Introduction to Machine Learning

Marcello Pelillo

University of Venice, Italy

Artificial Intelligence

a.y. 2017/18

Ca' Foscari University of Venice

What Is Machine Learning?

A branch of **Artificial Intelligence (AI)**.

Develops algorithms that can **improve their performance** using training data.

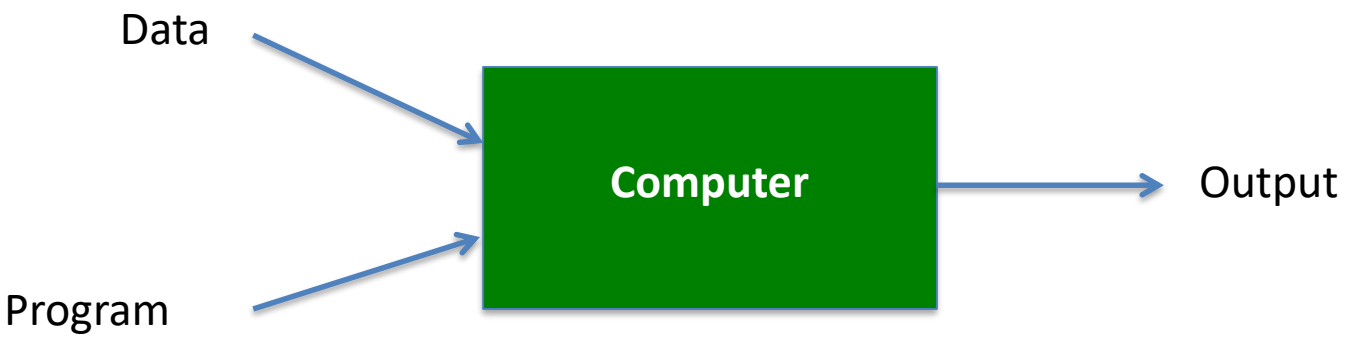
Typically ML algorithms have a (large) number of parameters whose values are learnt from the data.

Can be applied in situations where it is very challenging (= impossible) to define rules by hand, e.g.:

- Face detection
- Speech recognition
- Stock prediction
- …

Machines that Learn?

Traditional programming



Machine learning

Traditional Programming


```
if  (eyes == 2) &
    (legs == 4) &
    (tail == 1 ) &
    …
then Print "Cat!"
```

Machine Learning

Computer

Cat recognizer

"Cat"

Learning algorithm

Cat!

Data Beats Theory

«By the mid-2000s, with success stories piling up, the field had learned a powerful lesson: **data can be stronger than theoretical models**.

A new generation of intelligent machines had emerged, powered by a small set of statistical learning algorithms and large amounts of data.»

Nello Cristianini

The road to artificial intelligence: A case of data over theory

(New Scientist, 2016)

