Machine learning and physical (Earth system) modelling - course 1

julien.brajard@nersc.no

October 2021

NERSC

slides+notebook:*https://github.com/nansencenter/nersc_ml_course*

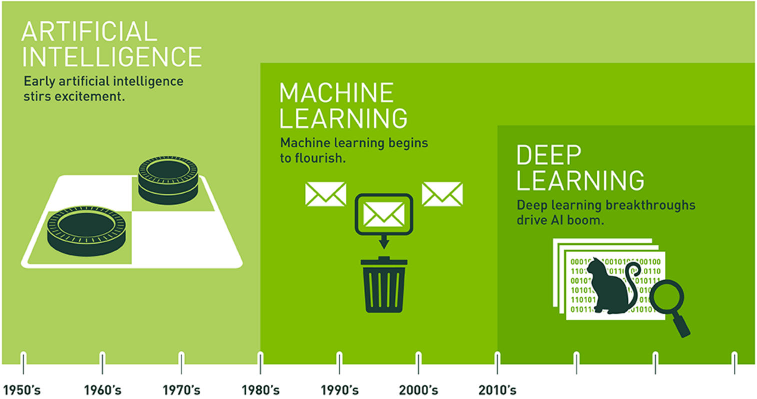
References

- Ian Goodfellow, Yoshua Bengio, and Aaron Courville.
  Deep Learning.
  MIT Press, 2016.
  http://www.deeplearningbook.org.
- Jake VanderPlas.
  Python Data Science Handbook: Essential Tools for Working with Data.
  O'Reilly Media, Inc., 1st edition, 2016.

Table of contents i

Introduction

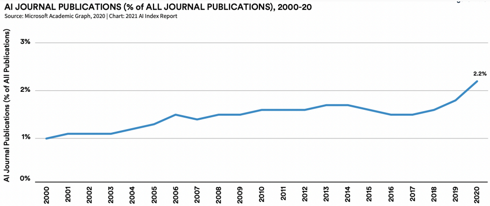
Scope of the lecture: Machine Learning

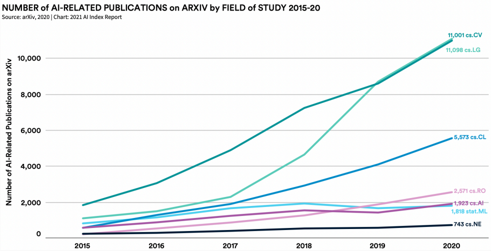


Since an early flush of optimism in the 1950s, smaller subsets of artificial intelligence – first machine learning, then deep learning, a subset of machine learning – have created ever larger disruptions.

Source: NVidia

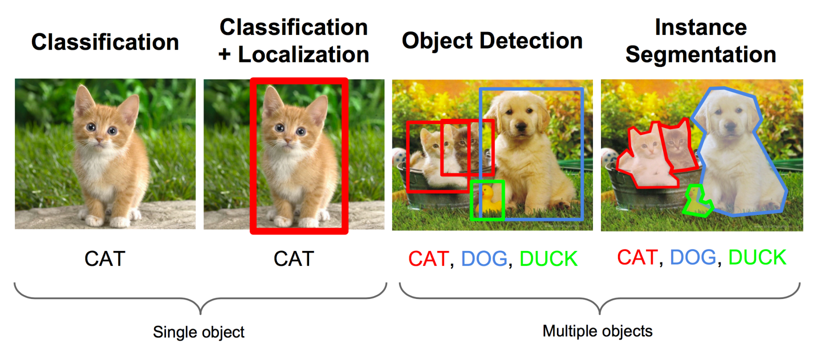
A (very) active field





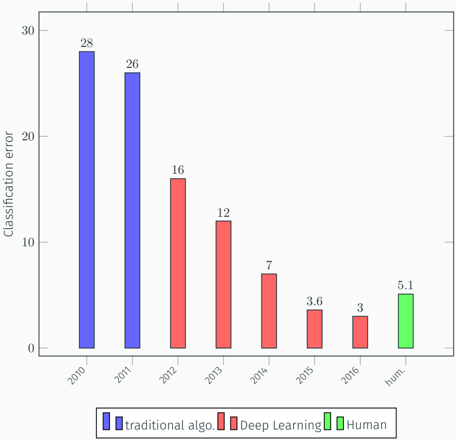
Zhang et al., “The AI Index 2021 Annual Report”

Example 1: Computer Vision



Li, Karpathy and Johnson, 2016, Stanford CS231n course

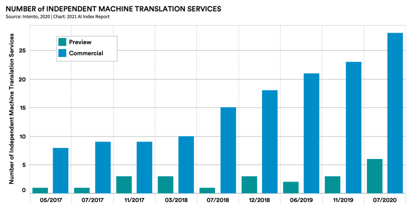
Example 1: Computer Vision



Deep learning architectures were based on Convolutional Neural Networks (CNN).

