MCnet Workshop
Lund, Sweden
June 22-27, 2020

Artificial Neural Networks – Historical Notes
Generics and High Energy Physics

Carsten Peterson
home.thep.lu.se/~carsten
Computational Biology & Biological Physics
Lund University, Sweden

Early development:

AI and Machine Learning

- The early years
- The "winter"
- Artificial Neural Networks (ANN) takes on
- ANN and thermodynamics (optimizers)
- Moving on - towards "Deep Learning"

High Energy Physics (early 90'ies):

- Jet origin identification
- W-mass reconstruction
- Track finding

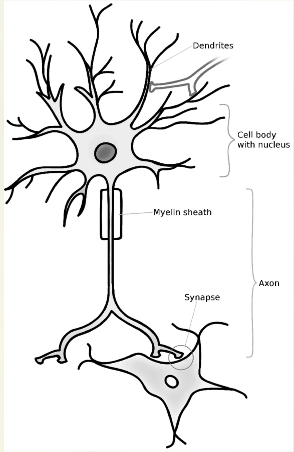
The early days – from logical rules to learning by experience

1950–85: Ambitions grew to create programs exhibiting Artificial Intelligence (AI). Initially based upon step-by-step rules with limitations for e.g. image recognition.

The early days – from logical rules to learning by experience

1950–85: Ambitions grew to create programs exhibiting Artificial Intelligence (AI). Initially based upon step-by-step rules with limitations for e.g. image recognition.

1943: W. McCullogh & W. Pitts developed a computational model of interacting neurons with threshold logics – launching pad for Artificial Neural Networks (ANN) and neuroscience.

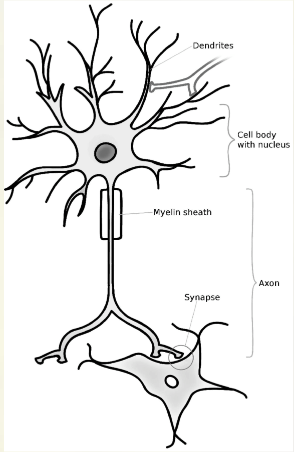


1949: D. Hebb proposed how neural activity affects the connections in between neurons - plasticity

The early days – from logical rules to learning by experience

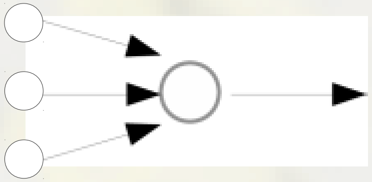
1950–85: Ambitions grew to create programs exhibiting Artificial Intelligence (AI). Initially based upon step-by-step rules with limitations for e.g. image recognition.

1943: W. McCullogh & W. Pitts developed a computational model of interacting neurons with threshold logics – launching pad for Artificial Neural Networks (ANN) and neuroscience.



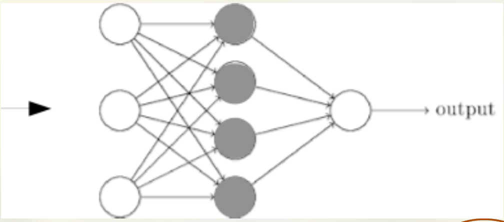
1949: D. Hebb proposed how neural activity affects the connections in between neurons.

1958: F. Rosenblatt developed a learning algorithm for setting the weights between the input and output neurons -- the Perceptron.



1969: M. Minsky & S. Papert published influential work showing limitations of the Perceptron with regard to logical rules, which strangled the whole field – the “winter”.

1975: P. Werbos resolved the problem by introducing an extra layer of “hidden nodes” as well as a method for adjusting the weights – backpropagation of errors.



The early days – from logical rules to learning by experience

1986: D. Rumelhart, G. Hinton & R. Williams streamlined the backpropagation method for usage in several applications.This Multilayer Perceptron (MLP) took off. The funding “winter” was replaced by “early spring”!

