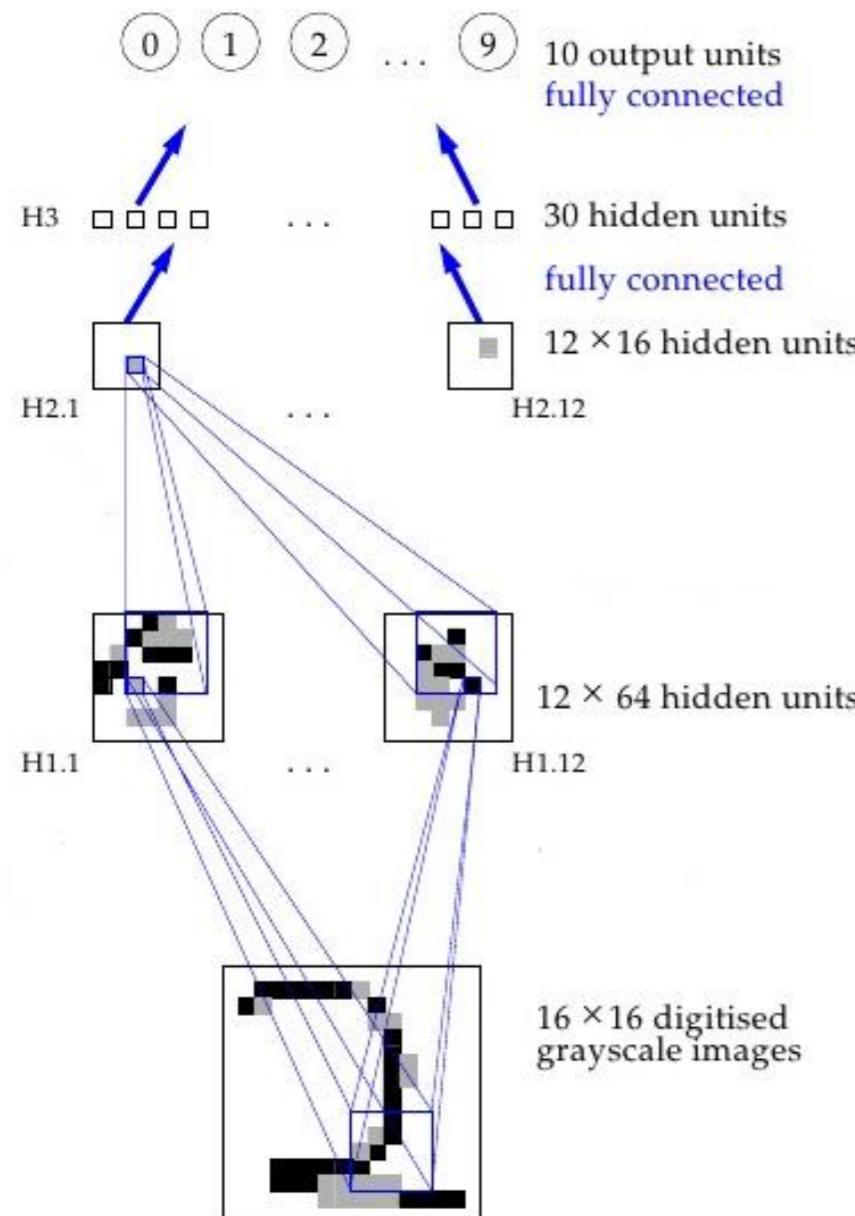
- In 1986, Rumelhart, Hinton Williams published a method in Nature
- They called the learning procedure "backpropagation"

Source: <https://www.cc.gatech.edu/~san37/post/dlhc-fnn/>

Year 1989

Annus mirabilis, sort of

- Multilayer feedforward networks are proved to be universal approximators



- LeNet was proposed and put into actual significant use: recognizing numbers
- It is a neural net + convolutional layers (weight sharing)