Example 2: Machine Translation

Objective : translate a text from a language to another.



Zhang et al., "The AI Index 2021 Annual Report"

- **Oct. 2013**: Pionneering scientic paper (Kalchbrenner, N., and Blunsom, P).
- **2016**: Neural machine translation outperform traditional approaches on public benchmarks
- **2017**: Major systems switch to neural machine translation (using deep recurrent neural networks)

Example 3: Playing Games

- **1997**: Deep Blue defeats Kasparov at Chess.
- **2016**: AlphaGo's victory again Lee Sedol at Go.
- **2017**: AphaGo Zero learns how to play Go only by playing against itself. It outperformed previous AlphaGo version (Reinforcement learning)
- **2017**: DeepStack beats professional human poker players.

Example 4: Protein folding

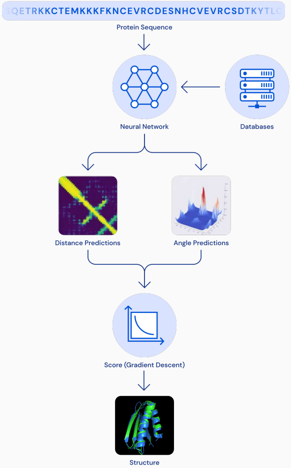


Diagram of Alpha Fold (source: Deepmind)

Senior et al. 2020

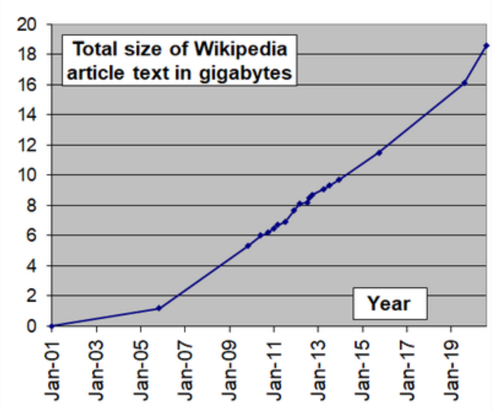
AI Art?

Edmond de Bellamy by Obvious(collective)

Generated using a Generative Adversarial Network.
Selling prince (Oct. 2018): $432,000

Reasons for these recent achievements?

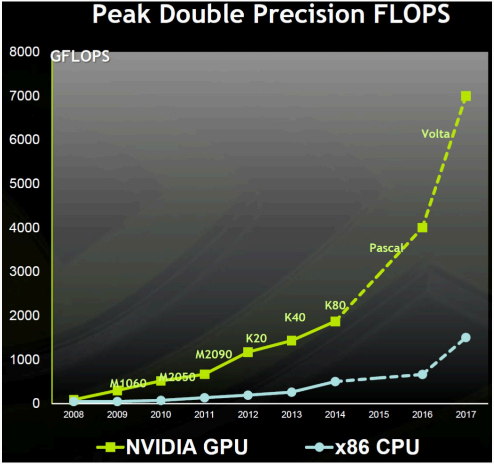
- Increasing of the datasets in size and quality



source: Wikipedia

Reasons for these recent achievements?

- Increasing of the datasets in size and quality
- Progress in computing resources.



source: NVDIA

Reasons for these recent achievements?

- Increasing of the datasets in size and quality
- Progress in computing resources.
- Scientific research on new algorithms (e.g adapted to image processing)

source: Deep Dream Generator

Reasons for these recent achievements?

- Increasing of the datasets in size and quality
- Progress in computing resources.
- Scientific research on new algorithms (e.g adapted to image processing)
- Very efficient software (GPU, cloud computing, automatic differentiation, ...)

TensorFlow Keras PyTorch

Reasons for these recent achievements?