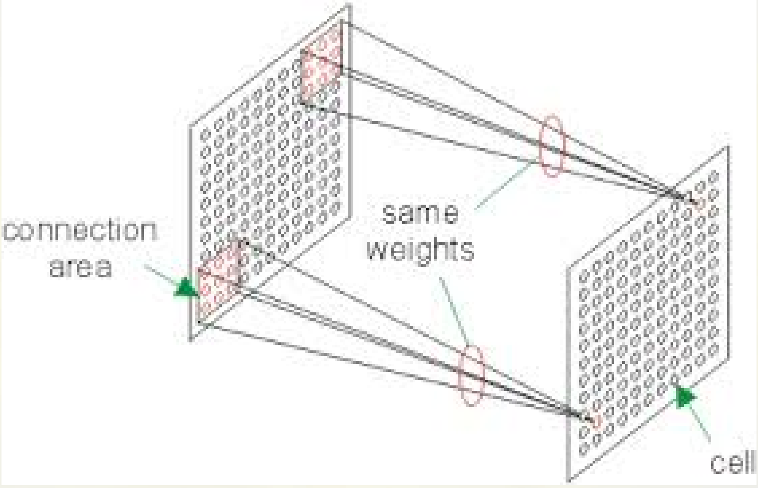
The departure from logical rules to “learning from examples” makes sense biologically. It enables the human recognition of objects (secs) using neurons with response times (microsecs) in 100 steps only.

The field took off!

CP joined

1989: K. Hornik “proved” that Multilayer Feedforward Networks are Universal Function Approximators – including continuous-valued outputs. Holds for most threshold functions.

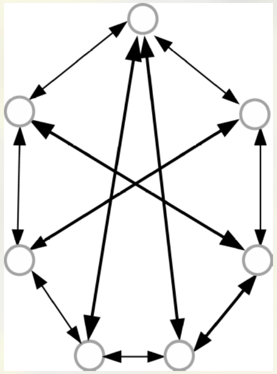
1992: J. Nowlan & G. Hinton simplified networks by soft weight-sharing.



The Hopfield model

1982: J. Hopfield developed a feedback network for an associative memory

Hebb rule learning (D.O. Hebb; 1949)



Associative memory

From a distorted word move to closest attractor

Local search

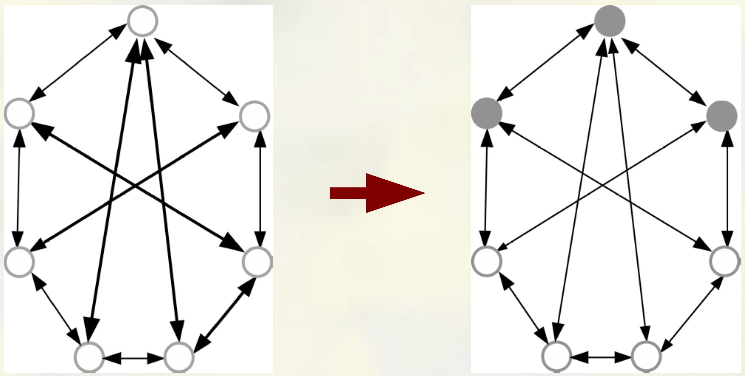
$$E = -1/2 \sum w_{ij} s_i s_j$$

$$\Delta E / \Delta s_i = \sum w_{ij} s_j$$

$$\Delta w_{ij} \propto p_i p_j$$

The Boltzmann machine

1983: D. Ackley, T. Sejnowski & G. Hinton developed the Boltzmann machine starting from the Hopfield model.



← Hidden/output nodes

← Input nodes

Local search

global search

$$P \propto e^{-E/kT}$$

Interesting alternative to feed-forward backpropagation

Important:
Provides distribution of weights

Learning rule adapted from Hebb

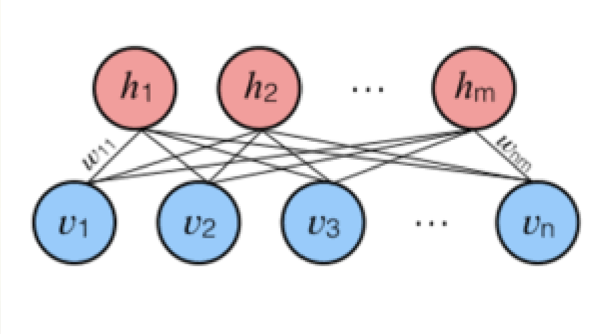
Very CPU consuming → Restricted Boltzmann (RBM) (Hinton; 2006)

The restricted Boltzmann machine for auto-encoding

Use Boltzmann machine with hidden nodes only to reduce number of DoF – develop autoencoding layer distributions (Hartman & Peterson; 1989)

Feasible with RBF
no hidden interconnections

Gateway to Deep Learning



Movie categories (Netflix)

Quantum MC

Statistical physics mean field approximation → can include hidden interconnections (Peterson & Anderson; 1987)

