The discoveries and inventions of John Hopfield and Geoffrey Hinton:

Hopfield network and Boltzmann machine

Sergey N. Dorogovtsev

UA & I3N

Scientific Background to the Nobel Prize in Physics 2024

"FOR FOUNDATIONAL DISCOVERIES AND INVENTIONS THAT ENABLE MACHINE LEARNING WITH ARTIFICIAL NEURAL NETWORKS"

The Nobel Committee for Physics

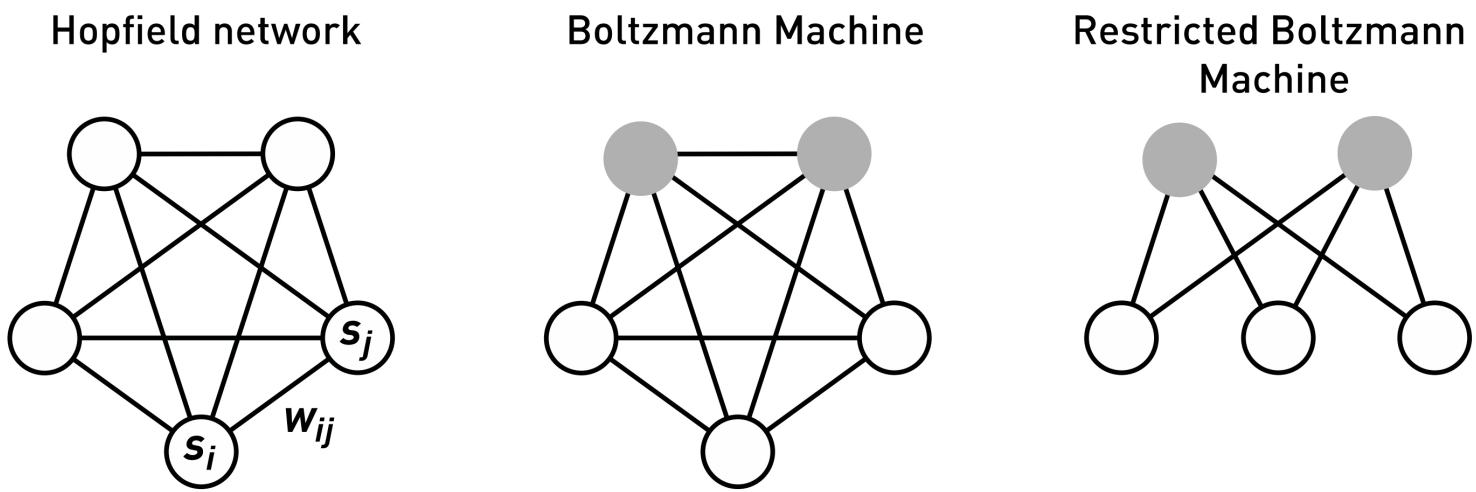


Figure 1. *Recurrent networks of N binary nodes s_i (0 or 1), with connection weights w_{ij}. (Left) The Hopfield model. (Centre) Boltzmann machine. The nodes are divided into two groups, visible (open circles) and hidden (grey) nodes. The network is trained to approximate the probability distribution of a given set of visible patterns. Once trained, the network can be used to generate new instances from the learned distribution. (Right) Restricted Boltzmann Machine (RBM). Same as the Boltzmann machine, but without any couplings within the visible layer or between hidden nodes. This variant can be used for layer-by-layer pre-training of deep networks.*

John Joseph Hopfield (1933)

J. J. Hopfield (1982). Neural networks and physical systems with emergent collective computational abilities.

Geoffrey Everest Hinton (1947)