- Increasing of the datasets in size and quality
- Progress in computing resources.
- Scientific research on new algorithms (e.g adapted to image processing)
- Very efficient software (GPU, cloud computing, automatic differentiation, ...)
- Free software and open data culture.

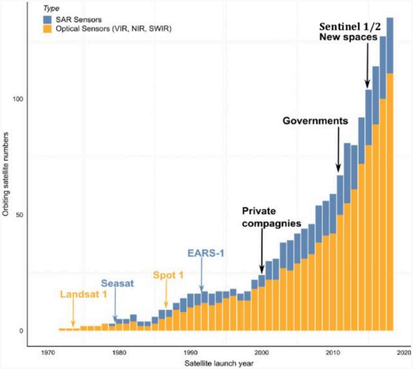
arXiv.org

GitHub

Apply Machine-Learning to physical (Earth-system) modelling?

Why is it a good idea?

- A increasing number of geophysical data (one spatial mission: 24 TB/day)



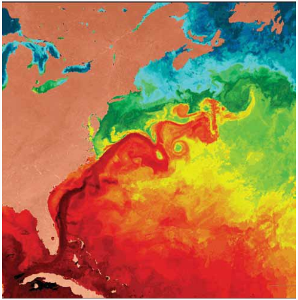
Earth System observavtion satellites

source: Tonneau et al. (2020)

Apply Machine-Learning to physical (Earth-system) modelling?

Why is it a good idea?

- A increasing number of geophysical data (one spatial mission: 24 TB/day)
- Data with highly significant spatial patterns



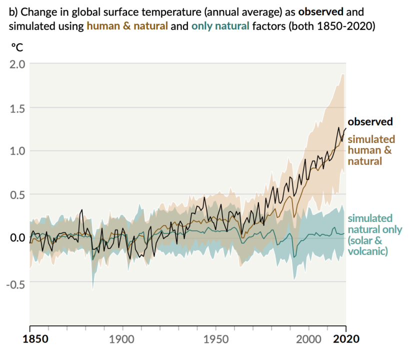
Sea Surface temperature of the gulf stream

source: Talley (2000)

Why is physical modelling specific?

NASDAQ Composite sock market index over the last 10 years

Figure 1: IPCC, AR6, WG1

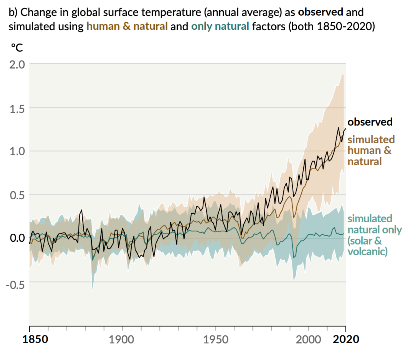


Why is physical modelling specific?

NASDAQ Composite sock market index over the last 10 years

Mostly unknown dynamical processes

Figure 1: IPCC, AR6, WG1



Mostly known dynamical processes (based on physical principles)

What about data assimilation?

Machine learning and data assimilation are closely linked.

Some references:

- Geer, A.J., 2021. Learning earth system models from observations: machine learning or data assimilation?. *Philosophical Transactions of the Royal Society A*, 379(2194)
- Brajard et al. 2019. Connections between data assimilation and machine learning to emulate a numerical model. *Proceedings of the 9th International Workshop on Climate informatics*
- Bocquet et al. 2019. Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models. *Nonlinear processes in geophysics.* 26(3).
- Abarbanel, H.D., Rozdeba, P.J. and Shirman, S., 2018. Machine learning: Deepest learning as statistical data assimilation problems. *Neural computation, 30(8)*.

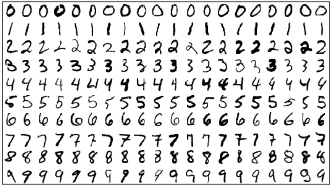
Generalities on Machine Learning

What is this about ?

Can we extract knowledge, make some predictions, determine a "model" using this large amount of data ?

What is this about ?

Can we extract knowledge, make some predictions, determine a "model" using this large amount of data ?

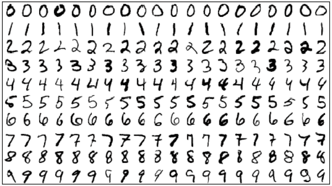


Base of images

$\longrightarrow$ Digit$\in \{0, \ldots, 9\}$

What is this about ?

