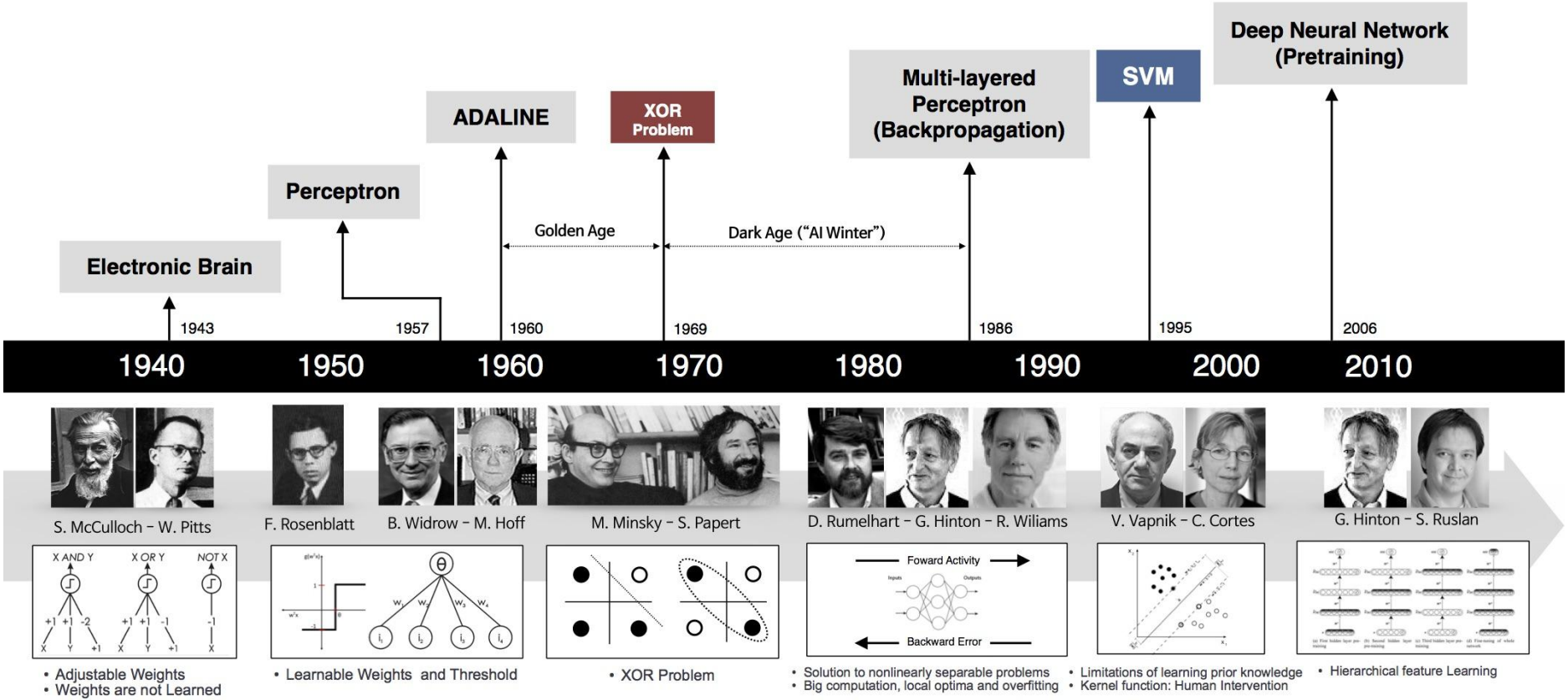
Huntsville Data Science, Artificial Intelligence, & Machine Learning