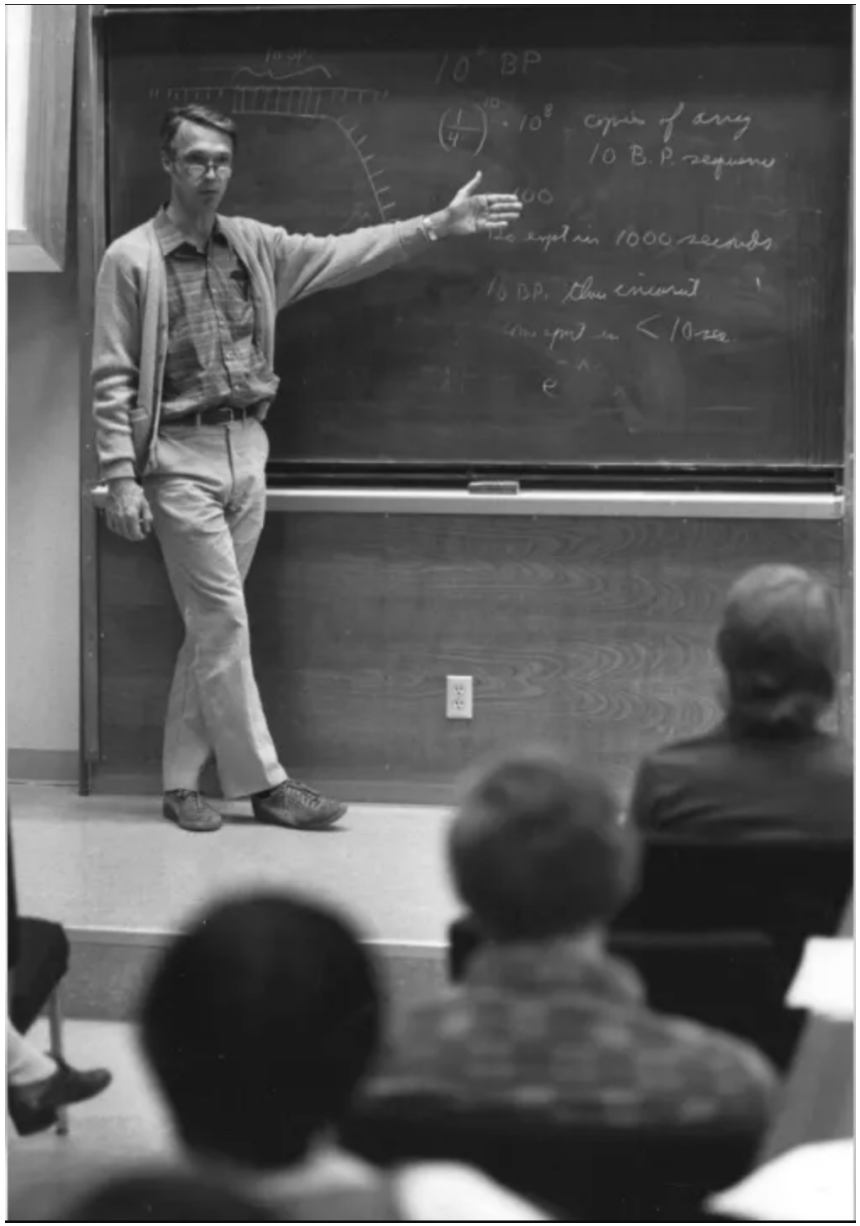
G. E. Hinton and T. J. Sejnowski (1983a). Optimal perceptual inference. (1983b). Analyzing cooperative computation.

S. E. Fahlman, G. E. Hinton, and T. J. Sejnowski (1983). Massively parallel architectures for AI: NETL, Thistle, and Boltzmann machines.

G. E. Hinton, T. J. Sejnowski, and D. H. Ackley (1984). Boltzmann machines: Constraint satisfaction networks that learn.

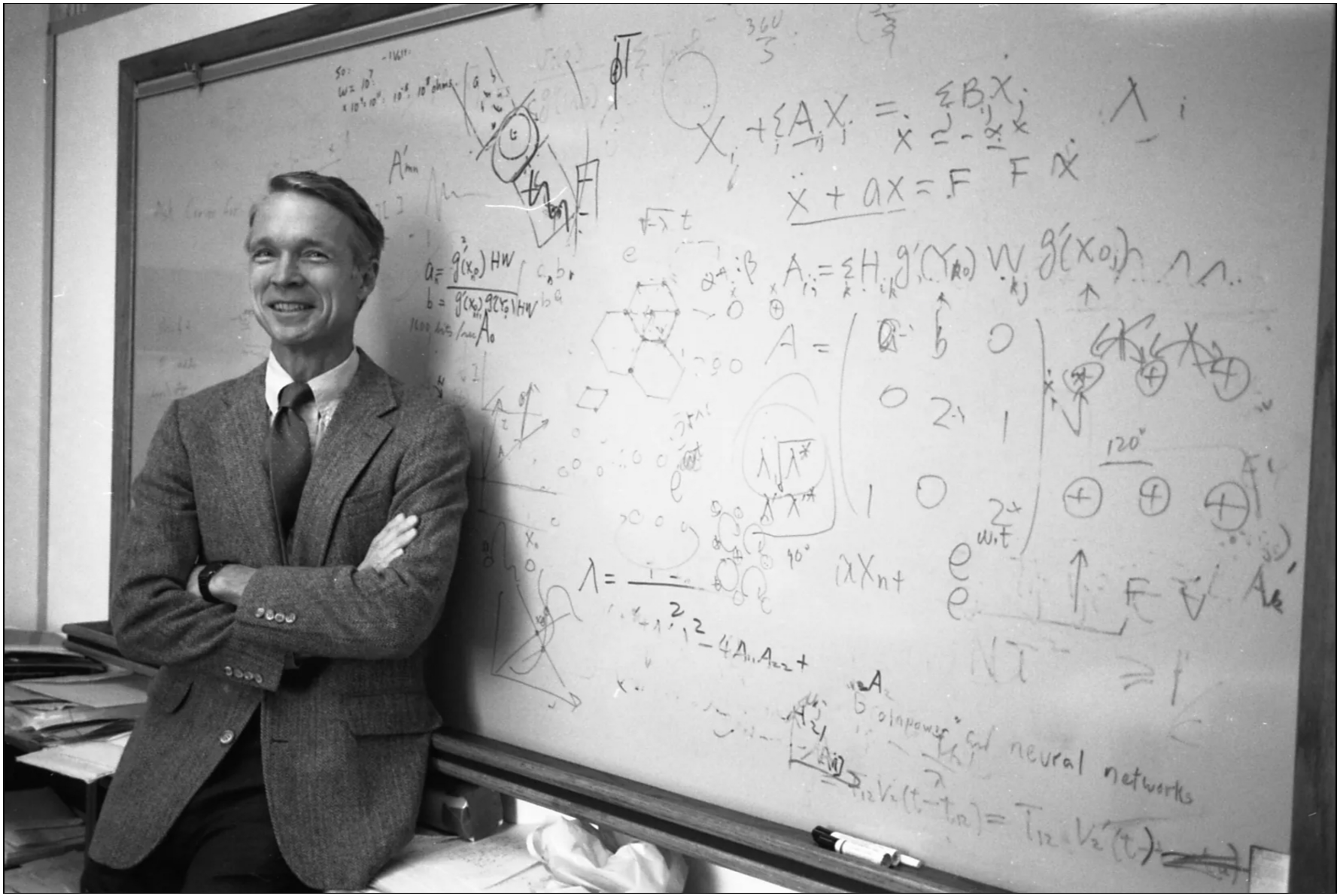
D. H. Ackley, G. E. Hinton, and T. J. Sejnowski (1985). A learning algorithm for Boltzmann machines.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams (1986). Learning representations by back-propagating errors.