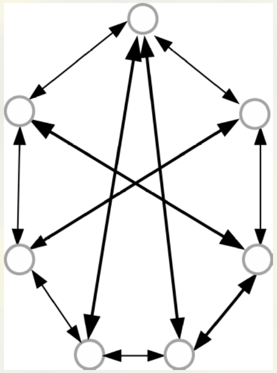
Variational procedures for Bayesian inference

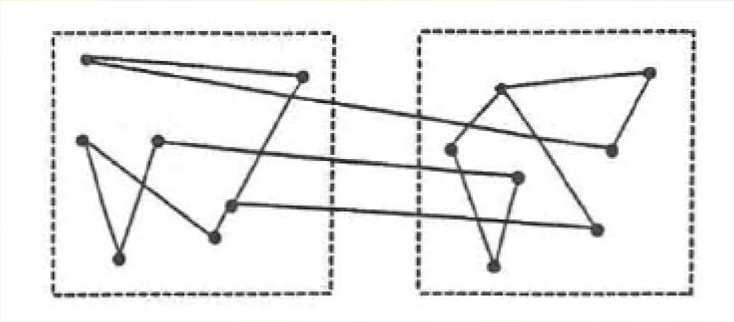
A sidetrack:
Neural optimizers

Solving optimization problems with ANN

1985: J. Hopfield & D. Tank developed a neural optimization scheme for difficult assignment problems (NP hard).

Example: Graph Bisection





$S_i = 1$ $S_i = -1$

$$E = -1/2 \sum w_{ij} s_i s_j$$

$$P \propto e^{-E/kT}$$

W_{ij} minimizes traffic in between & equal partition

In general one can also have s_{ia} for i->a (Potts spin)

Suitable for Quantum Computing

Putting it all together and more

2000 - : Ever increasing access to “big data”
Ambitions to handle many (i) input units and (ii) hidden layers

Yet, keep # of parameters low

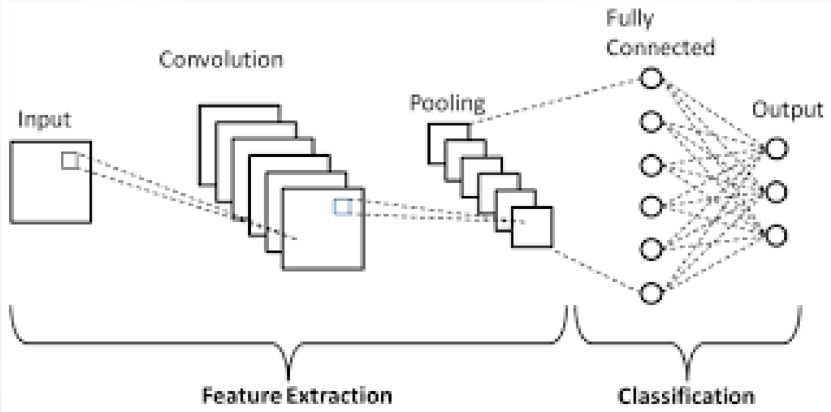
1995 – 2015: Y. Lecun, Y. Bengio & G. Hinton developed deep learning

Weight sharing

Autoencoders (RBM)

Pooling
ReLu
...
..

Deep Learning
Convolutional Neural Networks (CNN)



M. Ohlsson (next lectures)

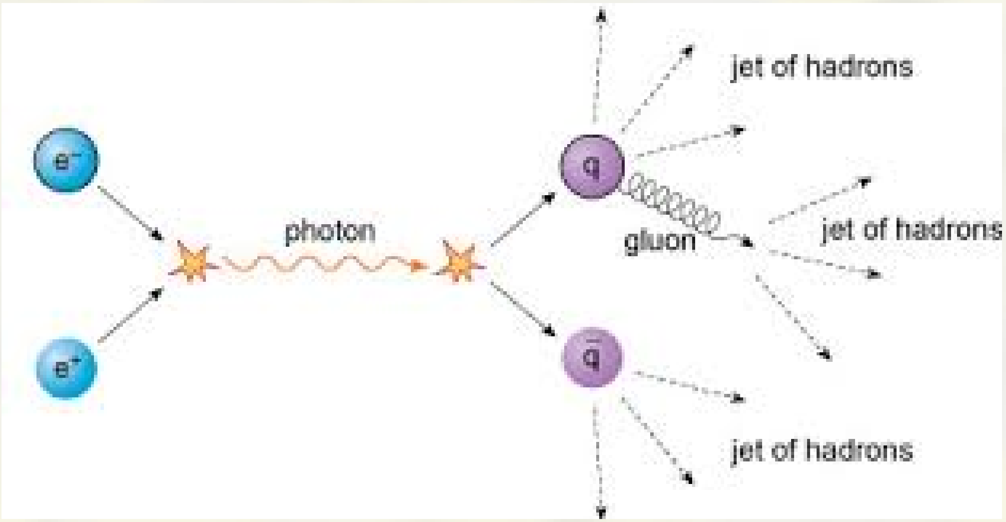
High energy physics applications – early days

