Lecture 1:

AI and Deep Learning

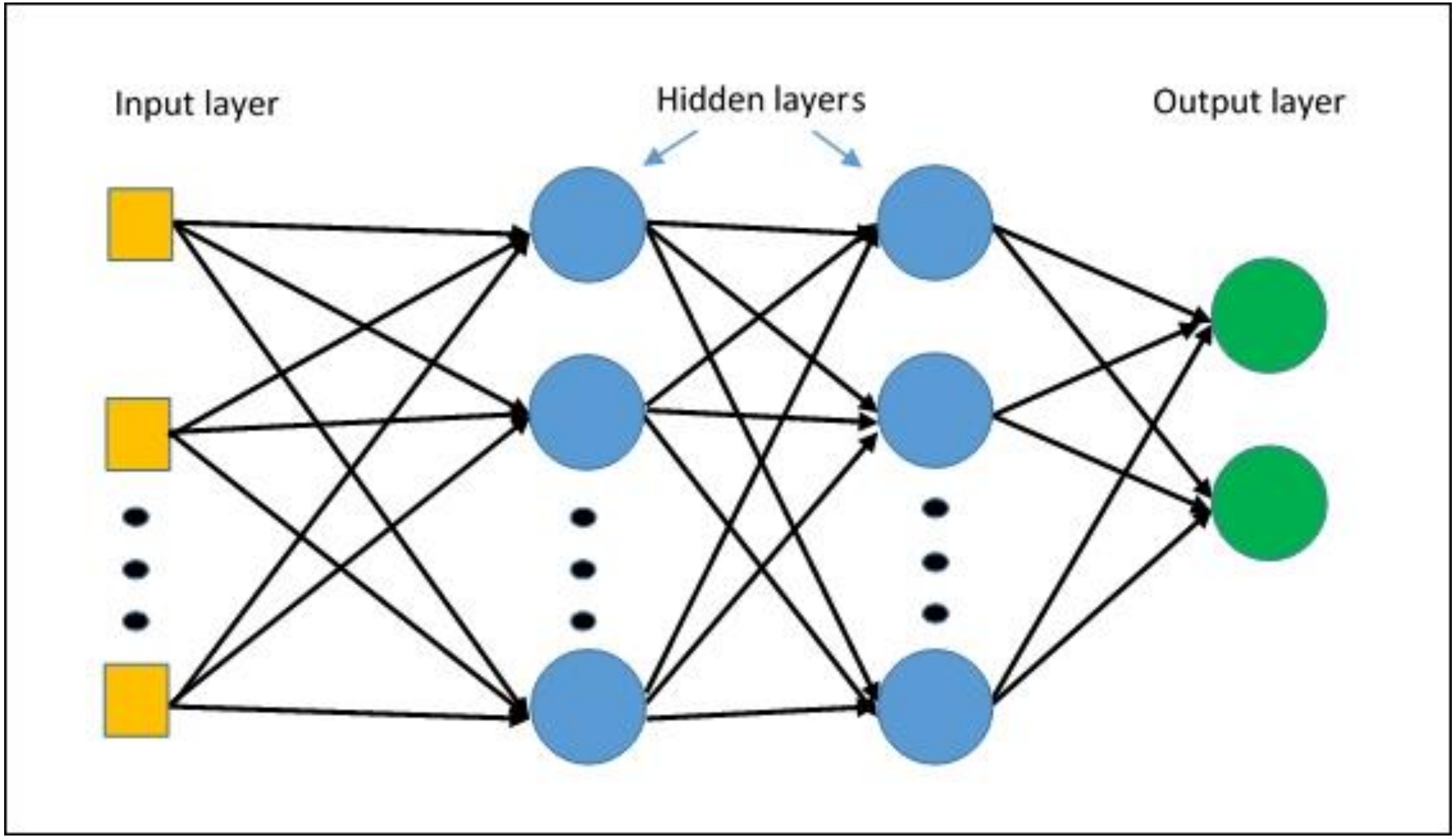
Simone Melzi, Marco tarini

Milano, 13/09/2021

PhD School – Learning on 3D geometries

LA STATALE Università degli Studi di Milano