Source: <http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning-part-2/>

Another winter dawns

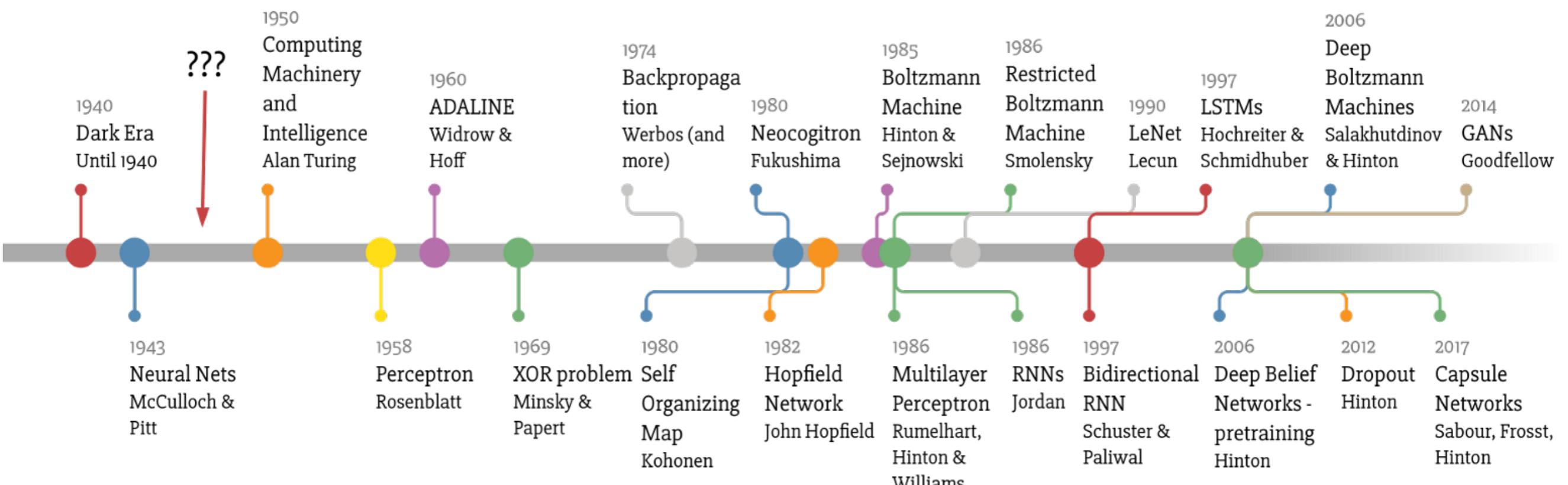
by backpropagation



Source: <https://amethix.com/2018/06/ai-winter-is-coming/>

- By late 80s, already knew that deep neural net was hard to train
- Deep neural nets trained with backpropagation did not work very well, not as well as nets with fewer layers
- Support vector machine came into fashion
- Random forests, with lovely mathematical theory behind proved to be effective