1990-1992: L. Lönnblad, T. Rögnvaldsson & C. Peterson developed ANN procedures for jet identification and W mass reconstrution based upon simulated data.

<u>Quark/gluon jet separation.</u> Feed an MLP ANN with hadronic 4-momenta of 4 leading hadrons (16 inputs) within identified jets

→ 85-90% generalization

Just a few % below Bayes limit!
State of the art at the time around 75%

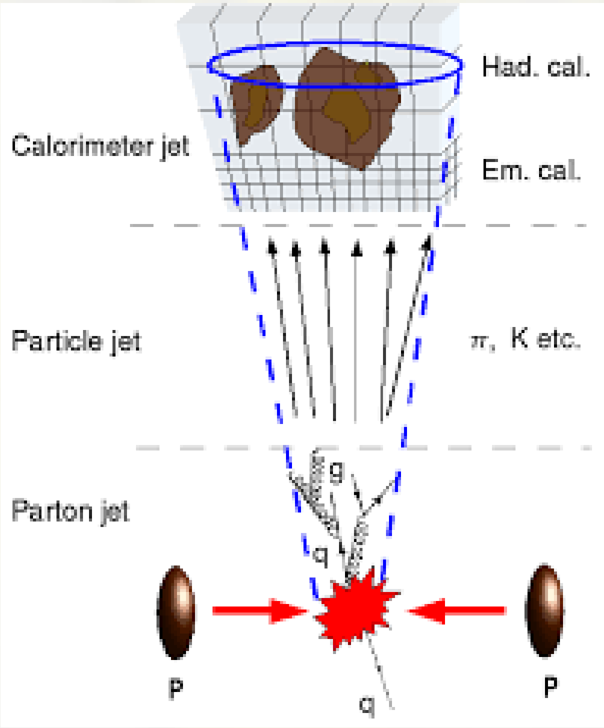


High energy physics applications – early days

W mass reconstruction

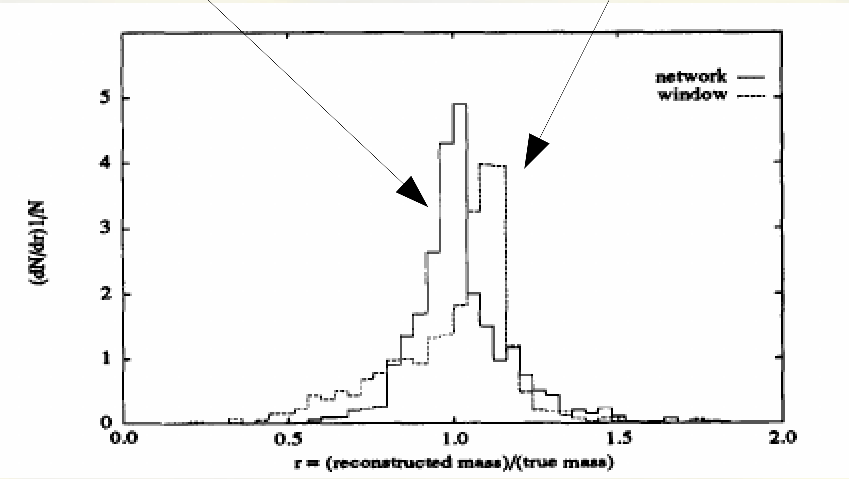
Simulate $p\bar{p} \rightarrow W \rightarrow q\bar{q} \rightarrow$ hadrons