John Hopfield, 1981

(studentaffairs.caltech.edu)



(pasadenastarnews.com)

John Hopfield, 1988

(studentaffairs.caltech.edu)

John Hopfield

(reddit.com/r/artificial)

Terrence Sejnowski and Geoffrey Hinton, 1980

Terrence Sejnowski (1947)

MA in physics from Princeton (adv. John Wheeler)
PhD in physics from Princeton (adv. John Hopfield)

John Hopfield $\Longrightarrow$ Terry Sejnowski

John Wheeler $\Longrightarrow$ Richard Feynman, Terry Sejnowski

(Bloomberg Businessweek)

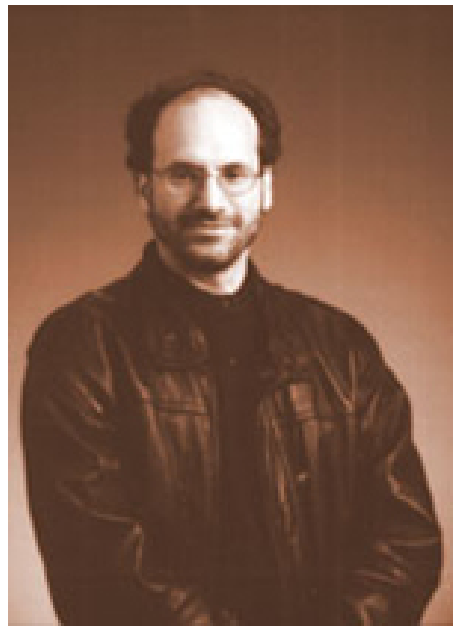
Geoffrey Hinton, 1986

(Bloomberg Businessweek)

Geoffrey Hinton, 1986

(news.stanford.edu)

David Rumelhart (1942–2011)



(cognitivesciencesociety.org)

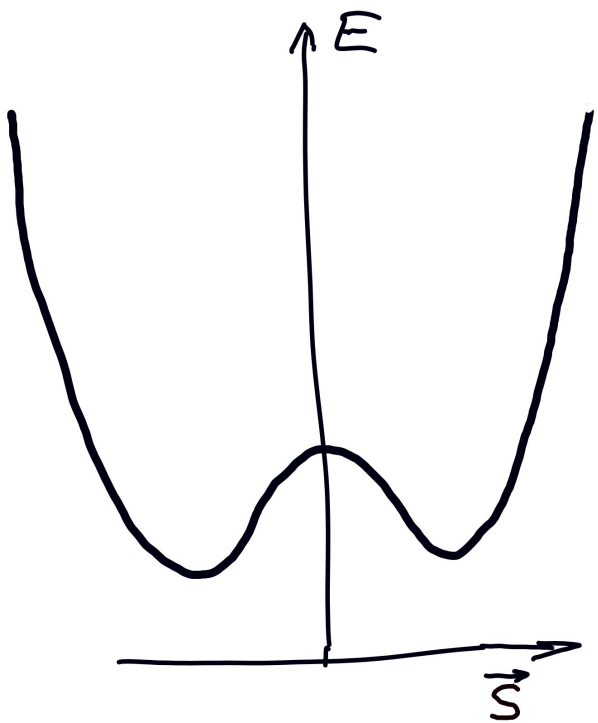
Paul Smolensky (1955)