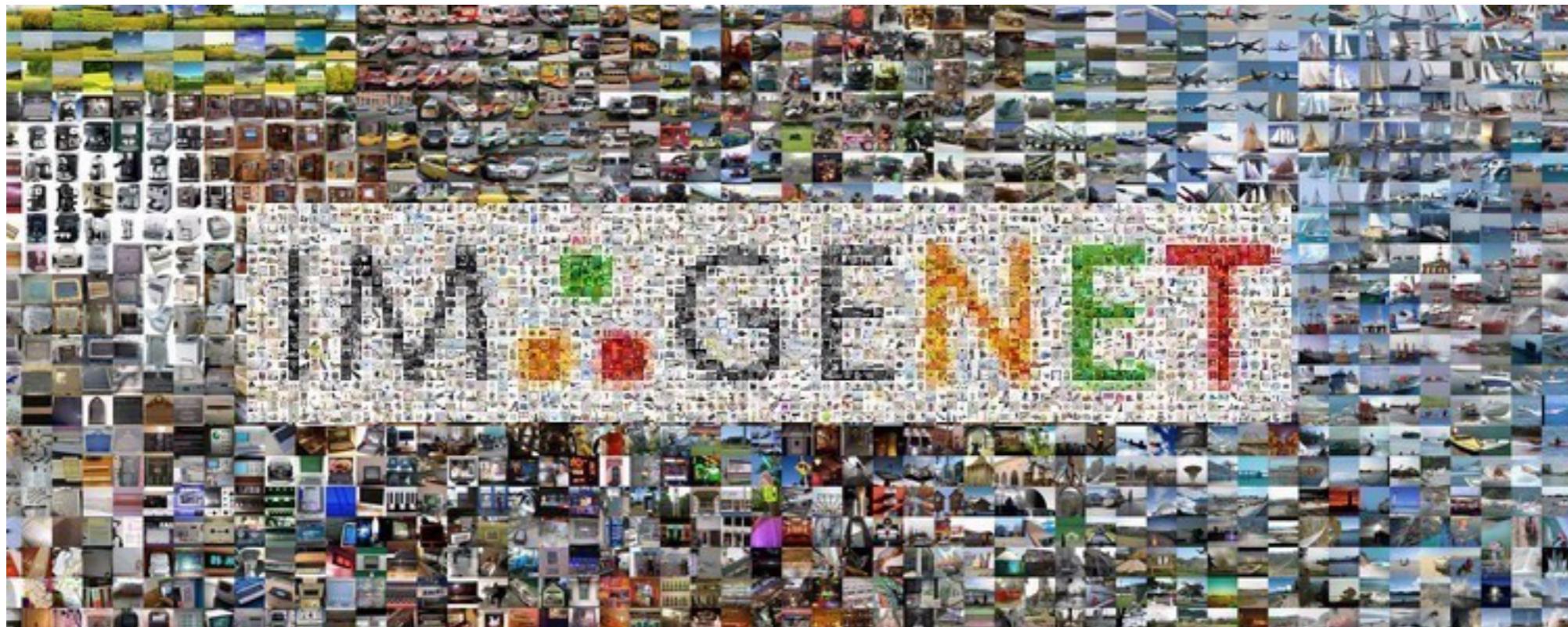
Time line Of deep learning



Source: <https://medium.com/@faviovazquez>

ImageNet

by Stanford Vision Lab



Source: https://gluon-cv.mxnet.io/build/examples_datasets/imagenet.html

Competition on ImageNet

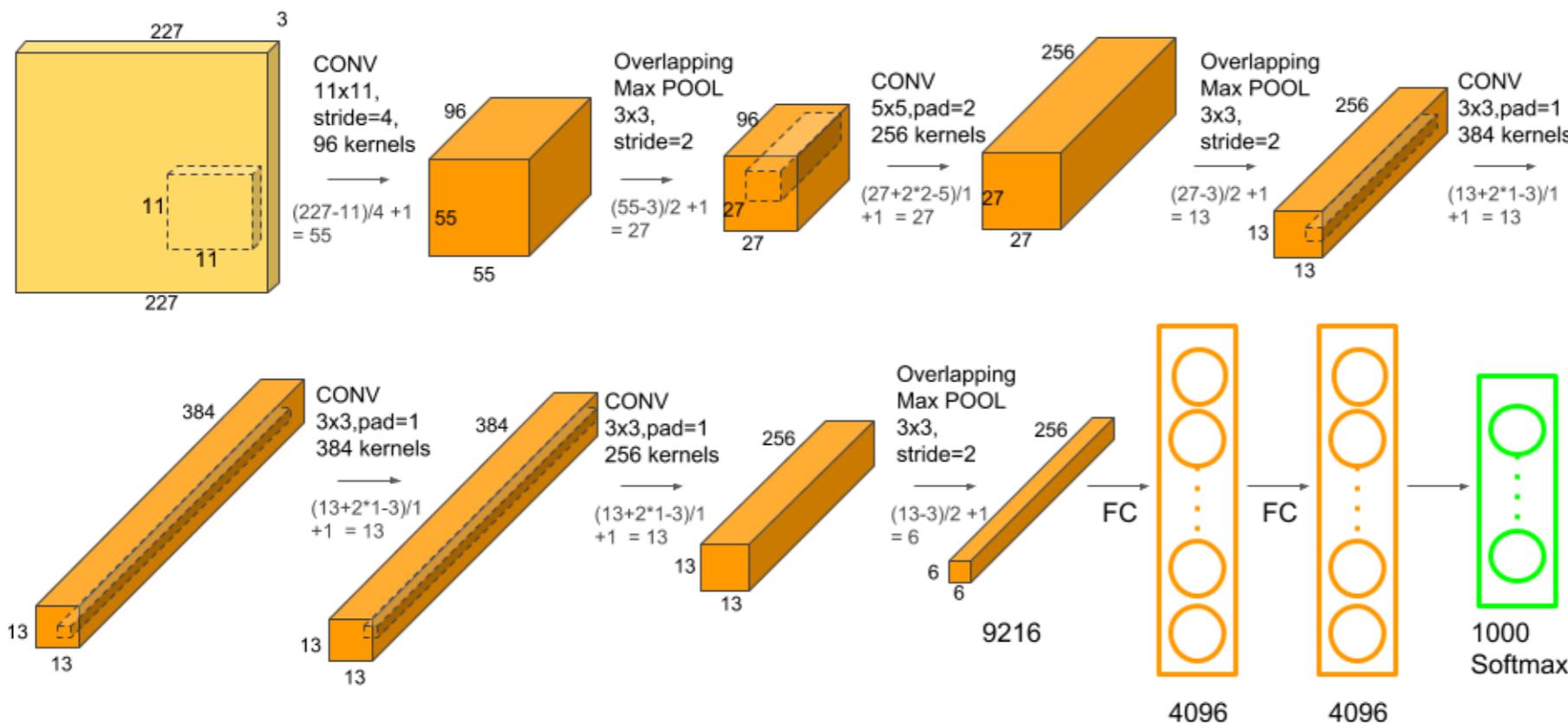
- ILSVRC: ImageNet Large Scale Visual Recognition Challenge
- Started from 2010; teams evaluate their algorithms on given data set, compete to achieve highest accuracy on visual recognition tasks.
- ILSVRC training dataset: 1000 object categories, 1.2 million images