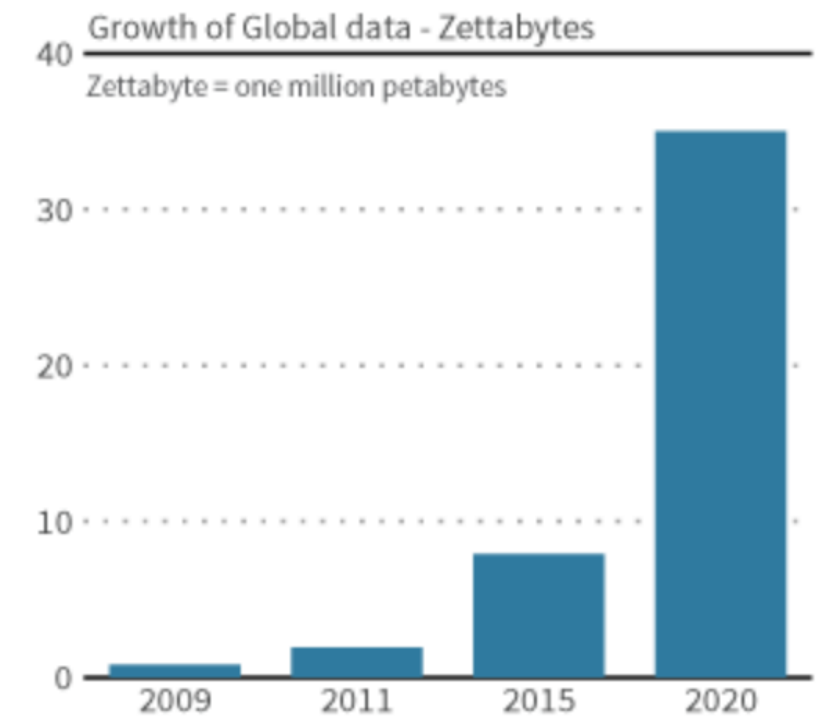
From Knowledge-Based Systems…

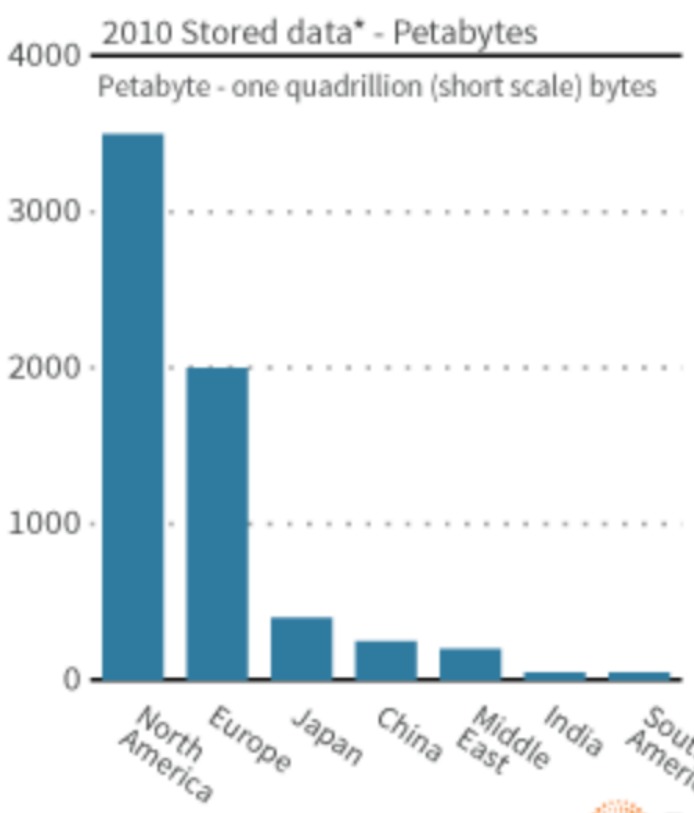
From: N. Cristianini, *On the current paradigm in artificial intelligence* (2014)

… to Machine Learning

From: N. Cristianini, *On the current paradigm in artificial intelligence* (2014)

Data Growth





*greater than

Sources: Nasscom -CRISIL GR&A analysis

REUTERS

What Kind of Data?

Video

Streaming video takes up more than 1/3 of the Internet traffic during normal television watching hours

72 hours of video are added to YouTube every minute

864,000 hours of YouTube video are uploaded each day

22 million hours of TV and movies are watched on Netflix each day

Zynga processes 1 petabyte of videogame content per day

Social media

More than 1.4 billion online consumers are spending 22 percent of their time on social platforms

172 million individuals visit Facebook each day

4.7 billion minutes spent on Facebook each day

532 million Facebook statuses updated each day

250 million photos uploaded to Facebook each day

30+ billion pieces of data are added to Facebook each month

40 million Twitter individual users each day

50 million tweets per day

32 billion searches performed on Twitter per month

22 million LinkedIn individual users each day

20 million Google+ individual users each day

17 million Pinterest individual users each day

2 million blog posts are written each day

Other digital platforms

1.3 exabytes of data sent and received by mobile Internet users each month

Average teenager sends 4,762 text messages per month

More iPhones are sold than babies born each day

294 billion emails are sent each day

72.9 products ordered per second on Amazon

18.7 million hours of music is streamed on Pandora each day

1,288 new apps are available to download each day

More than 35 million apps are downloaded each day

By 2018, there will be a demand for about 450,000 data scientists in the U.S., leaving a shortage of more than 150,000 positions

Sources: Mushroom Networks, Deep Blue Analytics, MBAOnline, IBM, Gartner

From: M. L. Stone, *Big Data for Media.* Reuters Institute for the Study of Journalism (Nov 2014)

The Philosophy of Data Science

«This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. Out with every theory of human behavior, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity.