February 07, 2018

Deep Learning History



Beginnings

Alan Turing - Computing Machinery and Intelligence (1950)

https://www.csee.umbc.edu/courses/471/papers/turing.pdf

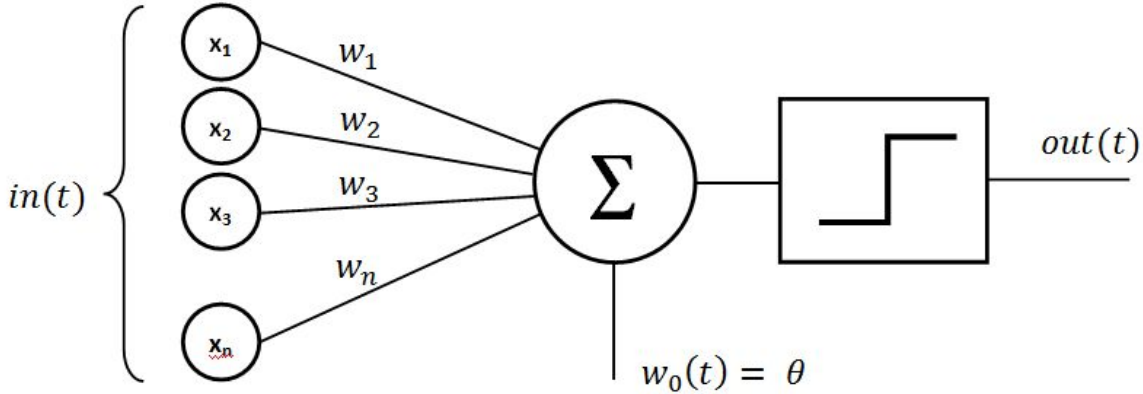
This paper introduce the concept of what is now known as the “Turing Test”

The state of the art at the time was the Threshold Logic Unit - mostly based on current knowledge of how neurons were thought to work.

Perceptron

The perceptron algorithm was invented in 1957 at the Cornell Aeronautical Laboratory by Frank Rosenblatt,funded by the United States Office of Naval Research The perceptron was intended to be a machine, rather than a program, and while its first implementation was in software for the IBM 704, it was subsequently implemented in custom-built hardware as the "Mark 1 perceptron". This machine was designed for image recognition: it had an array of 400 photocells, randomly connected to the "neurons". Weights were encoded in potentiometers, and weight updates during learning were performed by electric motors

.



AI Winter (1969)

Cause of AI Winter

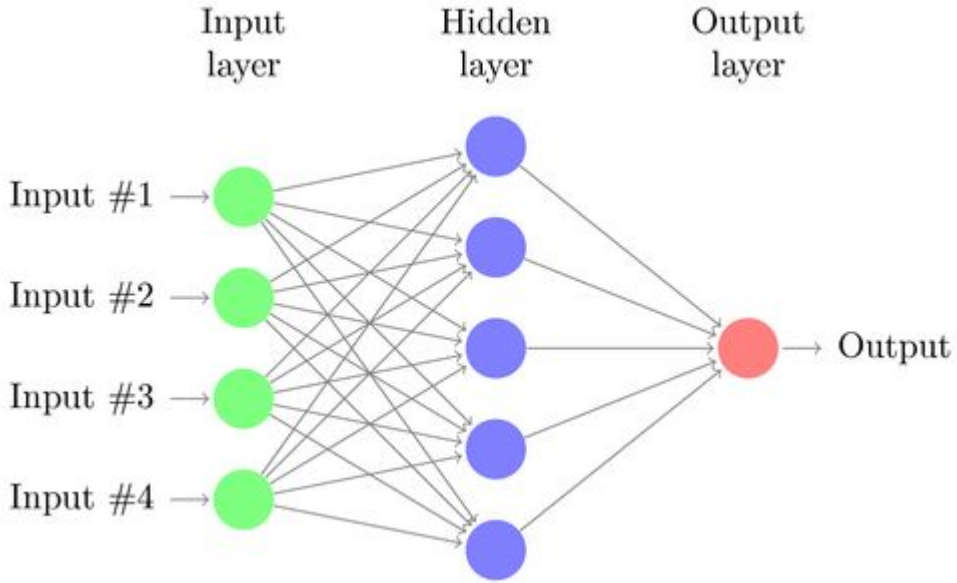
Perceptrons: an introduction to computational geometry is a book written by Marvin Minsky and Seymour Papert and published in 1969.

It offered a mathematical proof that the perceptron could not approximate an XOR function given an infinite training set.

I never liked XOR anyway.

Multilayer Perceptrons

By stacking several layers of perceptrons, researchers were able to overcome the XOR problem.