Source: <http://www.image-net.org>

- Team Hinton entered the competition

- They achieved an error rate of 15.3%, far far better than the next closest: 26.2%



Source: <https://neurohive.io/en/popular-networks/alexnet-imagenet-classification-with-deep-convolutional-neural-networks/>

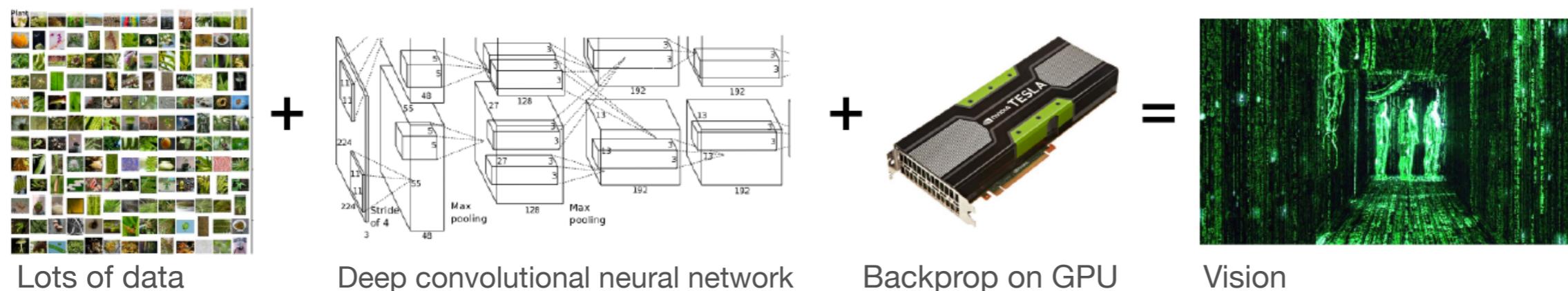
AlexNet

Key of success?

- The net structure is not the most important point

- Use of Rectified Linear Unit (ReLU) activation function
- Use of dropout
- GPU implementation (through CUDA)