With enough data, the numbers speak for themselves.»

Chris Anderson

The end of theory (Wired, 2008)

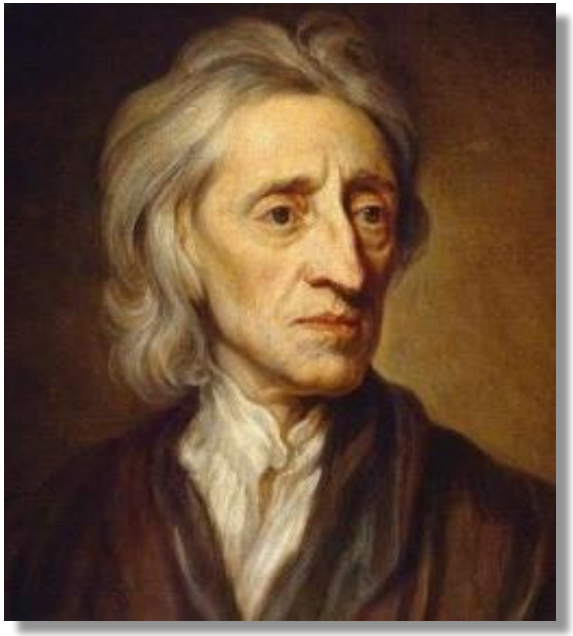
Back to *Tabula Rasa*

«Let us then suppose the mind to be, as we say, white paper void of all characters, without any ideas. How comes it to be furnished? Whence comes it by that vast store which the busy and boundless fancy of man has painted on it with an almost endless variety? Whence has it all the materials of reason and knowledge?

To this I answer, in one word, from EXPERIENCE.

In that all our knowledge is founded;

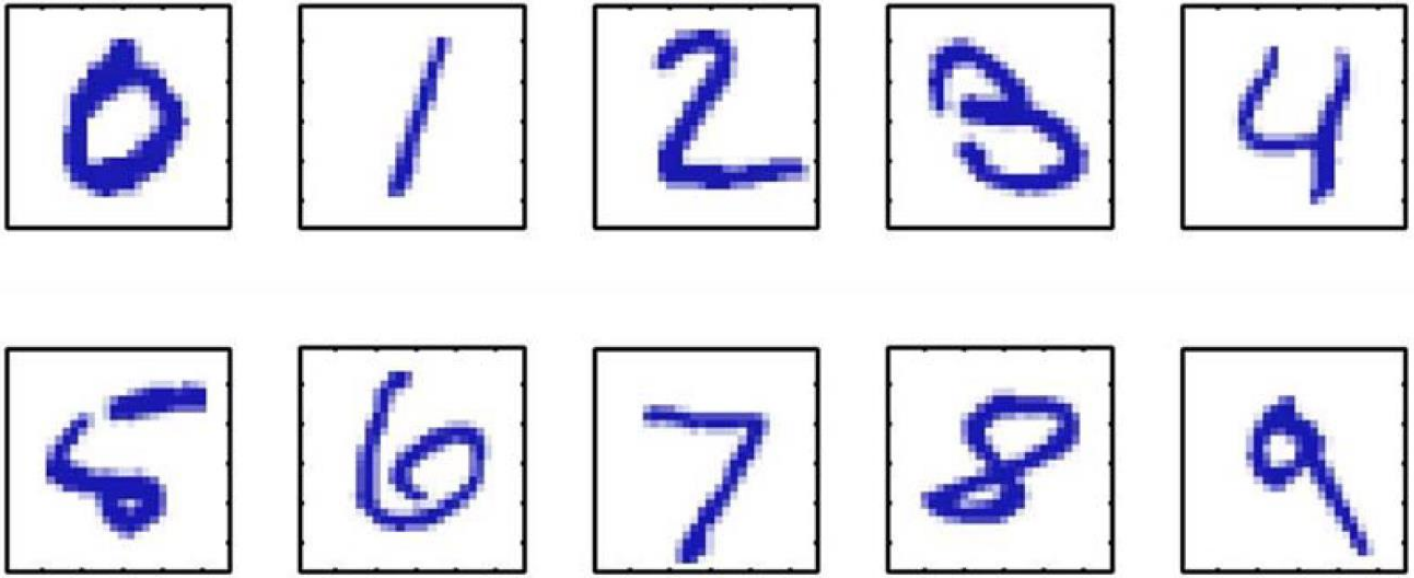
and from that it ultimately derives itself.»