480 cells / 3 hidden layers /
8*8 receptive fields with weight sharing



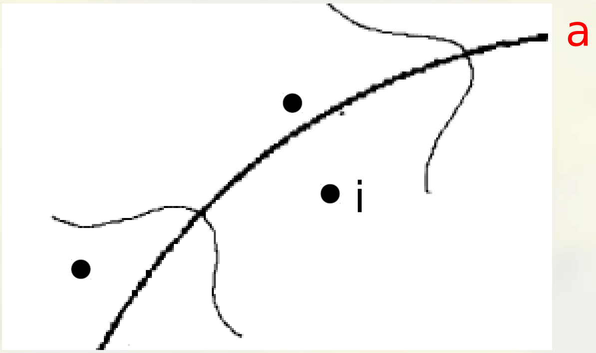
ANN

"Window" method

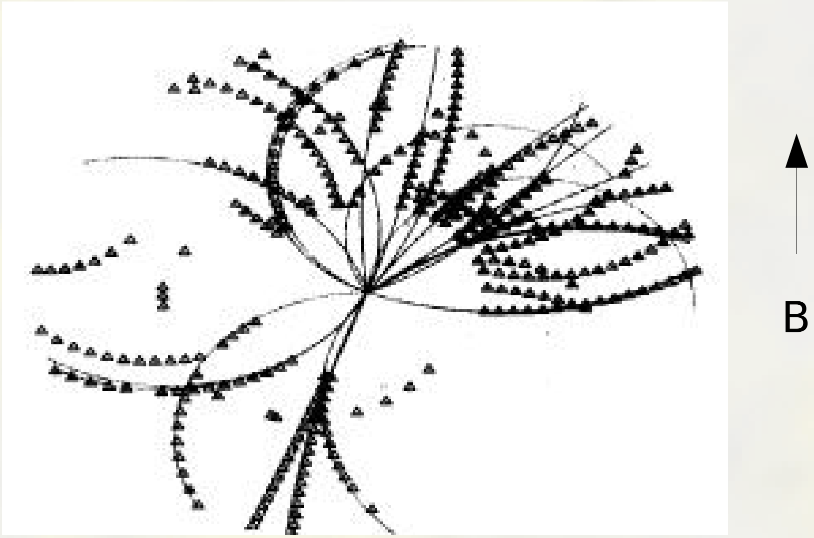


High energy physics applications – early days

1992: M. Ohlsson, C.Peterson and A. Yuille: Deformable templates for tracking



$$E[V_{ia};\ \underline{\theta_a,\ \kappa_a,\ \gamma_a}] = \sum_{i,a} V_{ia} M_{ia} + \lambda \sum_i \left(\sum_a V_{ia} - 1 \right)^2$$



xy-plane

Assign data point i to track a

V_{ia} binary; i belongs to a

M_{ia} = distance between i and a

Minimize with respect to both V_{ia} and M_{ia}

High energy physics applications – early days

Track finding with neural networks
C. Peterson
Nuclear Instruments and Methods **A279**, 537-545 (1989)

Finding gluon jets with a neural trigger
L. Lönnblad, C. Peterson and T. Rögnvaldsson
Physical Review Letters **65**, 1321-1324 (1990)

Using neural networks to identify jets
L. Lönnblad, C. Peterson and T. Rögnvaldsson
Nuclear Physics **B 349**, 675-702 (1991)

Self-organizing networks for extracting jet features
L. Lönnblad, C. Peterson, H. Pi and T. Rögnvaldsson
Computer Physics Communications **67**, 193-209 (1991)

Mass reconstruction with a neural network
L. Lönnblad, C. Peterson and T. Rögnvaldsson
Physics Letters **B 278**, 181-186 (1992)

Track finding with deformable templates - the elastic arms approach
M. Ohlsson, C. Peterson and A. L. Yuille
Computer Physics Communications **71**, 77-98 (1992)

An introduction to artificial neural networks
C. Peterson and T. Rögnvaldsson
Proc. 1991 CERN Summer School of Computing
CERN Yellow Report **92-02**, 113-170 (1992)

JETNET 3.0 - A versatile artificial neural network package
L. Lönnblad, C. Peterson, and T. Rögnvaldsson
Computer Physics Communications **81**, 185-220 (1994)

Abandoned ship? Not really

Planning for LHC required attention

Dispersed to other challenges

Theme of 1991CERN School of Computing

The field is now very much alive

Yet to further explore/develop

ANN (CNN) as an generic function approximator

Map known physics models onto ANN

Profitable with computationally demanding models
In particular those that need to be recalculated in real time (e.g. control, accelerators)

Capture rare events

Quantum MC template wave functions

Climate grid models

To increase resolution substantial speedup called for

In each slab approx 400 equations