Deep learning computer vision recipe



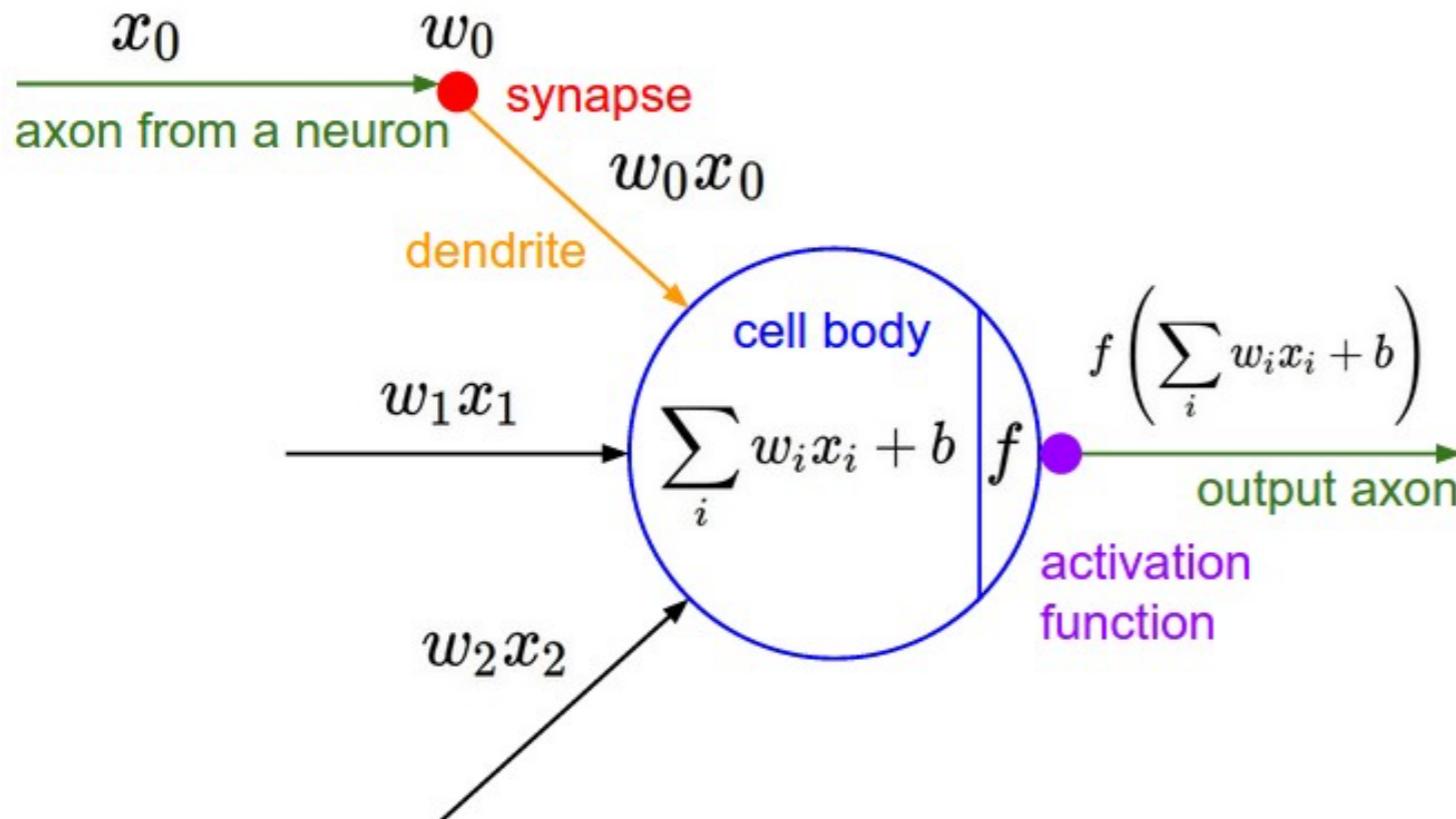
Source: <http://www.computervisionblog.com/2015/05/deep-learning-vs-big-data-who-owns-what.html>

ReLU

Simple is better

- The zero value output from ReLU introduces sparsity representation (more zeros in each layer, only the important neurons contribute)