SAPIENZA Università di Roma

What



We will have a **deeper** description …

Image from "Getting Started with TensorFlow", G. Zaccone, 2016

When

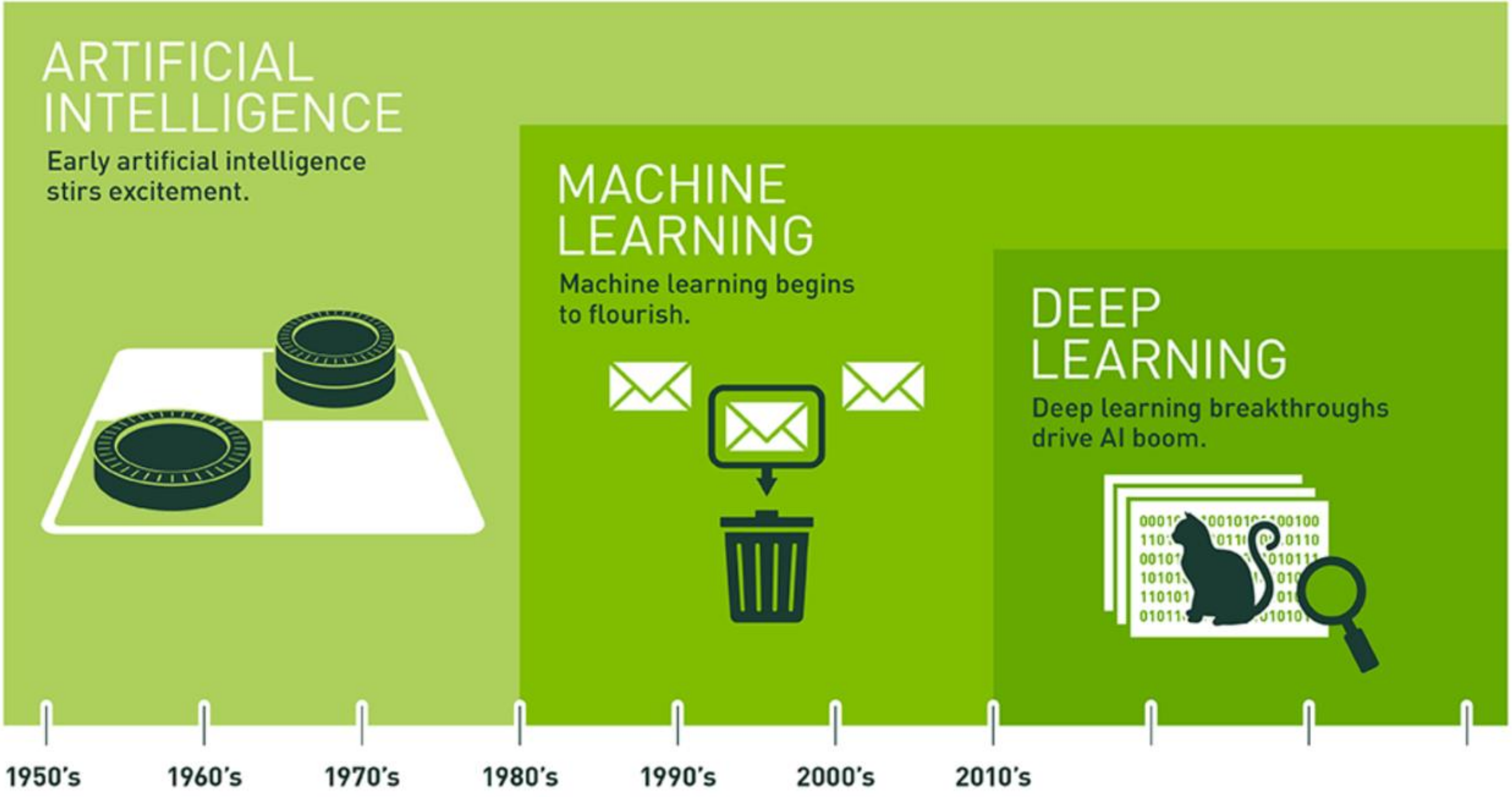
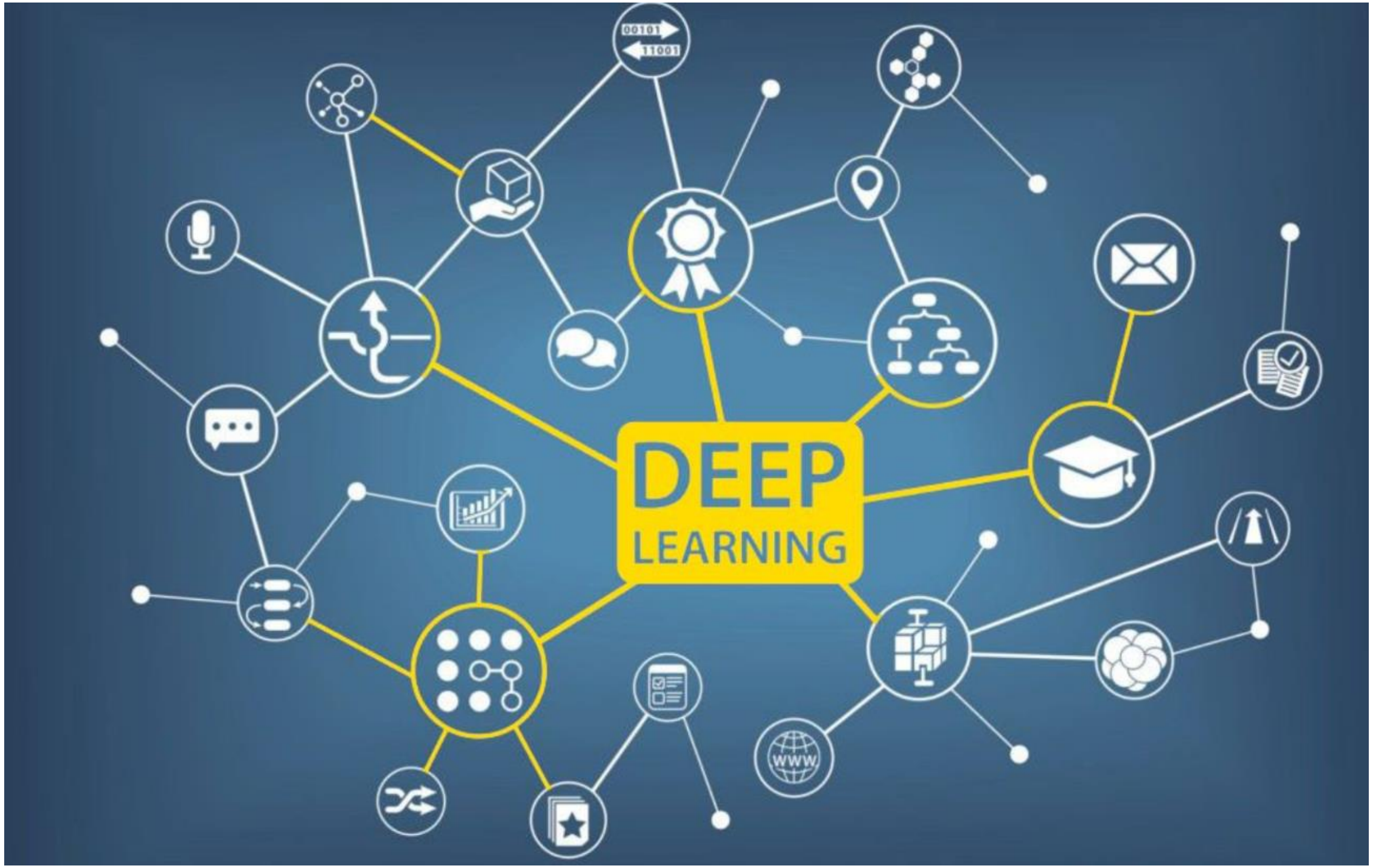


