Mathematics of data science - Course overview

Demetrio Labate

January 16, 2023

What is this course about?

This will be a research-oriented course designed for graduate students with an interest in doing research in **theoretical aspects** of algorithms that aim to extract information from **high-dimensional data.**

The topics of this area of research often lie at the intersection of Mathematics, Statistics, Computer Science and Electrical Engineering.

Traditionally, the analysis of data has been the domain of **signal processing**, whose theoretical underpinning could be found in linear algebra, Fourier analysis and the theory of Hilbert spaces.

What is this course about?

While classical and modern signal analysis was mostly concerned with 1-D (time-series), 2-D (images) and 3-D (videos) signals, emerging applications from medical imaging, electronic surveillance, social networks, etc., typically involve data which are **high-dimensional** and often **non-Euclidean.**

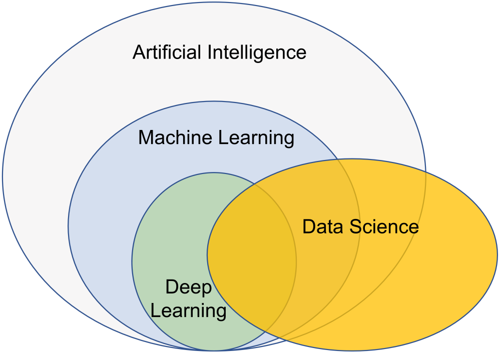
The paradigm shift occurring with the current notion of **'data science'** is the emphasis on the high-dimensionality of data.

This paradigm shifts has led to a new class of algorithms for efficient data representation, dimensionality reduction and feature selection for which the classical formalism of Hilbert spaces of traditional signal processing is often impractical or inadequate.

What is data science?

Machine learning is a discipline that uses computer algorithms and analytics to build predictive models.

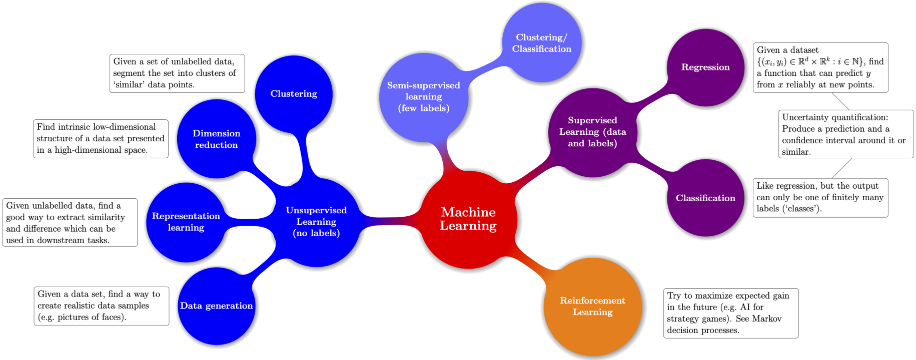
Deep learning is a subset of machine learning that deals with artificial neural networks, a class of algorithms inspired by the structure and function of the human brain.



Artificial intelligence aims to imitate the human brain and create machines that can perform and process tasks intelligently and independently