- ReLU is easier to calculate

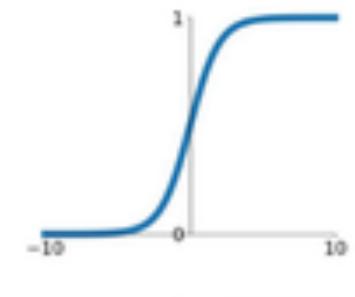


Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$

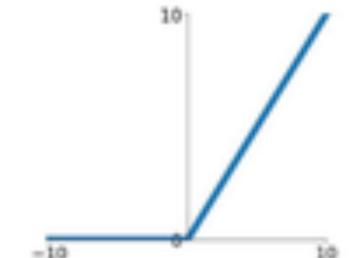
tanh

$$\tanh(x)$$



ReLU

$$\max(0, x)$$

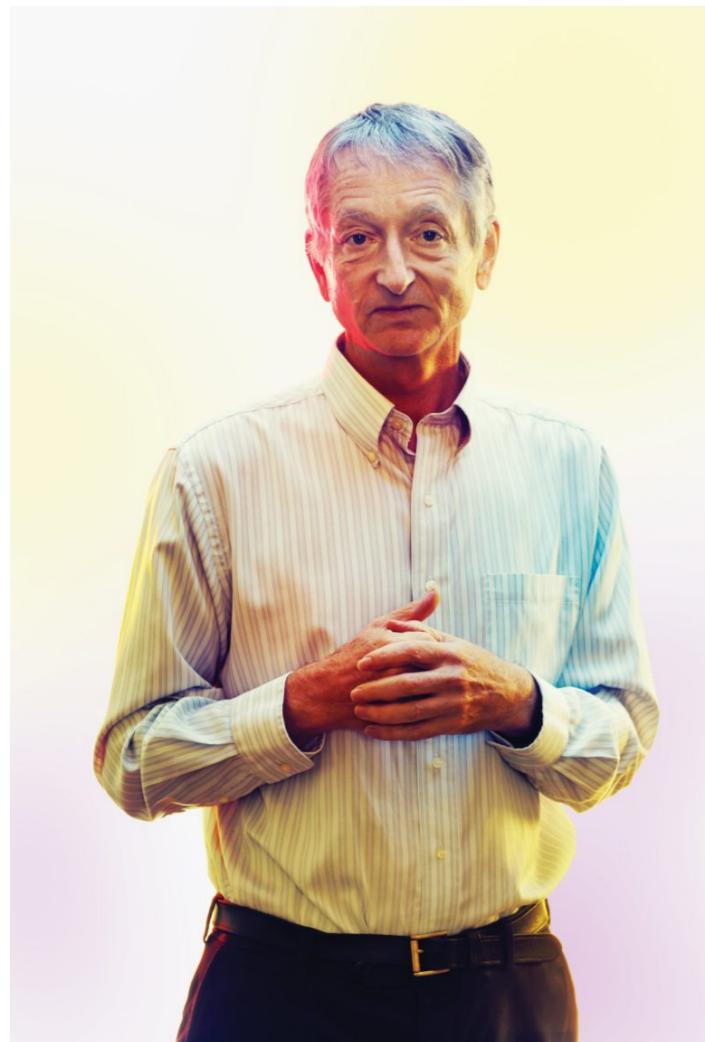


Source: <http://www.andreykurenkov.com/writing/ai/a-brief-history-of-neural-nets-and-deep-learning/>

The big question

Why did backpropagation fail in past?

- Labelled datasets were thousands times too small
- Computing power was millions times slower
- Initialized the weight in stupid ways
- Used the wrong type of non-linearity for activation function



Source: <https://torontolife.com/tech/ai-superstars-google-facebook-apple-studied-guy/>

Since then ...