Can we extract knowledge, make some predictions, determine a "model" using this large amount of data ?



Base of images

$\longrightarrow$ Digit $\in \{0, \ldots, 9\}$

- From high dimensional data (thousands to millions dimensions) to reduced dimensional data (less than 100)
- From disorganized data to comprehensive information
- Can we teach a machine how to do that ?

Two classes of Machine Learning problems

1. Regression: Determination of a quantitative variable from a set of data
   - The price of a building from various predictors (Surface, ...)
   - A physical value (Temperature, humidity, ...) in the future knowing the past
   - ...

Two classes of Machine Learning problems

1. Regression: Determination of a quantitative variable from a set of data
    - The price of a building from various predictors (Surface, ...)
    - A physical value (Temperature, humidity, ...) in the future knowing the past
    - ...
2. Classification: Determination of a class
    - A digit from a image
    - Identification of the content of an image
    - ...

Two types of objectives

1. Supervised learning: we have a set of labeled data with examples of targets.

Two types of objectives

1. Supervised learning: we have a set of labeled data with examples of targets.
2. Unsupervised learning: we only have unlabeled the data, we have no examples of what we want to obtain. We want to extract a ”useful” representation of these data, or some coherent categories.

Two types of objectives

1. Supervised learning: we have a set of labeled data with examples of targets.
2. Unsupervised learning: we only have unlabeled the data, we have no examples of what we want to obtain. We want to extract a ”useful” representation of these data, or some coherent categories.
    - Determine typical behaviors of clients in a supermarket knowing what the have bought.

Two types of objectives

1. Supervised learning: we have a set of labeled data with examples of targets.
2. Unsupervised learning: we only have unlabeled the data, we have no examples of what we want to obtain. We want to extract a ”useful” representation of these data, or some coherent categories.
    - Determine typical behaviors of clients in a supermarket knowing what the have bought.
3. Semi-Supervised Learning: Only a few subset of the data are labeled

Two types of objectives

1. Supervised learning: we have a set of labeled data with examples of targets.
2. Unsupervised learning: we only have unlabeled the data, we have no examples of what we want to obtain. We want to extract a ”useful” representation of these data, or some coherent categories.
    - Determine typical behaviors of clients in a supermarket knowing what the have bought.
3. Semi-Supervised Learning: Only a few subset of the data are labeled
4. Reinforcement Learning: We can initiate and observe the interaction of an agent with its environment. We want to optimize the behavior of the agent.

Two types of objectives

1. Supervised learning: we have a set of labeled data with examples of targets.
2. Unsupervised learning: we only have unlabeled the data, we have no examples of what we want to obtain. We want to extract a ”useful” representation of these data, or some coherent categories.
    - Determine typical behaviors of clients in a supermarket knowing what the have bought.
3. Semi-Supervised Learning: Only a few subset of the data are labeled
4. Reinforcement Learning: We can initiate and observe the interaction of an agent with its environment. We want to optimize the behavior of the agent.
    - Playing a chess game.

Formally

A Machine

$$y = \mathcal{M}(x, \theta)$$

- x: input
- y: output
- $\mathcal{M}$: a model (named ”machine”)
- θ : parameters of the model $\mathcal{M}$.

Machine learning consists in optimizing θ using a set of data. This is the training process.

The Machine Learning recipe

A Machine

$$y = \mathcal{M}(x, \theta)$$

What are the ingredients?

The Machine Learning recipe

A Machine

$$y = \mathcal{M}(x, \theta)$$

What are the ingredients?

- Some data
    - x, y : supervised learning
    - only x: unsupervised learning
    - x and some subset of y: semi-supervised learning

The Machine Learning recipe

A Machine

$$y = \mathcal{M}(x, \theta)$$

What are the ingredients?

- Some data
  - x, y : supervised learning
  - only x: unsupervised learning
  - x and some subset of y: semi-supervised learning
- An objective
  - y is quantitative: regression
  - y is a class: classification

The Machine Learning recipe

A Machine

$$y = \mathcal{M}(x, \theta)$$

What are the ingredients?

- Some data
  - x, y : supervised learning
  - only x: unsupervised learning
  - x and some subset of y: semi-supervised learning
- An objective
  - y is quantitative: regression
  - y is a class: classification
- A computational architecture (the machine)
  - linear
  - non-linear
  - neural networks, random forest, ...