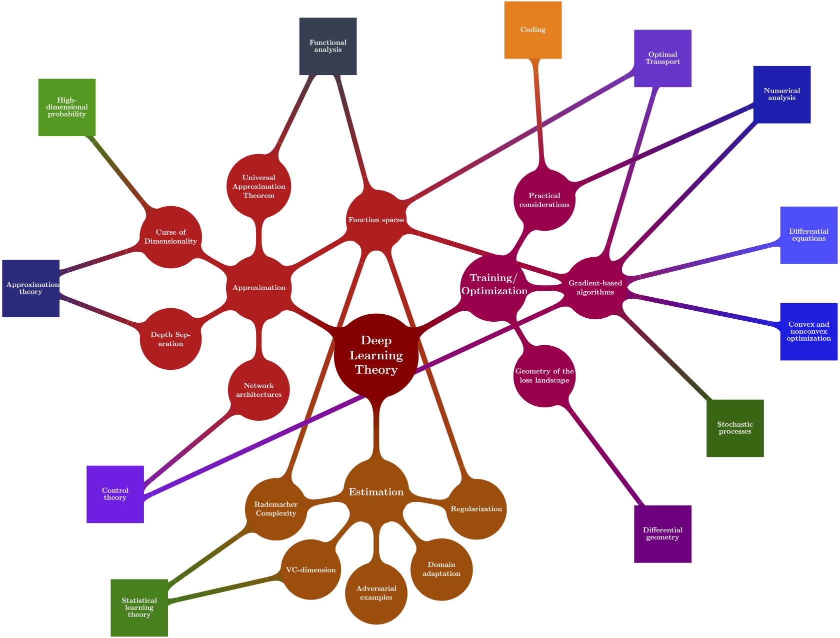
Data science is an inter-disciplinary field that uses scientific methods and algorithms to extract knowledge and insights from structured and unstructured data.

What is machine learning?



(Image by Stephan Wojtowytsch)

What is deep learning?



Some historical notes

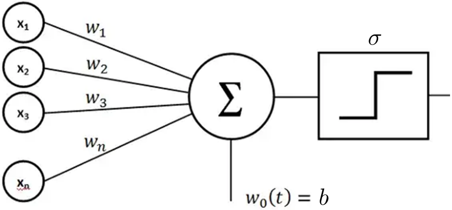
- 1943: The neurophysiologist Warren McCulloch and the mathematician Walter Pitts publish the paper *"A logical calculus of the ideas immanent in nervous activity"* proposing the first mathematical model of an **artificial neuron** with a simple input-output relationship.
  Given inputs $x_1, \ldots, x_n$, the inhibitory input z and a threshold T, the output is

$$y = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} x_n > T \text{ and } z = 0 \\ 0 & \text{otherwise} \end{cases}$$

- 1949: Donald Hebb published *"The Organization of Behaviour"*, proposing a model of synaptic plasticity where (biological) neural pathways strengthen (=adapt/learn) over each successive use.

Some historical notes

- 1958: The psychologist Frank Rosenblatt, inspired by the Hebbian theory of synaptic plasticity, proposed the **perceptron** (originally meant to be a machine rather than a program), a major improvement of the McCulloch-Pitts artificial neuron.

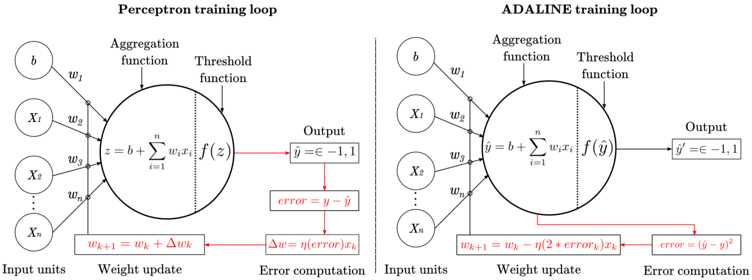


$$y = \sigma(w^t x + b) = \begin{cases} 1 & \text{if } w^t x + b \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

With respect to the McCulloch-Pitts model, the synaptic **weights** w_i need not be unitary or positive. In addition, the neuron takes an extra constant input, a weight b (the **bias**). An algorithm enables the perceptron to learn the synaptic weights from examples to carry out binary classification.

Some historical notes

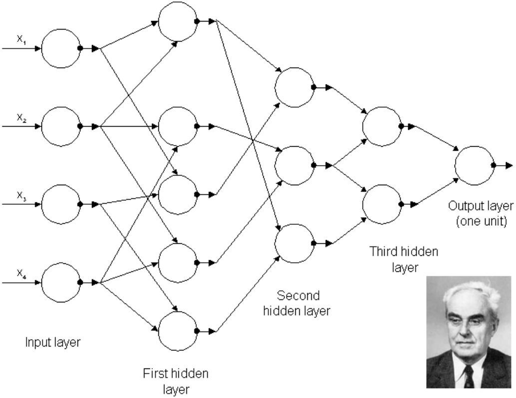
- 1959: Bernard Widrow and Marcian Hoff at Stanford U developed the first neural networks, called ADALINE and MADALINE, applied to real data problems (to remove echoes from a phone line), with the latter one consisting of 3 layers.