The Ising model

$$S_i = \pm 1$$

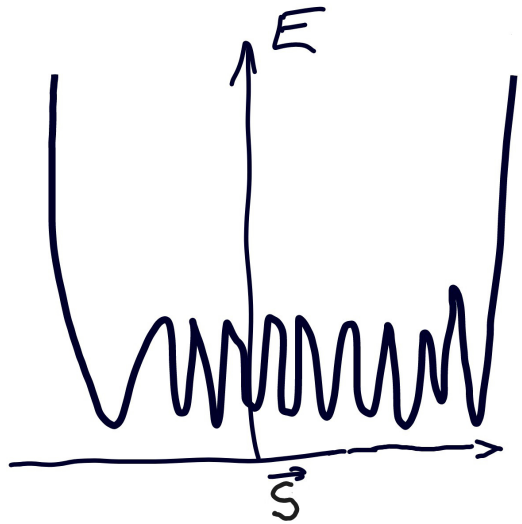
$$E(S_1, ..., S_N) = -\sum_{i>j} J_{ij} S_i S_j - \sum_i H_i S_i = -\frac{1}{2} \mathbf{S}^T \widehat{J} \mathbf{S} - \mathbf{H}^T \mathbf{S}$$

The ferromagnetic model, $J_{ij} > 0$:



Spin glasses

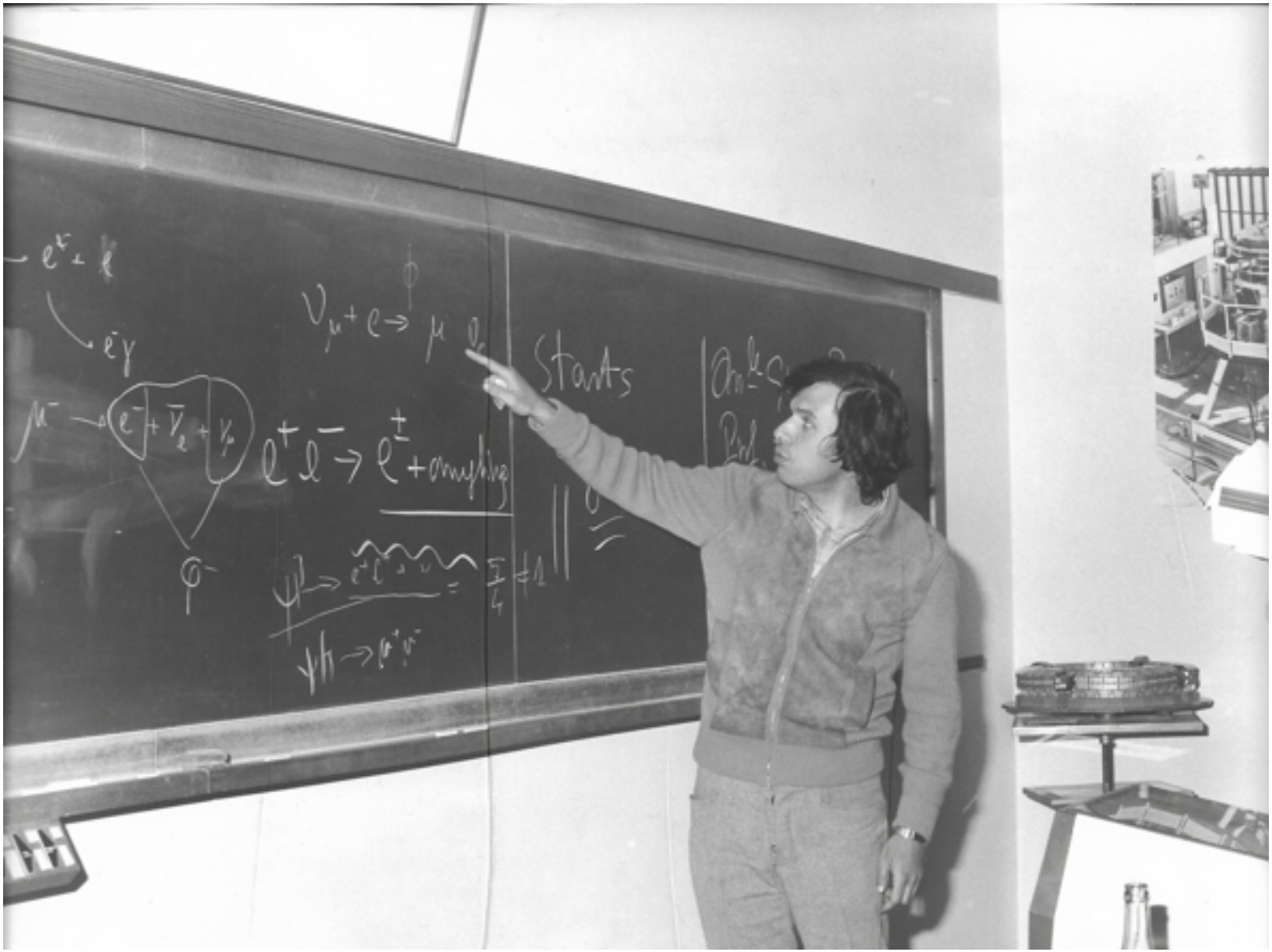
Mixture of positive and negative couplings:



The Sherrington–Kirkpatrick model (1975) was solved exactly (Parisi, 1979).

$$N_{\text{loc. min.}} \sim e^{\text{const}\, N}$$

Kirkpatrick, Gelatt Jr, and Vecchi (1983). Optimization by simulated annealing.