The Machine Learning recipe

A Machine

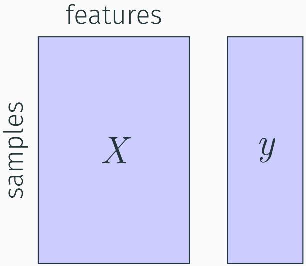
$$y = \mathcal{M}(x, \theta)$$

What are the ingredients?

- Some data
  - x, y : supervised learning
  - only x: unsupervised learning
  - x and some subset of y: semi-supervised learning
- An objective
  - y is quantitative: regression
  - y is a class: classification
- A computational architecture (the machine)
  - linear
  - non-linear
  - neural networks, random forest, ...
- A learning process
  - Estimation of θ

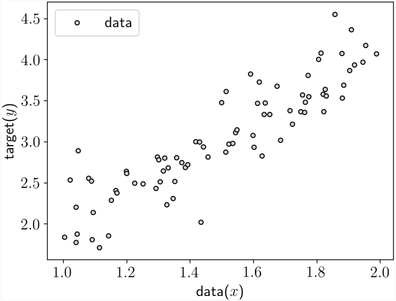
Multidimensional data

Generally, we have multidimensional data X and a one-dimensional target y.



An illustration

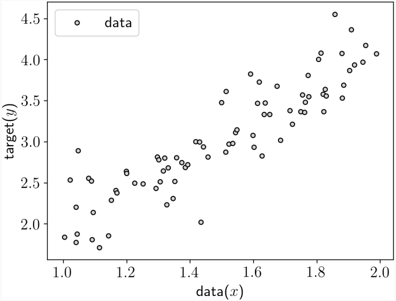
- Some Data



  - there are labeled y: supervised learning
  - y is quantitative: regression

An illustration

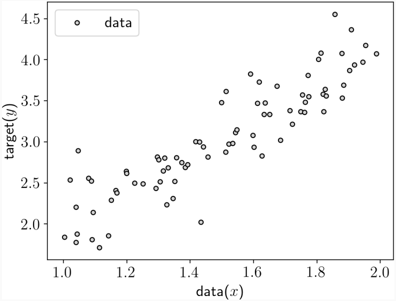
- Some Data



  - there are labeled y: supervised learning
  - y is quantitative: regression
- An Objective: Estimate $\hat{y}$ from x by minimizing $(\hat{y} - y)^2$ (Least-square objective)

An illustration

- Some Data

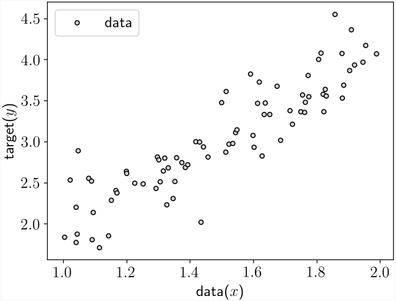


  - there are labeled y: supervised learning
  - y is quantitative: regression
- An Objective: Estimate $\hat{y}$ from x by minimizing $(\hat{y} - y)^2$ (Least-square objective)

- A model: $\hat{y} = \theta_1 X + \theta_0$ (linear)

An illustration

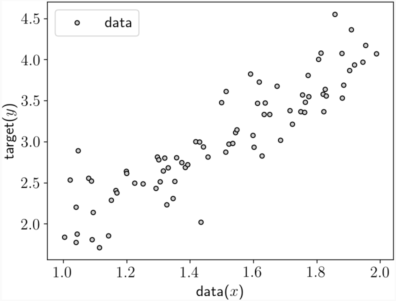
- Some Data



  - there are labeled y: supervised learning
  - y is quantitative: regression
- An Objective: Estimate $\hat{y}$ from x by minimizing $(\hat{y} - y)^2$ (Least-square objective)
- A model: $\hat{y} = \theta_1 X + \theta_0$ (linear)
- A learning process: $\theta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y$

An illustration

- Some Data



  - there are labeled y: supervised learning
  - y is quantitative: regression
- An Objective: Estimate $\hat{y}$ from x by minimizing $(\hat{y} - y)^2$ (Least-square objective)

- A model: $\hat{y} = \theta_1 X + \theta_0$ (linear)
- A learning process: $\theta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T y$

Result