Backpropagation (1986)

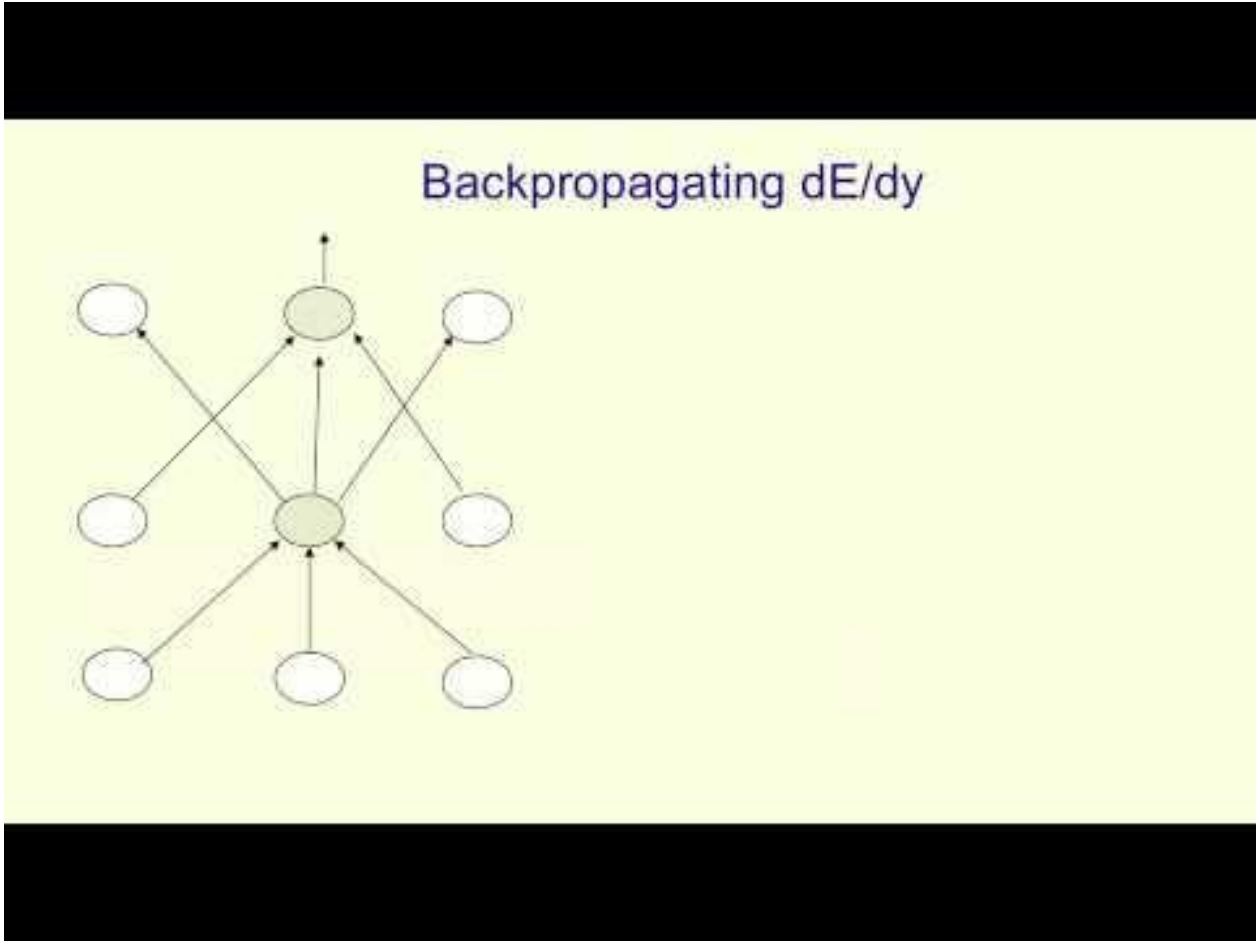
Geoff Hinton, along with David Rumelhart and Ronald Williams, published a paper entitled “Learning representations by back-propagating errors”

https://www.iro.umontreal.ca/~vincentp/ift3395/lectures/backprop_old.pdf

This provided a mechanism for training multilayer perceptron networks.