Image by Michael Copeland, NVIDIA

Slide credit E. *Rodolà*

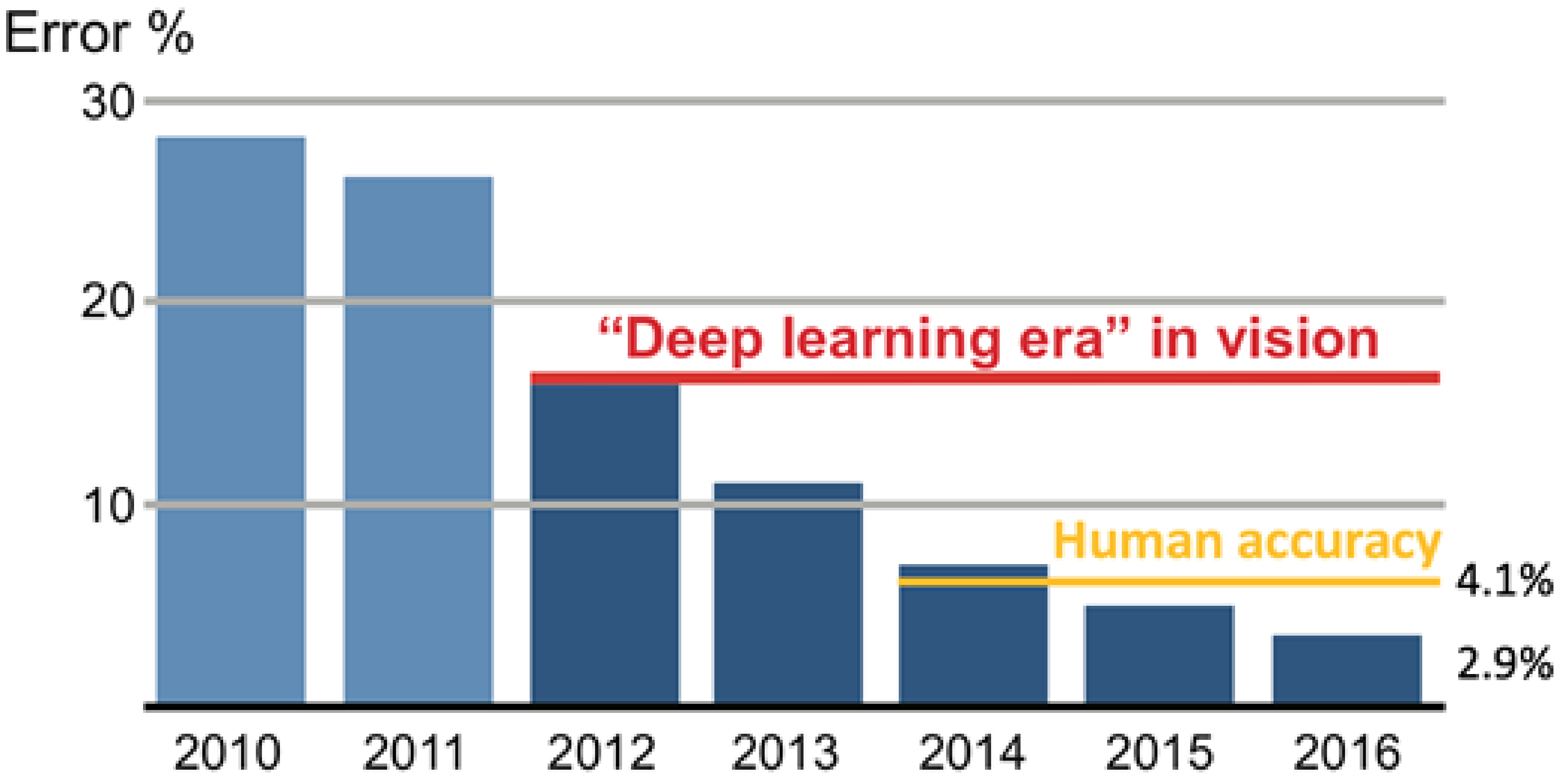
Where



Nowadays, Deep Learning and AI are everywhere

Slide credit E. *Rodolà*

Why



ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

Slide credit E. *Rodolà*

How

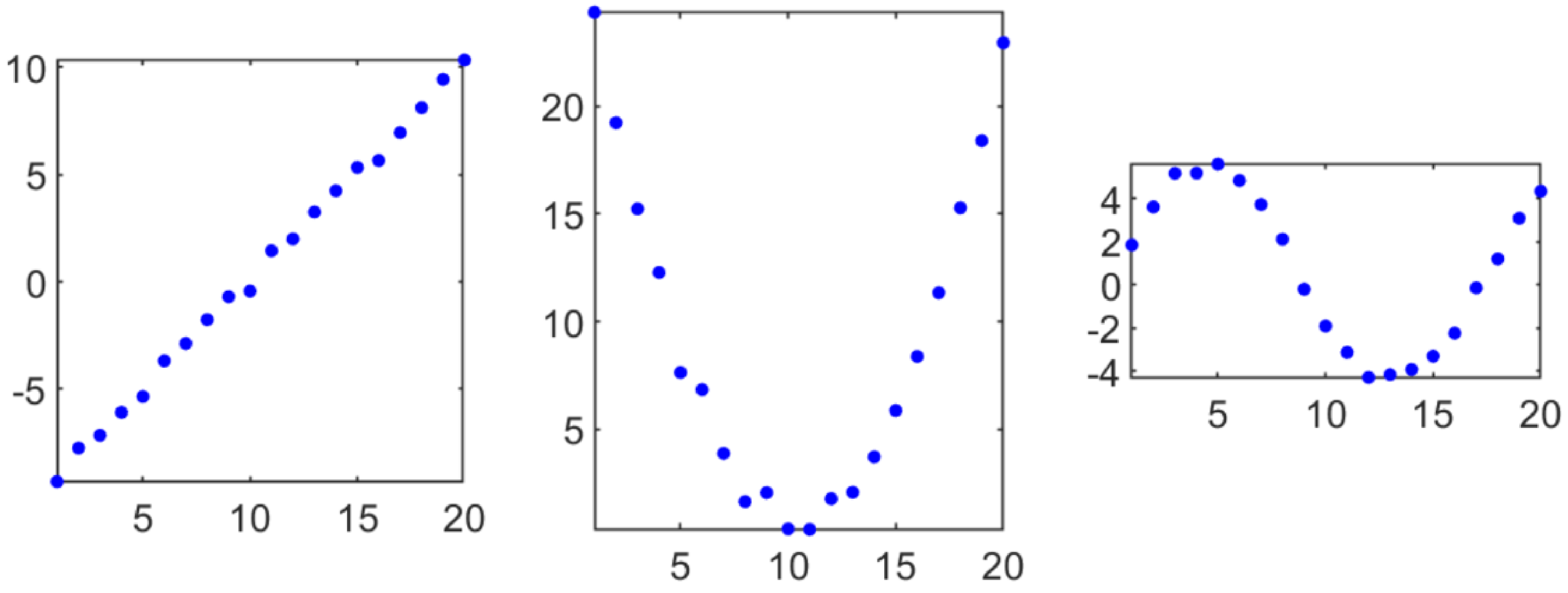
Exploiting and extracting information that is contained in the data.

Slide credit E. *Rodolà*

Learning

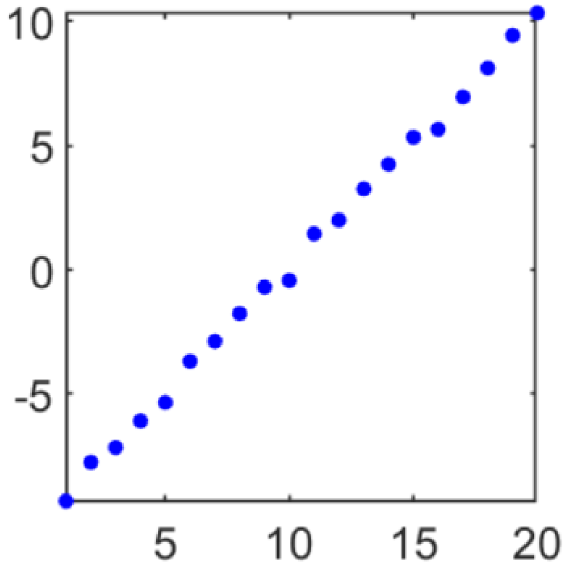
Learning = find a description of data

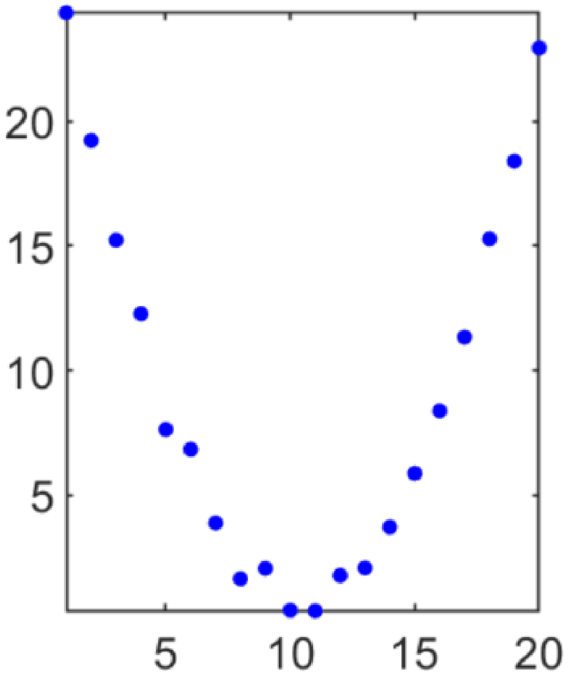
Or even better: describe the process (or model) that associates a given output from a given input