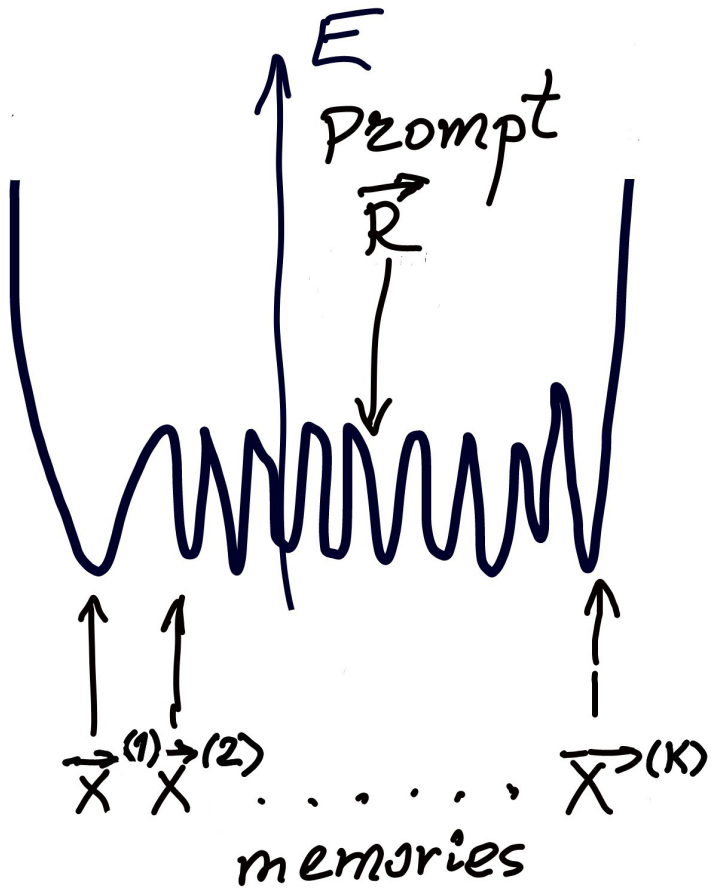


(corriere.it)

Giorgio Parisi

Associative memory

I. A. Richards (1924). *The Principles of Literary Criticism*



Hopfield network

$$E(S_1, ..., S_N) = -\frac{1}{2} \sum_{i,j=1}^{N} J_{ij} S_i S_j - \sum_i H_i S_i$$

The update rule:

$$S_i' = \text{sign}\left(\sum_j J_{ij} S_j + H_i \right)$$

Memories: $\mathbf{X}^{(\alpha)} = (X_1^{(\alpha)}, \ldots, X_N^{(\alpha)})^T$, $\alpha = 1, \ldots, K$

The Hebbian learning rule:

$$J_{ij} = \sum_{\alpha=1}^{K} X_i^{(\alpha)} X_j^{(\alpha)} \quad \text{if} \quad i \neq j$$

Hebb's law of association: the simultaneous activation of neurons strengthens their synaptic connections.

Hopfield network

$$E(\mathbf{S}; \widehat{J})$$

$$\frac{\partial}{\partial \widehat{J}} E(\mathbf{X}^{(\alpha)}; \widehat{J}) = 0, \qquad \alpha = 1, \ldots, K$$

$$\mathbf{R} \xrightarrow{\text{'gradient descent'}} \mathbf{X}^{(\beta)}$$

Only the coordinates of minima matter.

The basins of attraction are sufficiently large to repair an amount of errors of order N.

$$\text{storage capacity:} \qquad K_{\max} \approx 0.14N$$

When $K > K_{\max}$, spurious local minima emerge.

Hopfield network

$$E = -\frac{1}{2}\sum_{\alpha=1}^{K}\left(\sum_{j=1}^{N} X_j^{(\alpha)} S_j\right)^2 + \text{const}$$

$$S_i' = \text{sign}\left[\sum_{\alpha=1}^{K}\left[\left(X_i^{(\alpha)} + \sum_{j\neq i} X_j^{(\alpha)} S_j\right)^2 - \left(-X_i^{(\alpha)} + \sum_{j\neq i} X_j^{(\alpha)} S_j\right)^2\right]\right]$$

Modern Hopfield network

Krotov and Hopfield (2016): dense associative memory

$$E = -\frac{1}{2}\sum_{\alpha=1}^{K} F\left(\sum_{j=1}^{N} X_j^{(\alpha)} S_j\right)$$

$$S_i' = \text{sign}\left[\sum_{\alpha=1}^{K}\left[F\left(X_i^{(\alpha)} + \sum_{j\neq i} X_j^{(\alpha)} S_j\right) - F\left(-X_i^{(\alpha)} + \sum_{j\neq i} X_j^{(\alpha)} S_j\right)\right]\right]$$