Also about this time, the **universal approximation theorem** states that a feed-forward network with a single hidden layer containing a finite number of neurons (i.e., a multilayer perceptron), can approximate continuous functions on compact subsets of $\mathbf{R}^n$, under mild assumptions on the activation function.

Hinton explains...



Learning before “Deep”

Gradient based learning (1998) - Yann Lecun - http://yann.lecun.com/exdb/publis/pdf/lecun-01a.pdf

CNN from Yann Lecun (AT&T Bell Labs) could recognize handwritten digits.

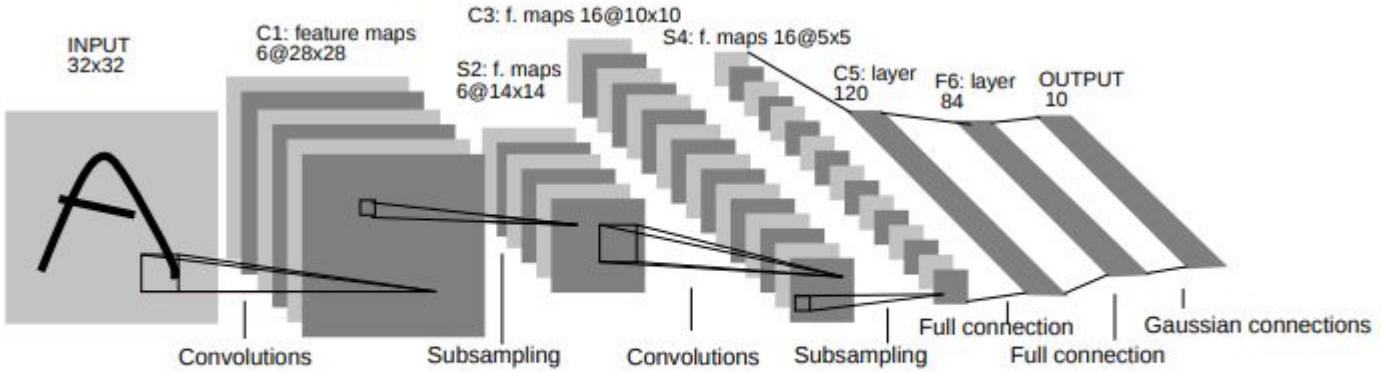


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Rolling in the Deep

Deep Learning (2006) - Again with Geoff HInton. The idea was to train a simple 2-layer unsupervised model like a restricted boltzman machine, freeze all the parameters, stick on a new layer on top and train just the parameters for the new layer. You would keep adding and training layers in this greedy fashion until you had a deep network, and then use the result of this process to initialize the parameters of a traditional neural network. Using this strategy, people were able to train networks that were deeper than previous attempts, prompting a rebranding of ‘neural networks’ to ‘deep learning’.

Filling the void with Hardware and Data

Imagenet (2009) - millions of labeled images created and published by Fei-Fei Li at Stanford

MNIST - Handwritten digits

Google House Numbers from street view

Flickr 30k Image dataset

More - https://github.com/ChristosChristofidis/awesome-deep-learning#datasets

GPU - used for multi-core floating point calculation

Custom chipsets from Intel (Nervana) https://ai.intel.com/ and NVIDIA

https://www.nvidia.com/en-us/deep-learning-ai/

Exponential Improvements

Alexnet (2012) - Won the Large Scale Visual Recognition Challenge(LSVRC) with an error rate 10% lower than the previous year. Used dropout to reduce overfitting and a rectified linear activation unit (ReLU)

Generative Adversarial Networks (2014) - Ian Goodfellow

Speech recognition:

https://venturebeat.com/2017/05/17/googles-speech-recognition-technology-now-has-a-4-9-word-error-rate/

https://techcrunch.com/2017/08/20/microsofts-speech-recognition-system-hits-a-new-accuracy-milestone/

https://hacks.mozilla.org/2017/11/a-journey-to-10-word-error-rate/

Future History

Open source libraries are allowing new products to be developed very quickly.

For instance:
https://towardsdatascience.com/hey-google-where-is-my-pet-tensorflow-object-detection-contribution-9c1d1fdd0443

Google's TensorFlow
Microsoft's CNTK
Amazon's DSSTNE
Theano
Torch
Caffe
Deeplearning4J

Questions/Comments?