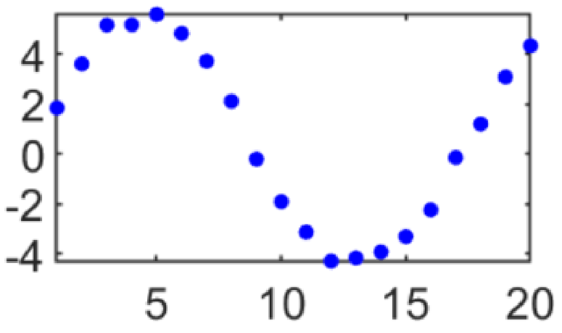


Slide credit E. *Rodolà*

An example







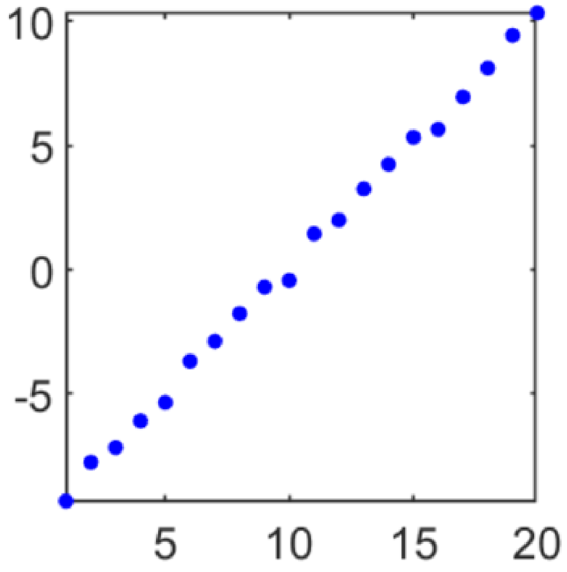
$y = ax + b$

$y = ax^2 + bx + c$

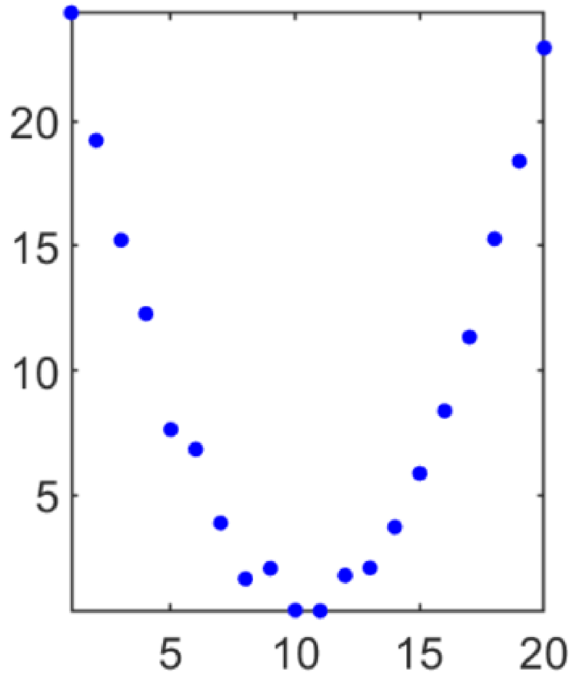
$y = ax^3 + bx^2 + cx + d$

Slide credit E. *Rodolà*

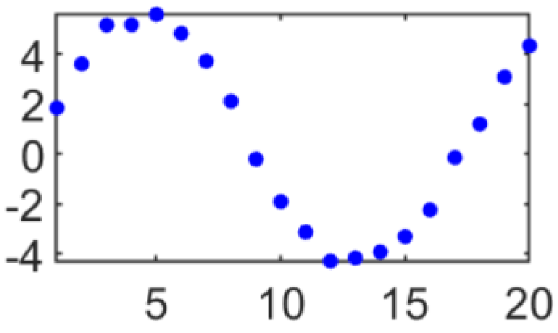
Not only data



$$y = ax + b$$



$$y = ax^2 + bx + c$$

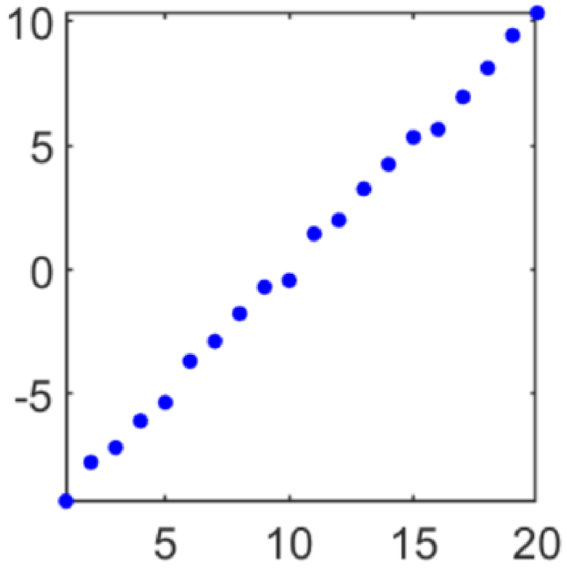


$$y = ax^3 + bx^2 + cx + d$$

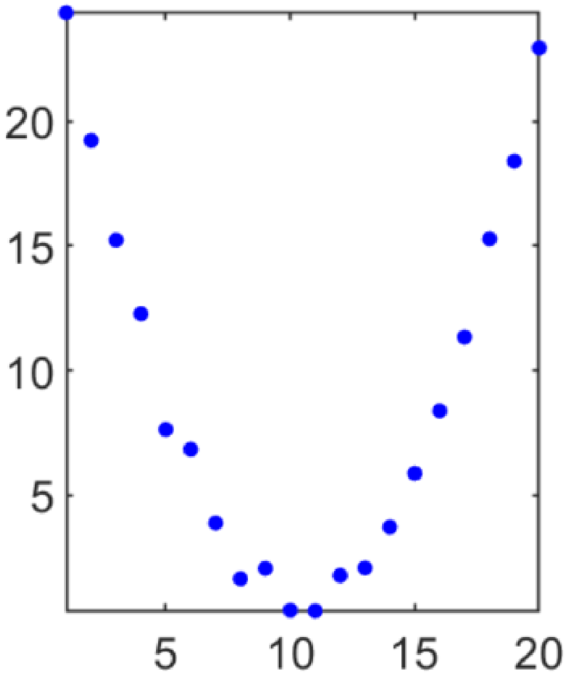
In addition, we could know something about the data or about the process (Prior Knowledge)

Slide credit E. *Rodolà*

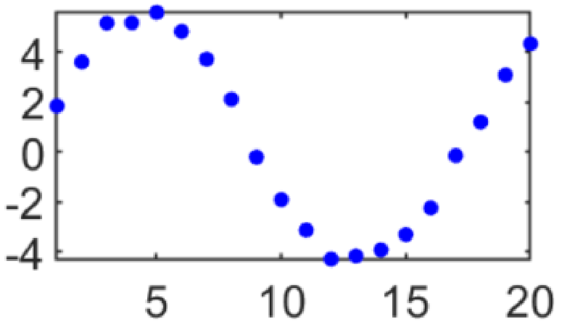
Not only data



$$y = ax + b$$



$$y = ax^2 + bx + c$$



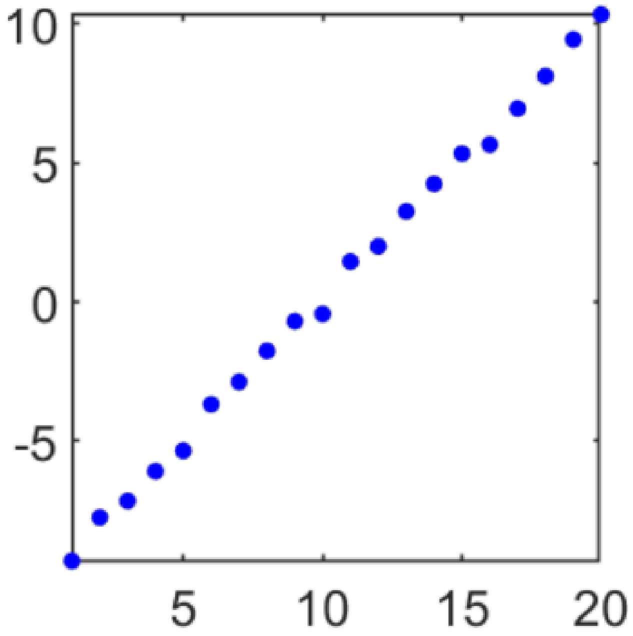
$$y = a\sin(x) + bx + c$$

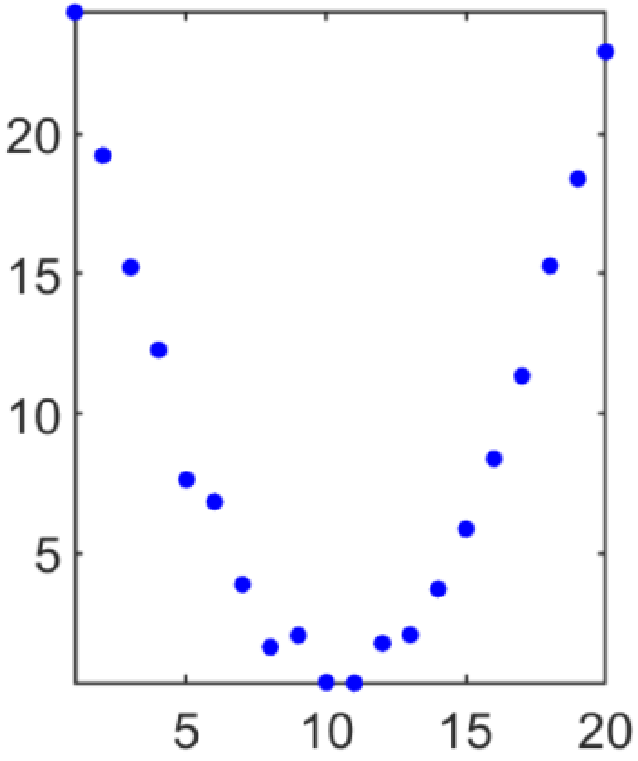
In the third example, we could know that the process is **periodic**

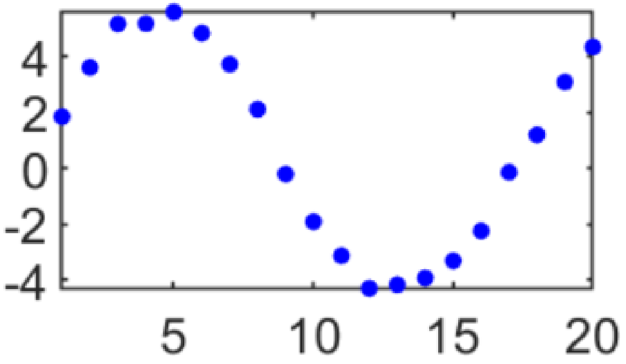
Slide credit E. *Rodolà*

Model = map (function)