Modern Hopfield network

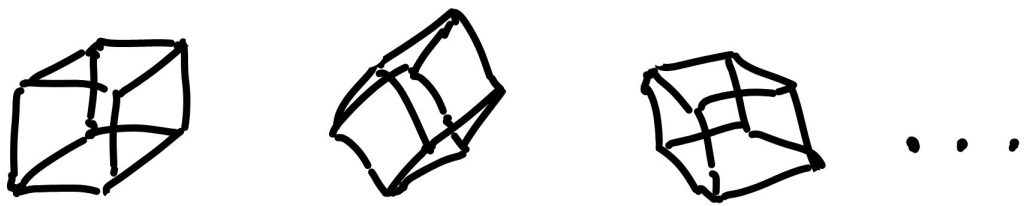
For $F(x) = x^n$, $\quad K_{\max} \sim N^{n-1}$

For $F(x) = e^x$, $\quad K_{\max} \sim e^{\text{const}\, N}$

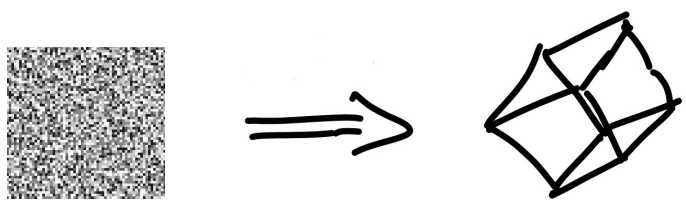
Boltzmann machine

A BM is a stochastic (probabilistic) Hopfield network with hidden units, in which visible units are used both for input and output.

Unsupervised learning (i.e., unlabelled data)

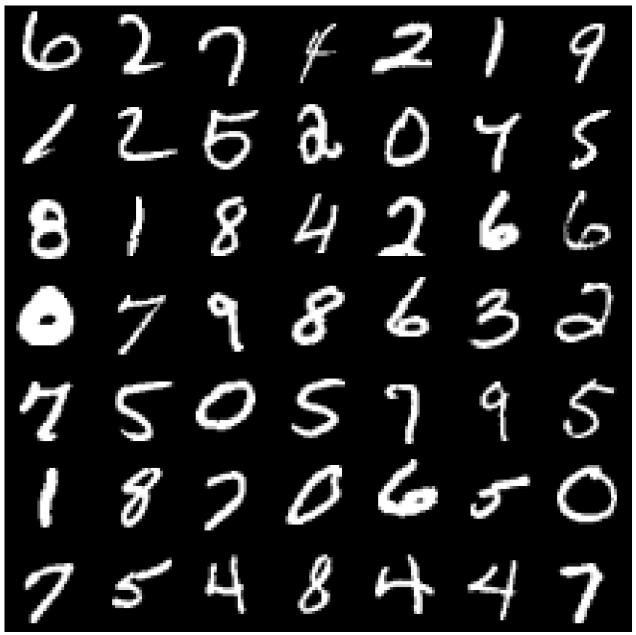


A BM models data—generates images that look like real images



2-layer Boltzmann machine

MNIST digit dataset



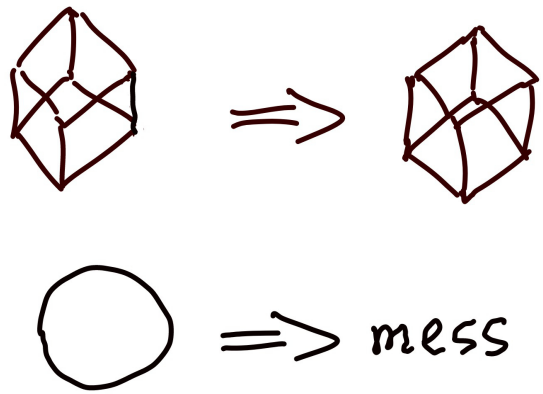
Images the net sees
when it is perceiving the world.

Images that the net generates
when it is dreaming.

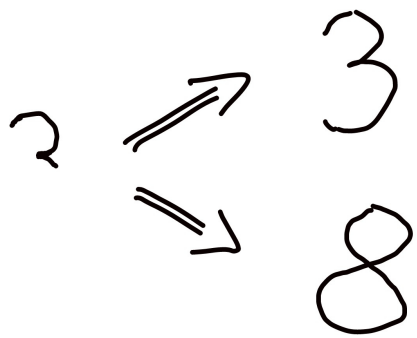
Salakhutdinov and Hinton (2009)

Boltzmann machine

A BM can detect unusual inputs



The completion task



Boltzmann machine

$$\mathbf{s} = (s_1, \ldots, s_N),\ s_i = 0, 1$$

$$\mathbf{s} = \mathbf{v} \cup \mathbf{h}$$

input: $P_{\text{data}}(\mathbf{v}) \Longrightarrow$ output: $P_{\text{out}}(\mathbf{v})$ as close to $P_{\text{data}}(\mathbf{v})$ as is possible

Boltzmann machine

$$E(\mathbf{s}; \widehat{w}, \mathbf{b}) = -\frac{1}{2}\sum_{i,j=1}^{N} w_{ij} s_i s_j - \sum_i b_i s_i$$