John Locke

An Essay Concerning Human Understanding (1690)

Example:
Hand-Written Digit Recognition



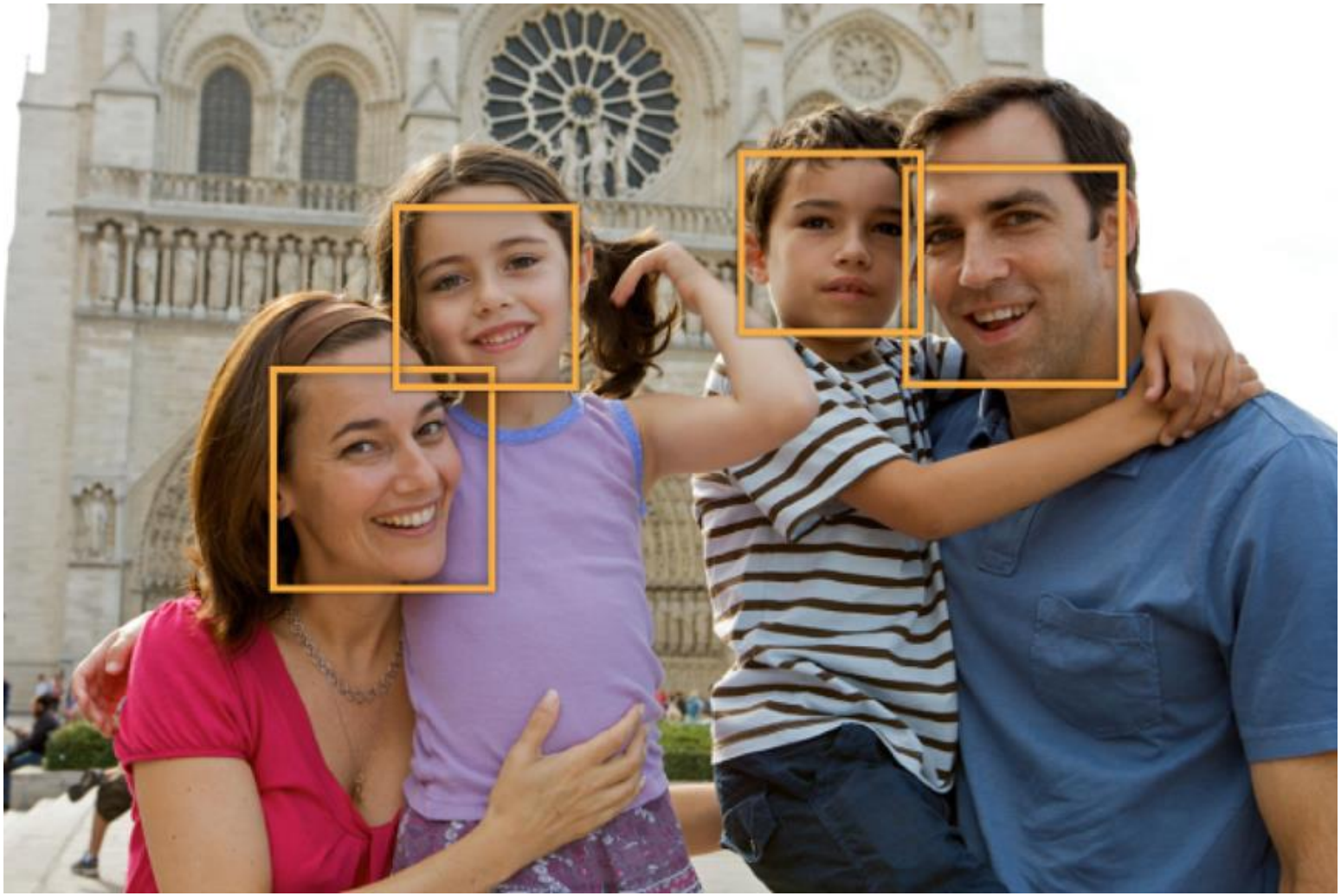
Images are 28 x 28 pixels

Represent input image as a vector $\mathbf{x} \in \mathbb{R}^{784}$