Learning = discovery the **map** from **input** to **output**







$y = ax + b$

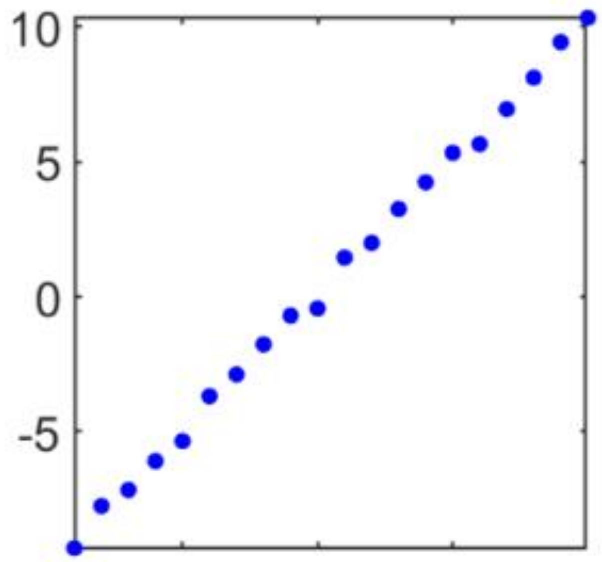
$y = ax^2 + bx + c$

$y = a\sin(x) + bx + c$

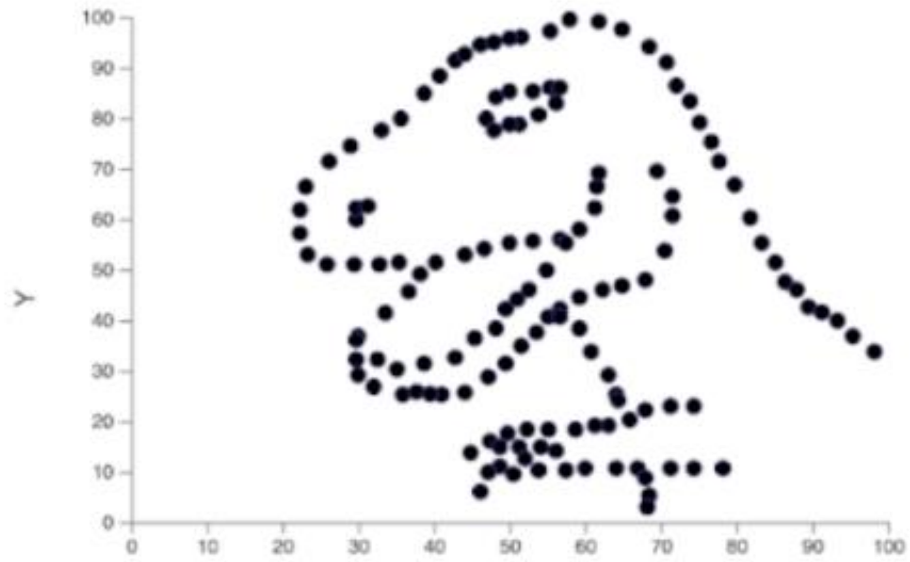
Slide credit E. *Rodolà*

Data structure

Key assumptiomn: the data has an underlying structure



$$y = ax + b$$



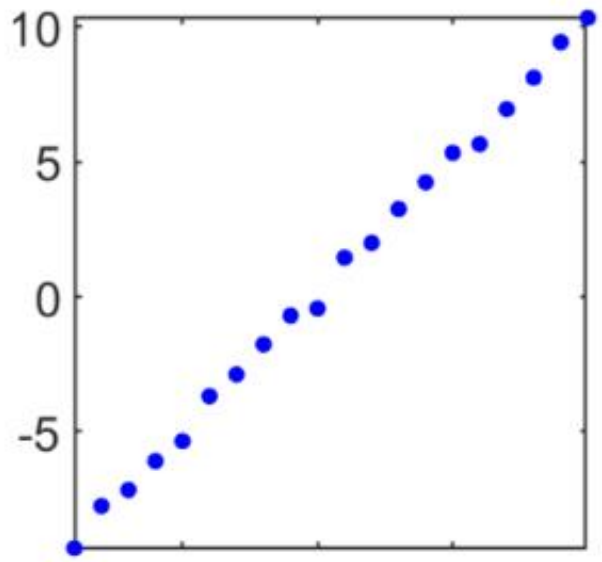
?

This structure is not usually well-described by simple functions

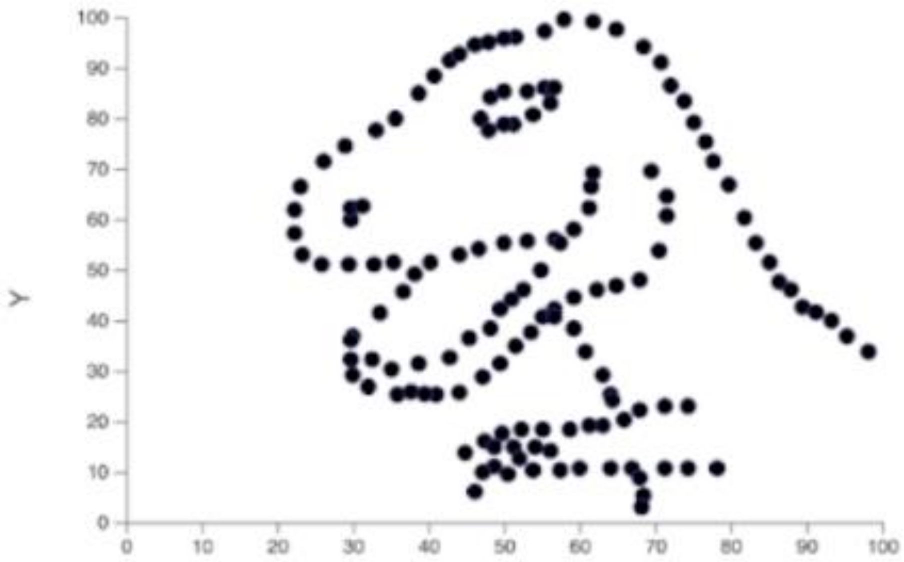
Slide credit E. *Rodolà*

Data structure

Key assumptiomn: the data has an underlying structure



$$y = ax + b$$



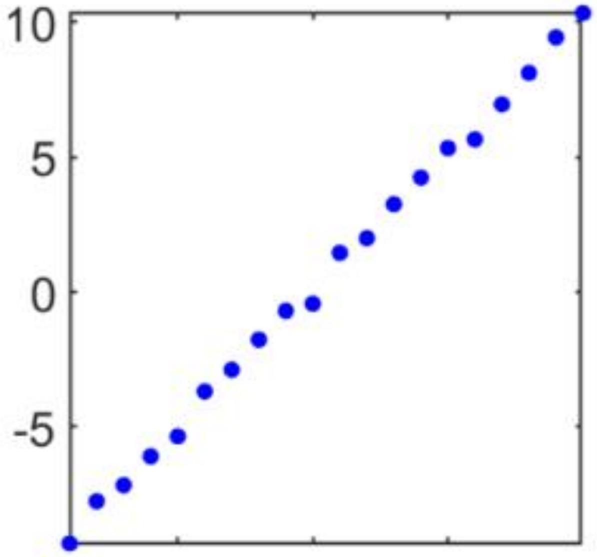
?

This structure is not usually well-described by **LINEAR** functions

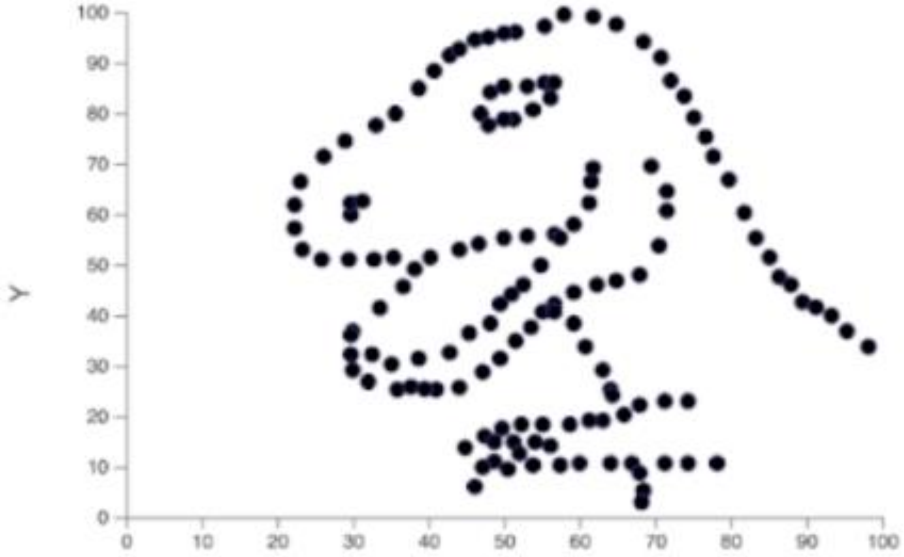
Slide credit E. *Rodolà*

Data structure

Key assumptiomn: the data has an underlying structure



$$y = ax + b$$



?