Equilibrium, finite temperature

$$P(\mathbf{s}, T) = \frac{1}{\Sigma} e^{-E(\mathbf{s})/T}$$

Consecutively update randomly chosen units:

$$\begin{aligned} \text{prob}(s_i = 1) &= \frac{e^{-E(s_i=1)/T}}{e^{-E(s_i=1)/T} + e^{-E(s_i=0)/T}} \\ &= \frac{1}{1 + e^{-[E(s_i=1)-E(s_i=0)]/T}} \end{aligned}$$

Boltzmann machine

$$P_{\text{out}}(\mathbf{v}) = \sum_{\mathbf{h}} P(\mathbf{s}) \qquad (\mathbf{s} = \mathbf{v} \cup \mathbf{h})$$

Unsupervised learning:

A BM finds weights and biases such that $P_{\text{out}}(\mathbf{v})$ appears to be as close to $P_{\text{data}}(\mathbf{v})$ as is possible.

In other words:

A BM finds weights and biases that provide a Boltzmann distribution in which the data vectors have high probability.

$$P_{\text{data}}(\mathbf{v}) = \sum_{\mathbf{h}} P_{\text{constrained}}(\mathbf{s}) \qquad (\mathbf{s} = \mathbf{v}_{\text{data}} \cup \mathbf{h})$$

The energies of minima matter—in contrast to a Hopfield network.

Boltzmann machine

'Learning the weights is figuring out how the hidden neurons should be used to model the structure in the images it perceives'.

(Hinton's Nobel Lecture)

$P_{\text{data}}(\mathbf{v})$ is an observed distribution,
$P_{\text{out}}(\mathbf{v})$ is a model distribution.

The Kullback–Leibler divergence (1951):

$$D[\{P_{\text{data}}(\mathbf{v}), P_{\text{out}}(\mathbf{v})\}] \equiv \sum_{\mathbf{v}} P_{\text{data}}(\mathbf{v}) \ln \left[\frac{P_{\text{data}}(\mathbf{v})}{P_{\text{out}}(\mathbf{v})} \right] \geq 0$$

Boltzmann machine

$$\frac{\partial D[\{P_{\text{data}}(\mathbf{v}), P_{\text{out}}(\mathbf{v})\}; \widehat{w}, \mathbf{b}]}{\partial w_{ij}} = -\frac{1}{T}\left(\langle s_i s_j \rangle_{\text{data}} - \langle s_i s_j \rangle_{\text{model}}\right)$$

$$\frac{\partial D[\{P_{\text{data}}(\mathbf{v}), P_{\text{out}}(\mathbf{v})\}; \widehat{w}, \mathbf{b}]}{\partial b_i} = -\frac{1}{T}\left(\langle s_i \rangle_{\text{data}} - \langle s_i \rangle_{\text{model}}\right)$$

$\langle s_i s_j \rangle =$ fraction of time when i and j are active simultaneously.

- $\langle\ \rangle_{\text{data}}$: at the current $\widehat{w}$ and $\mathbf{b}$, the average is over the constrained equilibrium states $\mathbf{s} = \mathbf{v}_{\text{data}} \cup \mathbf{h}$ and then over all input vectors.
- $\langle\ \rangle_{\text{model}}$: at the current $\widehat{w}$ and $\mathbf{b}$, the average is over the equilibrium states $\mathbf{s}$ without any constraint.

Boltzmann machine

Gradient descent:

$$\frac{\partial x}{\partial t} = -\gamma \frac{\partial E(x)}{\partial x}$$

$$x_{t+1} = x_t - \gamma \frac{\partial E(x = x_t)}{\partial x}$$

$$\Delta w_{ij} \propto \langle s_i s_j \rangle_{\text{data}} - \langle s_i s_j \rangle_{\text{model}}$$

$$\Delta b_i \propto \langle s_i \rangle_{\text{data}} - \langle s_i \rangle_{\text{model}}$$

In simple terms:

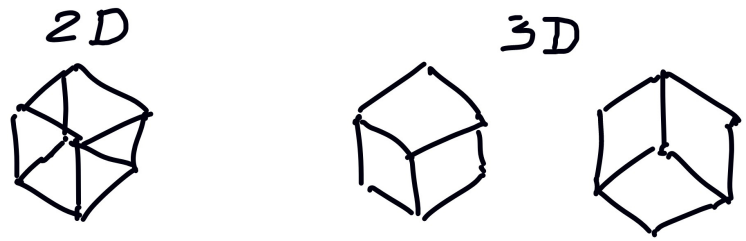
The first term is analogous to the storage term (the Hebbian learning rule) in a Hopfield network,

the second term is like the term for getting rid of spurious minima.

Boltzmann machine

- Highly distributed: every weight and bias is determined by the full set of input vectors, and they determine the entire energy landscape.
- A Boltzmann machine without hidden units — fails.
- Hidden units—latent variables—represent the features = interpretations = 'explanations' of the input.