- 1964: Vladimir Vapnik and Alexey Chervonenkis invented the original **Support Vector Machine** (SVM) algorithm to solve linear classification problems. Later, in 1992, Boser, Guyon and Vapnik extended the approach to nonlinear classification.

Some historical notes

- 1965: Ivakhnenko and Lapa proposed the first **Multilayer Perceptron**, with polynomial activation functions. In each layer, they selected the best features through statistical methods and forwarded them to the next layer They did not use backpropagation to train their network end-to-end but used layer-by-layer least squares fitting where previous layers were independently fitted from successive layers.



Some historical notes

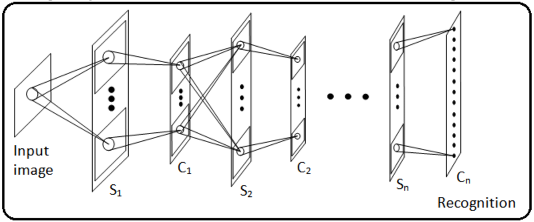
- 1969: Marvin Minsky, founder of the MIT AI Lab, and Seymour Papert, director of the lab, published the book "*Perceptrons*" where they argued that the perceptron approach to neural networks could not be translated effectively into multi-layered neural networks.

  The authors implied (erroneously) that, since a single perceptron is incapable of implementing functions such as the XOR logical function, larger networks would have similar limitations.

  The impact of this publication was so powerful that it dried up funding to an extent that, for the next 10–12 years (the so-called **AI winter**), virtually no research institutions would take on any project about neural networks.

Some historical notes

- 1980: Fukushima introduced the *neocognitron*, a muti-layer neural networt containing **convolutional layers** and downsampling layers, for tasks of pattern recognition.

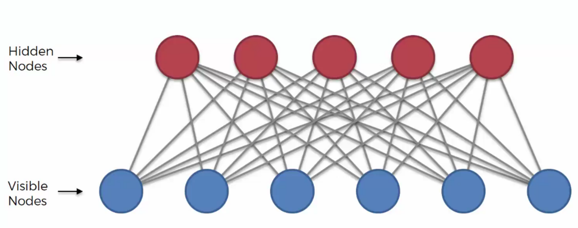
  

  This architecture was inspired by the work of Hubel and Wiesel on the visual cortex ("Receptive fields of single neurons in the cat's striate corte", 1959).

- 1986: Rumelhart, Hinton, and Williams popularized **backpropagation** to train a multilayer neural network. The original theory was derived in the context of control theory by Kelley in 1960 and by Bryson in 1961. In 1974 Werbos first suggested its application to train artificial neural networks.

Some historical notes

- 1986: Paul Smolensky invented the **restricted Boltzmann machine** (RBM), initially called "harmonium". This is a generative stochastic artificial neural network that can learn a probability distribution over its set of inputs.

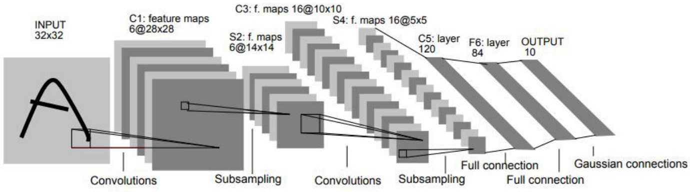


An RBM takes the inputs and translates them to a set of numbers that represents them (forward pass). Then, these numbers are translated back to reconstruct the inputs (backward pass). Through several forward and backward passes, an RBM is trained to reconstruct the input data

In the mid-2000, RBMs rose to prominence after Hinton and collaborators invented fast learning algorithms for them with applications in dimensionality reduction, classification, collaborative filtering and feature learning.

Some historical notes

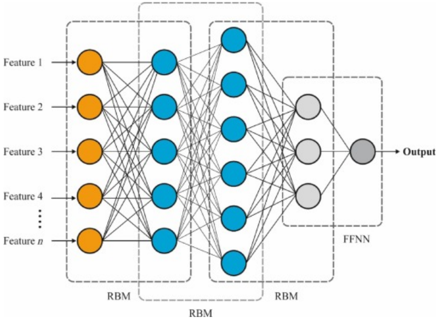
- 1989: LeCun et al. proposed a 5-layer **Convolutional Neural Network** (CNN), called LeNet, trained using backpropagation, for handwriting digit recognition. It was the first CNN architecture that used back-propagation to practical applications.



- 1998: LeCun et al. introduced the (now famous) MNIST dataset and demonstrated that CNNs outperformed all competing models for the task of handwriting digit recognition.

Some historical notes

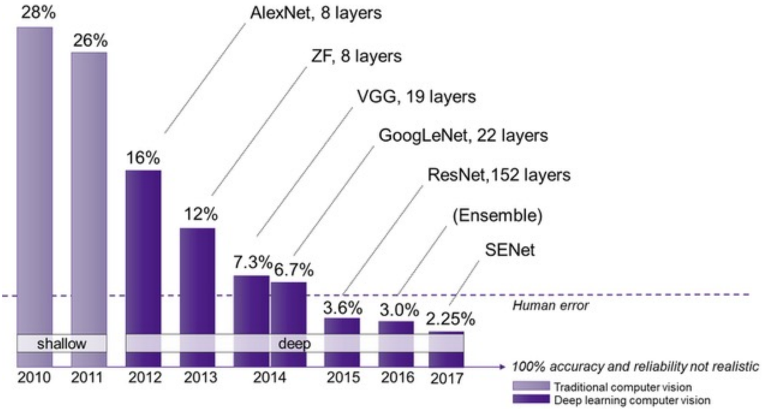
- 2004: Oh and Jung show that standard neural networks can be greatly accelerated on GPUs (20 times faster than CPUs).
- 2006: Hinton, Osindero and Teh introduced **deep belief network** - special multilayer neural networks that can be viewed as a composition of unsupervised networks such as RBMs, where each sub-network's hidden layer serves as the visible layer for the next.



This seminal paper popularized with the notion of **deep learning**.

Some historical notes

- 2012: AlexNet (a CNN) won the ImageNet Large Scale Visual Recognition Challenge, consisting of recognizing about 10,000 object categories from a set of over 10,000,000 images.



In the same year, CNNs were reported to significantly improve on the best performance for multiple image databases.

Topics of the course

The topics covered in this course include:

- Mathematics of signal processing
  - Fourier series
  - Wavelets
  - Scattering transform
- Expressive power of neural networks
  - Universal approximation theorems
  - Shallow vs deep neural networks
  - The manifold assumption
- Statistical learning theory
  - The PAC learning framework
  - Rademacher complexity and VC dimension
- Support Vector Machines (SVM)
  - Linear SVM
  - Nonlinear SVM
- Geometry of high-dimensional data
- Manifold learning

Student evaluation

Student evaluation is based on two assignments: (1) class participation and (2) final project.

1. **Class participation:** Every week, I will assign simple proofs or numerical tests or literature searches.
2. **Final project:** It requires the critical reading of one or more fundamental research paper in an area closely related to topics of the course. I will select the papers in coordination with the students. I will set up several deadlines during the semester to verify the completion of a number of intermediate objectives finalized to the preparation of a written report and a 15-to-20-min in-class presentation.

Student evaluation

Here is a tentative list of topics for the final project.

- Universal approximation theorems
- Barron spaces
- Geometric deep learning
- Graph neural networks
- Deep Belief Networks
- Generative adversarial networks.
- Manifold learning
- Intrinsic dimension
- Sketching
- Wasserstein gradient flows
- Reinforcement learning.
- Deep learning and inverse problems
- Implicit bias of optimization algorithms
- Recurrent neural networks and ODE/PDE