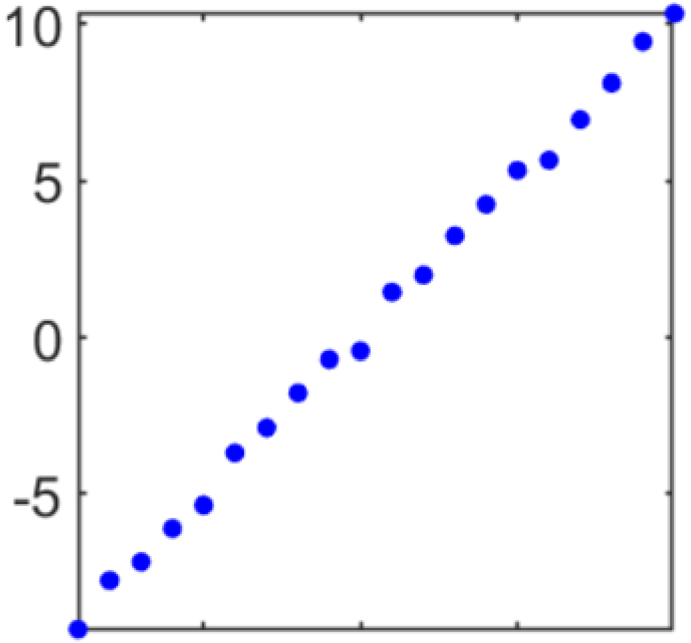
Data and functions could be not 1D

Slide credit E. *Rodolà*

Not a unique representation

There are different models that could represent the same data

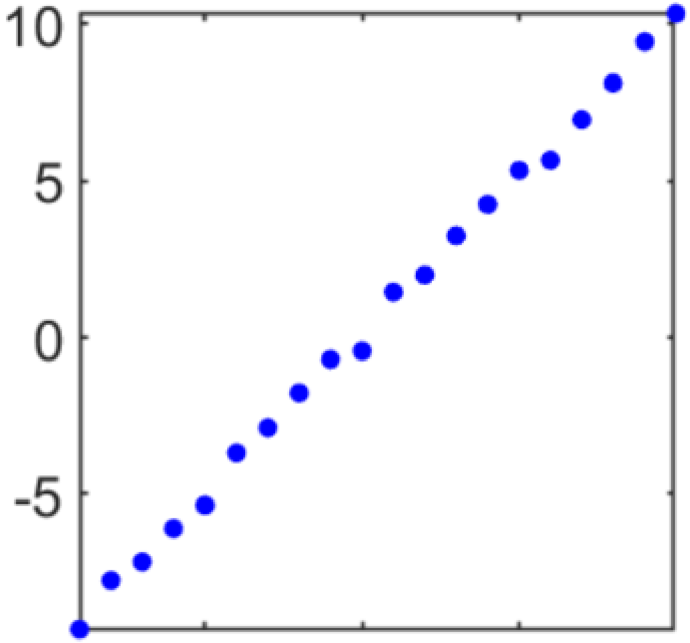
as a **function**



$$y = ax + b$$

$=$

as a **parametric curve**

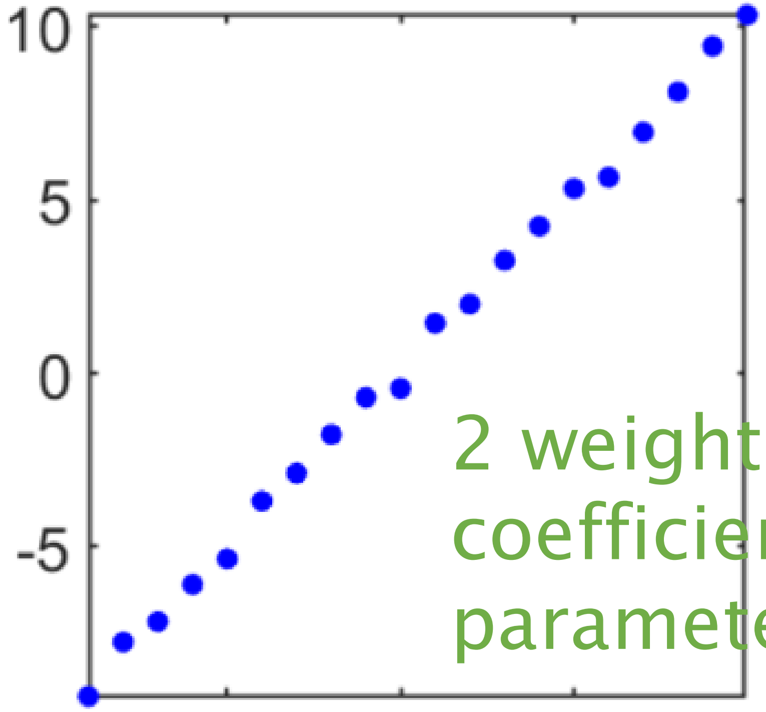


$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_x \\ a_y \end{pmatrix} t + \begin{pmatrix} b_x \\ b_y \end{pmatrix}$$

Not a unique representation

There are different models that could represent the same data

as a **function**

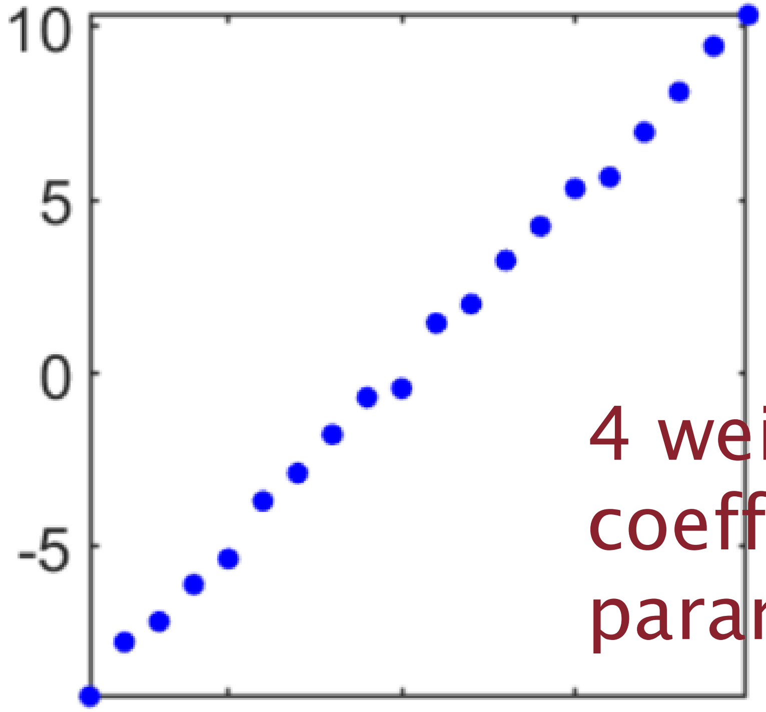


2 weights
coefficients
parameters

$$y = ax + b$$

=

as a **parametric curve**



4 weights
coefficients
parameters

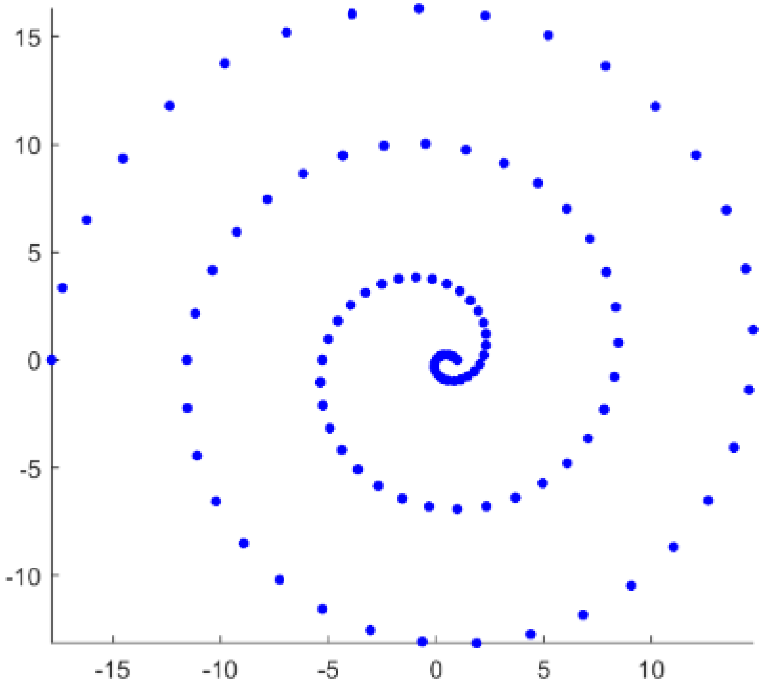
$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} a_x \\ a_y \end{pmatrix} t + \begin{pmatrix} b_x \\ b_y \end{pmatrix}$$

Slide credit E. *Rodolà*

The right representation

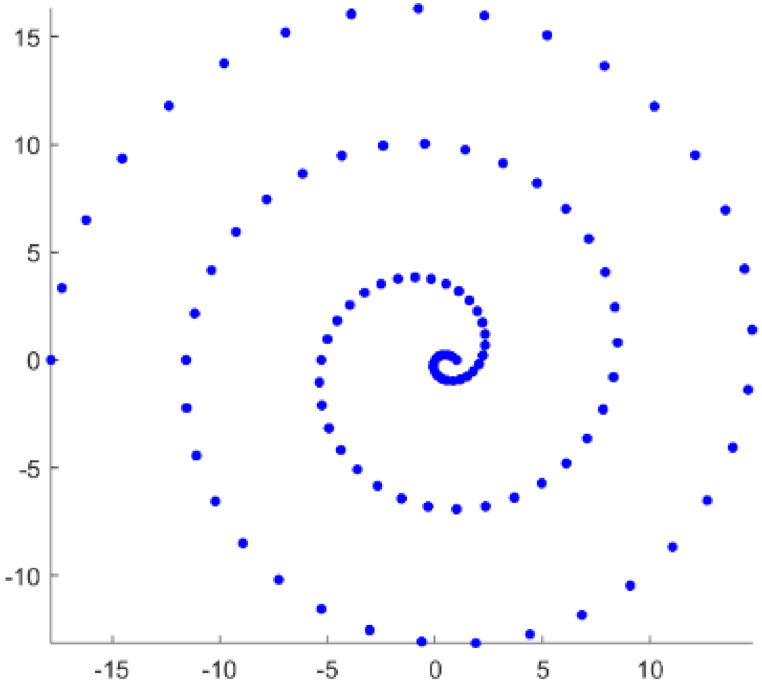
There are different models that could represent the same data

as a **function**



=

as a **parametric curve**



y is not a function of x

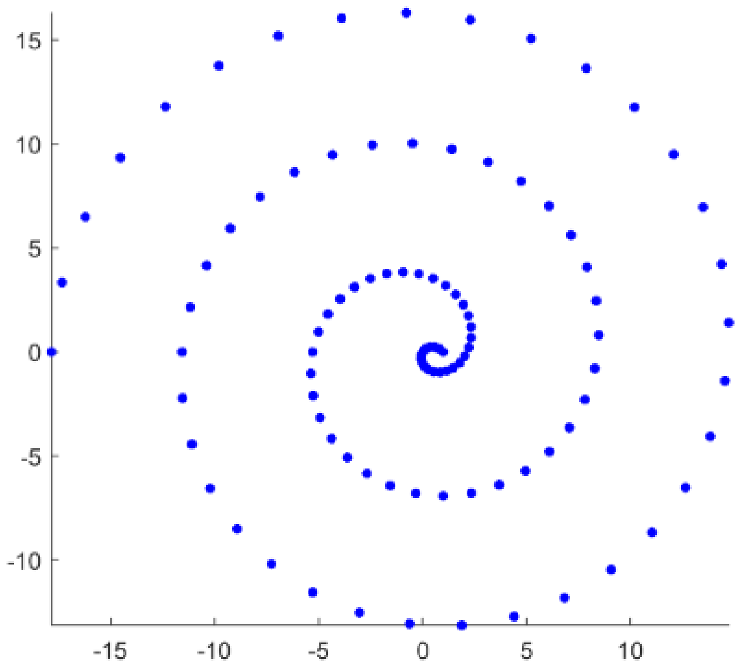
$$\begin{pmatrix} x \\ y \end{pmatrix} = a \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} (a - t)$$

Slide credit E. *Rodolà*

The right representation

There are different models that could represent the same data

as a **function**

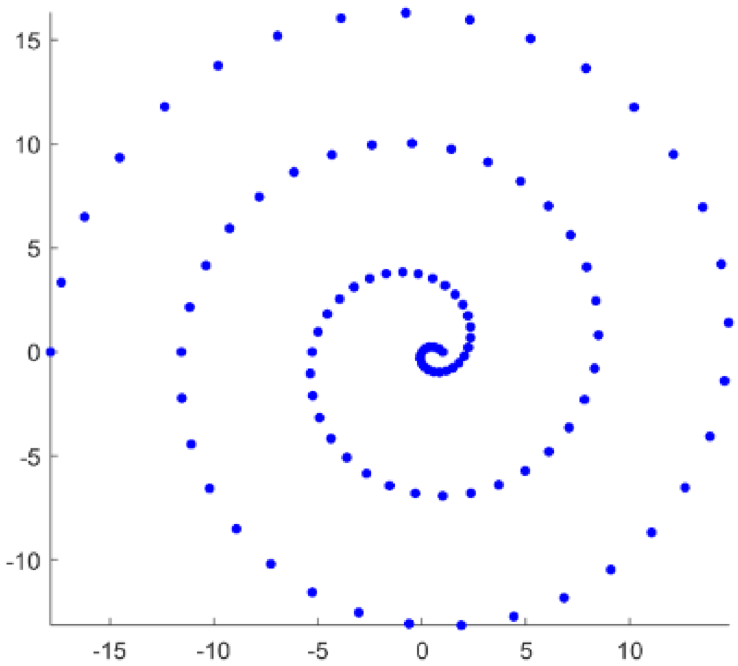


$$r = a\theta$$

(polar coordinates)

=

as a **parametric curve**



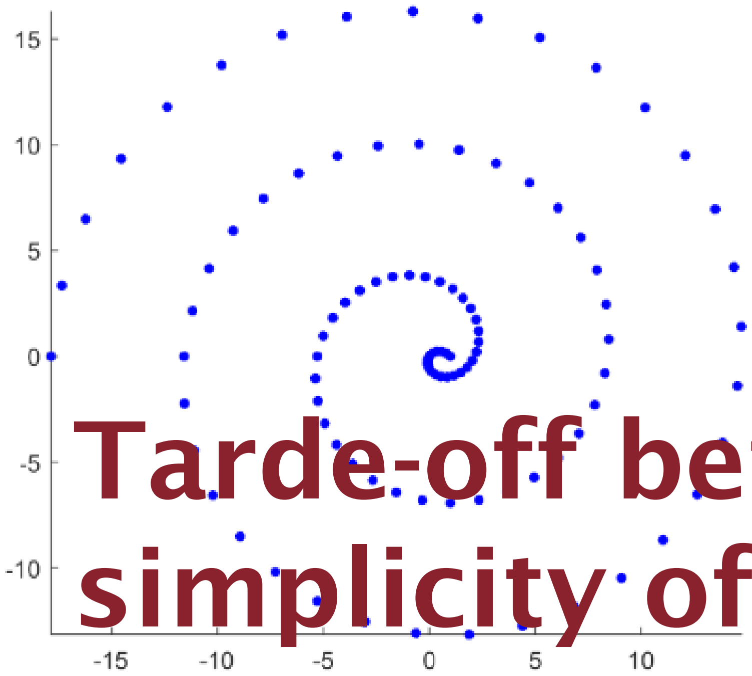
$$\begin{pmatrix} x \\ y \end{pmatrix} = a \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} (a - t)$$

Slide credit E. *Rodolà*

The right representation

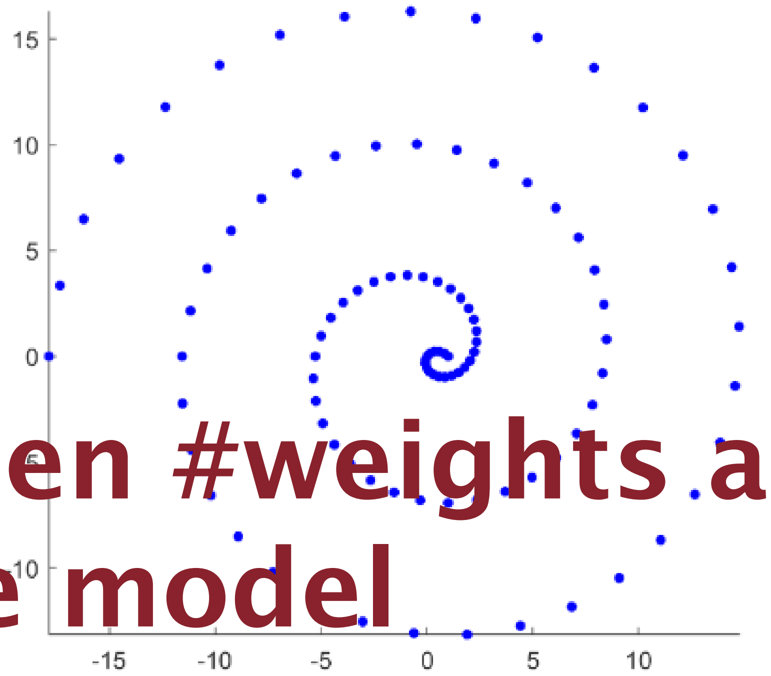
There are different models that could represent the same data

as a **function**



$=$

as a **parametric curve**



Tarde-off between #weights and simplicity of the model

$$r = a\theta$$

linear!

$$\begin{pmatrix} x \\ y \end{pmatrix} = a \begin{pmatrix} \cos t \\ \sin t \end{pmatrix} (a - t)$$

Slide credit E. *Rodolà*

Data dimensionality

Data can have more than dimension 1 or 2

A $h \times w$ image is represented by a vector of size hw each entry of which is the gray value at the corresponding pixel

$\in \mathbb{R}^{w \times h} \cong \mathbb{R}^{wh}$

A ~1 megapixel image (grayscale) has $\sim 10^6$ dimensions

Not all these dimensions are informative

Slide credit E. *Rodolà*

Need for Data

A dataset of natural images will be extremely sparse in $\mathbb{R}^{h \times w}$

And some regions of this space will be observed very frequently

Tarde-off between #dimensions and Amount of data required

Slide credit E. *Rodolà*

Need for Priors

Priors help to better understand the data

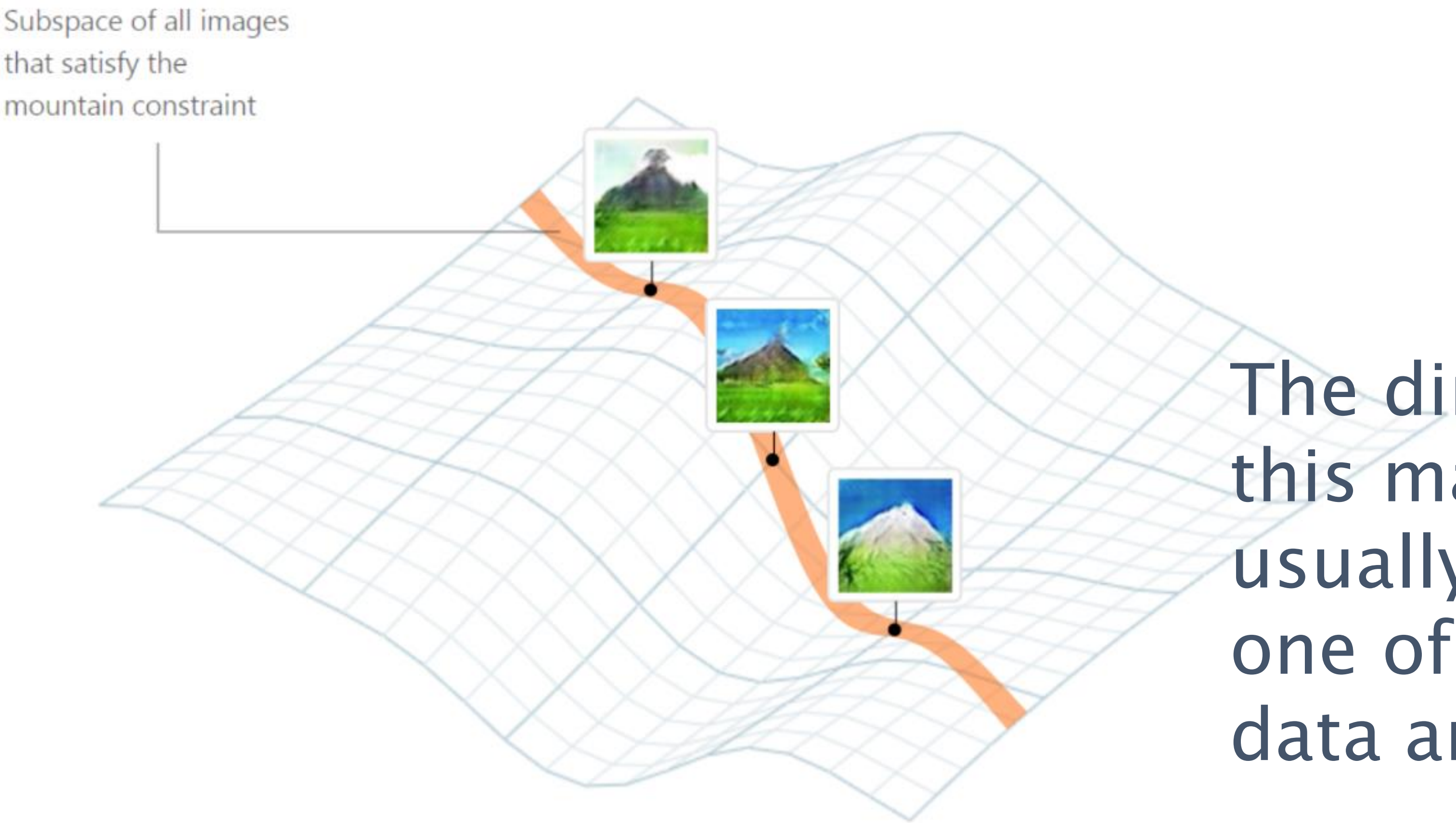
If we can assume some priors on the data we will be able to select a more meaningful representation to model the data-distribution

One common prior assumed in deep learning is the **Manifold hypothesis**

Slide credit E. *Rodolà*

Manifold Hypothesis

The input data lives in a some underlying non-Euclidean structure called a Manifold



The dimensionality of this manifold are usually smaller than the one of the space where data are represented.

Slide credit E. *Rodolà*

AI and applications

The AI techniques are exploited in several applications

- Economy and finance
- Social analysis
- Agriculture
- Cybersecurity
- Education
- Healthcare
- Media
- Commerce (e-commerce)
- Manufacturing
- Automotive
- …

We usually imagine them applied to audio signals or images

AI and Images (2D)

Many outstanding results have been achieved on applications that involve images (computer vision)

Due to the huge amount of data of this kind available ImageNet contains 1.281.167 training images, 50.000 validation images and 100.000 test images

Why Geometric deep Learning?

Pixels (**Euclidean**)

Slide credit E. *Rodolà*

Why Geometric deep Learning?

Geometry (Non-**Euclidean**) Pixels (**Euclidean**)

Slide credit E. *Rodolà*

Geometric deep Learning

1. Data are not organized on a fixed grid or template

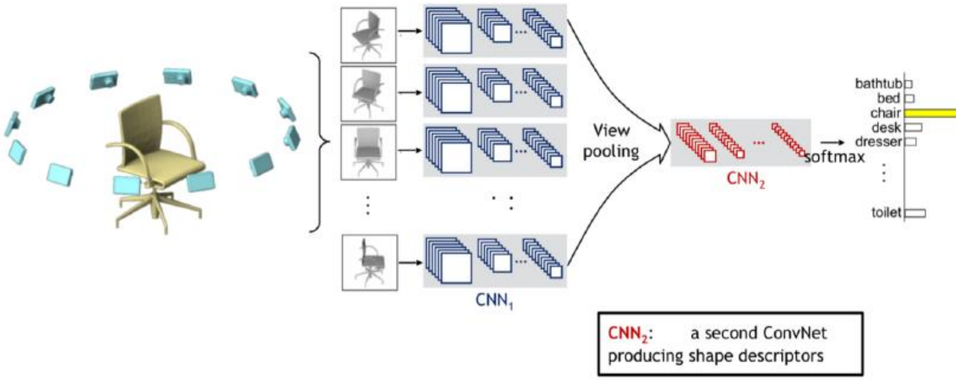
2. Different representation are possible (as we will see)

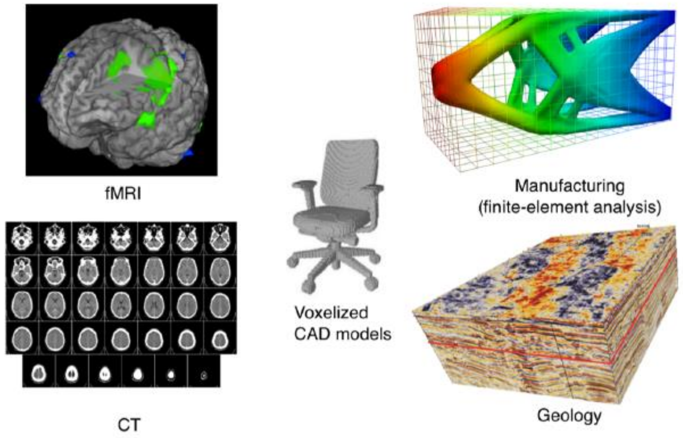
3. More Rigid transformations are possible

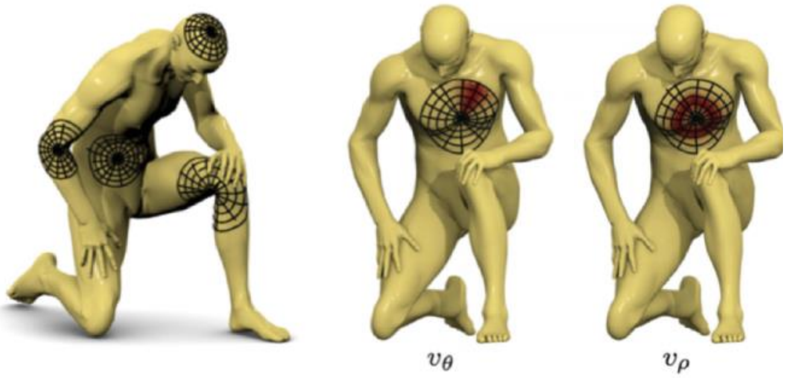
4. Limited amount of data available

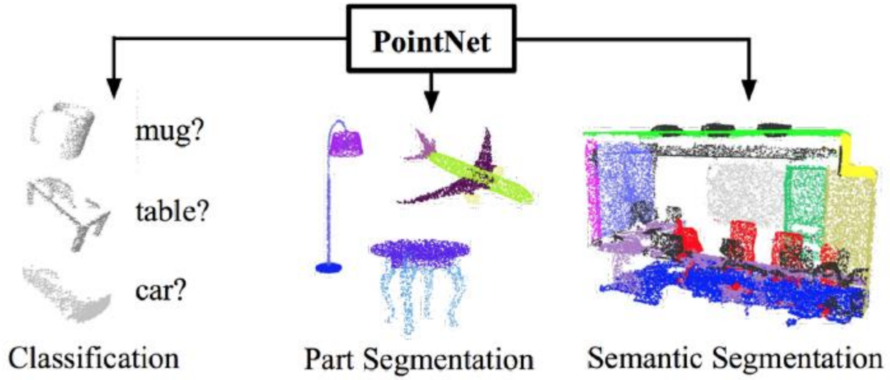
5. Limited amount of information available (labels, segmentations,..)

We can learn on 3D geometries









The main scope of this course is to see some of the solutions that have been proposed in the last decade