theano

K Keras

Caffe

DEEPMLEARNING4J

TensorFlow

P Y T O R C H

mxnet



Deep learning: The After

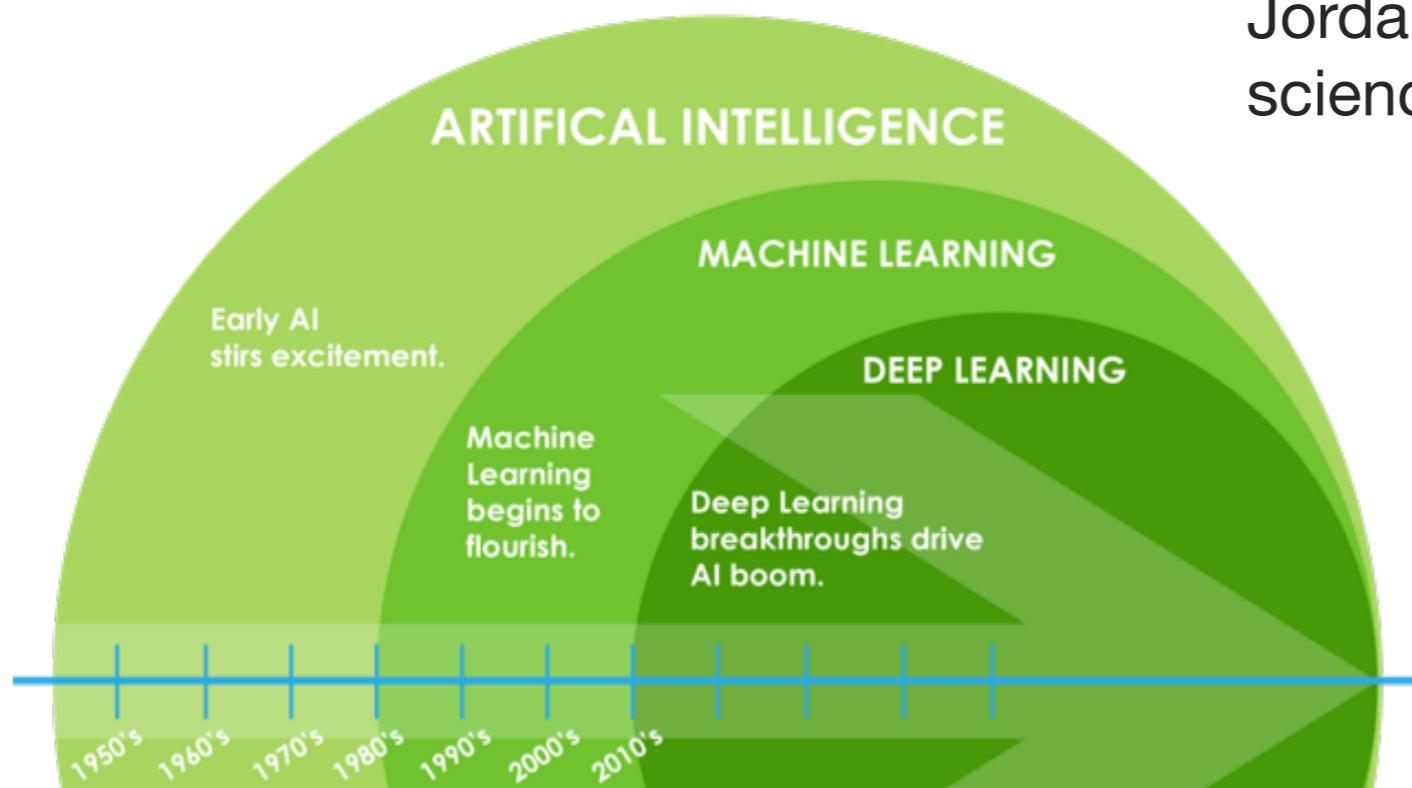
Confusion?

The differences among the few terms

- Arthur Samuel coined the term 'machine learning' in 1959 while at IBM

- Rina Dechter introduced the term deep learning in 1986

- Machine learning and statistics are closely related, thus Michael I. Jordan suggested the term 'data science' to refer to the overall field



Source: <https://buzzrobot.com/difference-between-artificial-intelligence-machine-learning-and-deep-learning-ccfd779eca7b>

Confusion?