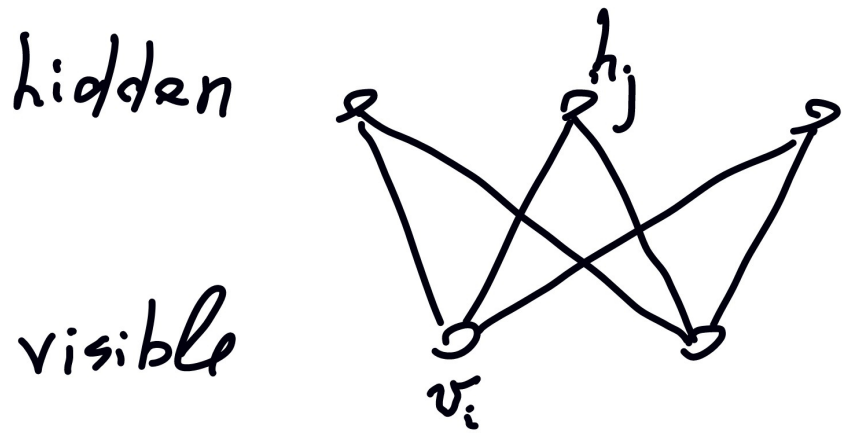
'The energy of a configuration represents the badness of the interpretation'. (Hinton's Nobel lecture) There can be different interpretations:



- Number of hidden units? — Overfitting?
- The original algorithm is ultimately slow.

Restricted Boltzmann machine

Paul Smolensky (1986)



- Parallel update
- If visible units are clamped, i.e., $\mathbf{v} = \mathbf{v}_{\text{data}}$, then the hidden units reach equilibrium in one step,

and so the value $\langle v_i h_j \rangle_{\text{data}}$ can be found in one step.

Restricted Boltzmann machine

$$\underline{\text{visible}} \rightarrow \underline{\text{hidden}} \rightarrow \underline{\text{visible}} \rightarrow \underline{\text{hidden}} \rightarrow \ldots$$

$$\langle v_i h_j \rangle_{\text{data}} \equiv \langle v_i h_j \rangle_0 \rightarrow \langle h_j v_i \rangle_{0'} \rightarrow \langle v_i h_j \rangle_1 \rightarrow \ldots \rightarrow \langle v_i h_j \rangle_\infty,$$

where $\langle v_i h_j \rangle_\infty = \langle v_i h_j \rangle_{\text{model}}$.

$$\Delta w_{ij} \propto \langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{model}} = \langle v_i h_j \rangle_0 - \langle v_i h_j \rangle_\infty$$

$\underline{\text{The contrastive divergence algorithm}}$ (Hinton, 2002)

$$\Delta w_{ij} \propto \langle v_i h_j \rangle_0 - \langle v_i h_j \rangle_1$$

Netflix Prize competition.

A front runner team during Oct.–Dec. 2006:

A. Mnih, R. Salakhutdinov, and G. Hinton.

2012 *AlexNet*: A. Krizhevsky, G. Hinton, and I. Sutskever.

the statistical physics of algorithms

Reading and watching

- Scientific Background to the Nobel Prize in Physics 2024.
  https://www.nobelprize.org/uploads/2024/11/advanced-physicsprize2024-3.pdf
- John Hopfield and Geoffrey Hinton (8 Dec. 2024). 2024 Nobel Prize lectures.
  https://www.youtube.com/watch?v=lPIVl5eBPh8
- Dmitry Krotov (2023). Dense associative memory in machine learning.
  https://www.youtube.com/watch?v=K3JOQVflvyM
- G. E. Hinton (2007). Boltzmann machine. *Scholarpedia*, **2**, 1668.
- Geoffrey Hinton (2012). Neural Networks and Machine Learning [Coursera 2013].
  Lecture 11: Hopfield nets and Boltzmann machines.
  https://www.youtube.com/watch?v=IP3W7cI01VY
  Lecture 12: Restricted Boltzmann machines (RBMs).
  https://www.youtube.com/watch?v=SY7ilsii2YM
- Lenka Zdeborová (2024). Statistical physics of machine learning.
  https://www.youtube.com/watch?v=TLHYwbrhGJc


universidade de aveiro
universidade de aveiro
universidade de aveiro
universidade de aveiro
United Nations Educational, Scientific and Cultural Organization
International Year of Light 2015
i3N
INSTITUTO DE NANOESTRUTURAS, NANOMODELAÇÃO E NANOFABRICAÇÃO
i3N
INSTITUTE FOR NANOSTRUCTURES, NANOMODELLING AND NANOFABRICATION
universidade de aveiro
theoria poiesis praxis
POCH
COMPETE 2020
PROGRAMA OPERACIONAL COMPETITIVIDADE E INTERNACIONALIZAÇÃO
PORTUGAL 2020
UNIÃO EUROPEIA
Fundos Europeus Estruturais e de Investimento
POPH
POTENCIAL HUMANO
QUADRO DE REFERÊNCIA ESTRATÉGICO NACIONAL
PORTUGAL 2007-2013
GOVERNO DE PORTUGAL
UNIÃO EUROPEIA
Fundo Social Europeu
COMPETE
PROGRAMA OPERACIONAL FACTORES DE COMPETITIVIDADE
FCT
Fundação para a Ciência e a Tecnologia
FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA