

Introduction to Machine Learning

Humans have long dreamed of creating *machines that can think and act independently*. For many years this has been the aim of **Artificial Intelligence (AI)**. In the early days of AI, many problems that are difficult to solve by humans (e.g., large summations or multiplications, solution of systems of equations) turn out to be easier for computers as long as humans could define a list of tasks that machines could perform at faster speed and higher precisions than humans can do themselves.

On the other hand, tasks that are very easily solved by adult humans and even kids (e.g., recognizing animals in pictures or singing a song) turned out to be very difficult for computers. The main reason of such difficulties lies in the fact that humans cannot explain in words (and with a simple set of instructions) how they have learned to accomplish these tasks. This is where instead the second era of AI solutions, belonging to the field of **Machine Learning (ML)**, have shown astonishing results in the last decade. Instead of relying on hard-coded rules, these algorithms operate in a similar fashion to human beings as they learn from *experience*. In other words, given enough *training data* in the form of inputs (e.g., photos) and outputs (e.g., label of the animal present in the photo), ML algorithms can learn a complex nonlinear mapping between them such that they can infer the output from the input when provided with unseen inputs.

A large variety of ML algorithms have been developed by the scientific community, ranging from the basic *linear* and *logistic regression* that we will see in our fourth lecture, decision tree-based statistical methods such as *random forrest* or *gradient boosting*, all the way to *deep neural networks*, which have recently shown to outperform previously developed algorithms in many fields (e.g., computer science, text analysis and speech recognition, seismic interpretation). This subfield has grown exponentially in the last few years and it is now referred to as **Deep Learning** and will be subject of most of our course. In short, Deep learning is a particular kind of machine learning that represent the world as a nested hierarchy of increasingly complicated concepts the more we move away from the input and towards the output of the associated computational graph. Whilst sharing the same underlying principle of *learning from experience in the form of a training data*, different algorithms presents their own strengths and limitations and a machine learning practitioner must make a careful judgment at any time depending on the problem to be solved.

AI_ML_DL

Terminology

Machine Learning is divided into 3 main categories:

- **Supervised Learning:** learn a function that maps an input to an output ($X \rightarrow Y$). Inputs are also referred to as features and outputs are called targets. In practice we have access to a number of training pairs

$\{\mathbf{x}_i, \mathbf{y}_i\}$ $i = 1, \dots, N$ and we learn $\mathbf{y}_i = f_\theta(\mathbf{x}_i)$ where f_θ is for example parametrized via a neural network. Two main applications of supervised learning are

- *Classification*: the target is discrete
- *Regression*: the target is continuous
- **Unsupervised Learning**: learn patterns from unlabelled data. These methods have been shown to be able to find compact internal representation of the manifold the input data belongs to. Such compact representations can become valuable input features for subsequent tasks of supervised learning. In the context of deep learning, unsupervised models may even attempt to estimate the entire probability distribution of the dataset or how to generate new, independent samples from such distribution. We will get into the mathematical details of these families of models in the second part of our course.
- **Semi-supervised Learning**: it lies in between the other learning paradigms as it learns from some examples that include a target and some that do not.

Input data can also come in 2 different types:

- **Structured data**: tables (e.g., databases)
- **Unstructured data**: images, audio, text, ...

Examples of applications in geoscience are displayed in the figure below.

GEOSCIENTIFIC APPLICATIONS

A number of available data types in various geoscientific contexts is also displayed.

GEOSCIENTIFIC DATA

History

Finally, we take a brief look at the history of Deep Learning. This field has so far experienced three main waves of major development (and periods of success) interspersed by winters (or periods of disbelief):

- **'40 - '50**: first learning algorithms heavily influenced by our understanding of the inner working of the human brain. Mostly linear models such as the McCulloch-Pitts neuron, the perceptron by Rosenblatt, and the adaptive linear element (ADALINE). The latter was trained on an algorithm very similar to Stochastic Gradient Descent (SGD). These models showed poor performance in learning complex functions (e.g., XOR) and led to a drop in popularity of the field.
- **'80 - '90**: these years saw the creation of the Multi Layer Perceptron (MLP), the neocognitron (the ancestor of the convolutional layer), the

first deep neural networks (e.g., LeNet for MNIST classification), the first sequence-to-sequence networks and the LSTM layer.

- **from 2010 till now:** a major moment for the history of this field can be traced back to 2012, when a deep convolution neural network developed by Krizhevsky and co-authors won the ImageNet competition lowering the top-5 error rate from 26.1 percent (previous winning solution not based on a neural network) to 15.3 percent. Since then the field has exploded with advances both in terms of model architectures (AlexNet, VGG, ResNet, GoogleLeNet, ...) optimization algorithms (AdaGrad, RMSProp, Adam, ...), applications (computer vision, text analysis, speech recognition, ...). Moreover, recent developments in the area of unsupervised learning have led to the creation of dimensionality reduction and generative algorithms that can now outperform any state-of-the-art method that is not based on neural networks.

If you want to dig deeper into the history of this field, an interesting read can be found [here](#).