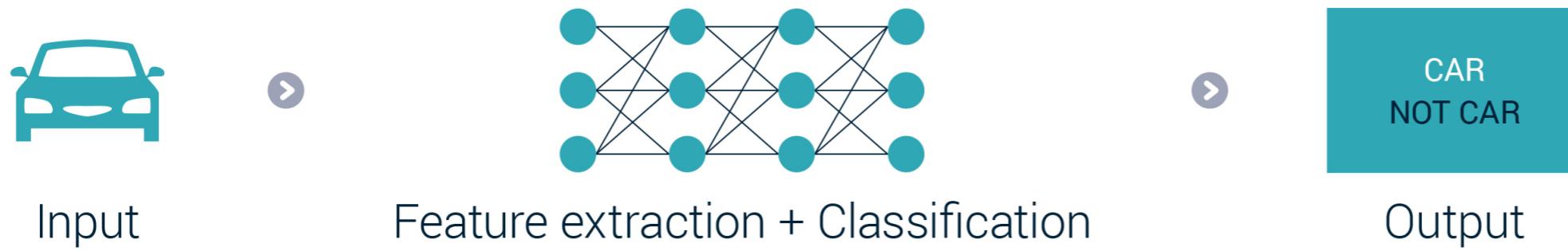
The differences among the few terms

- Feature: a number or a vector that describes something about the input

Machine Learning



Deep Learning

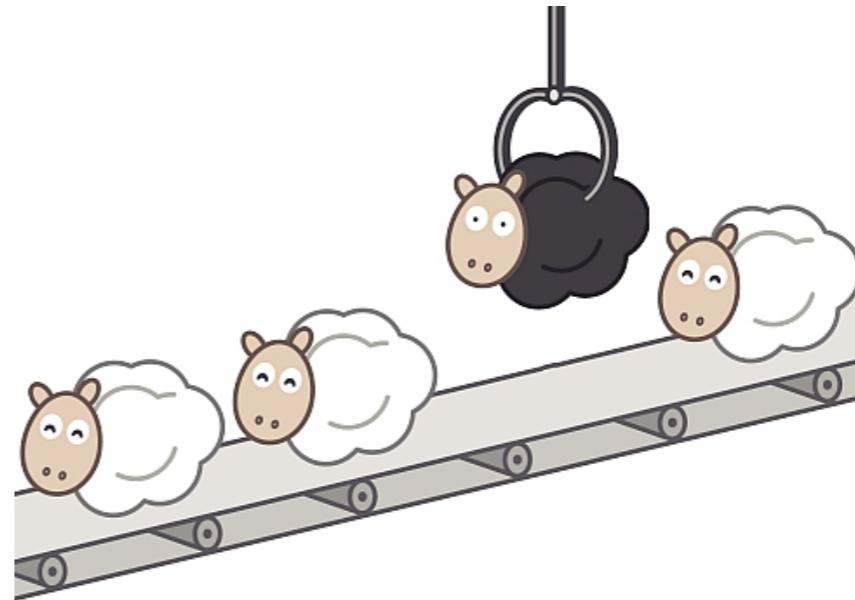


Source: <https://verhaert.com/difference-machine-learning-deep-learning/>

Application

Three main categories

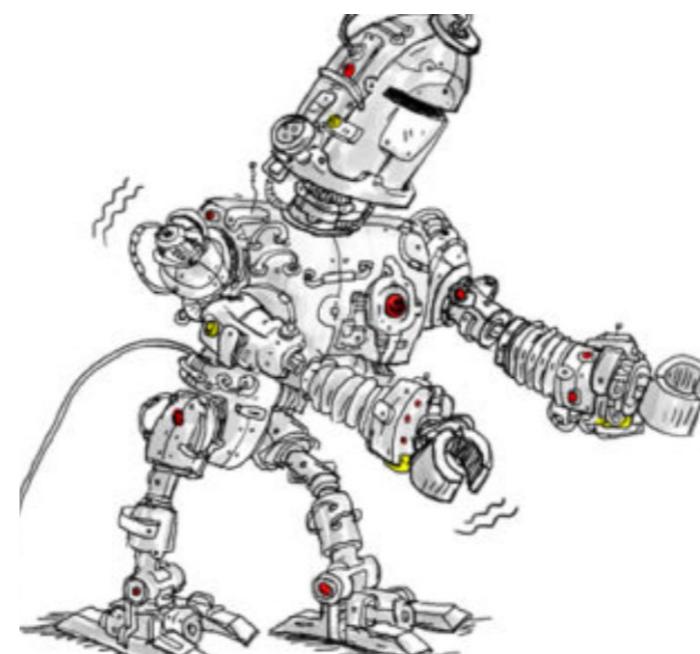
Identify



Source: <http://blog.ss8.com>



Create



Act

Source: <http://www.williamalone.com/articles/create-html5-canvas-javascript-drawing-app/>

Source: <https://bitsandatoms.co/tag/reinforcement-learning/>