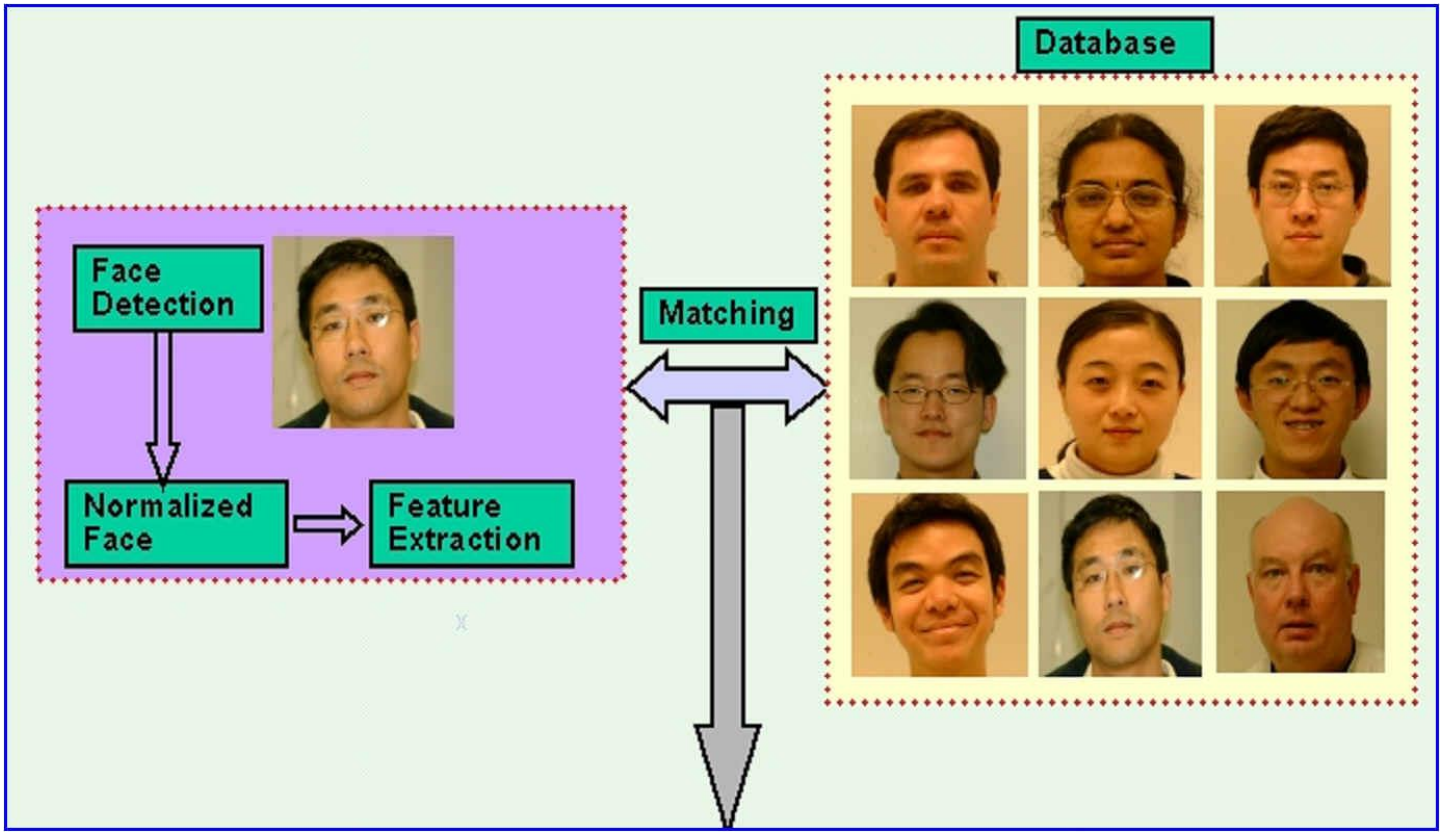
Learn a classifier $f(\mathbf{x})$ such that,

$$f : \mathbf{x} \rightarrow \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

Example:
Face Detection



Example:
Face Recognition

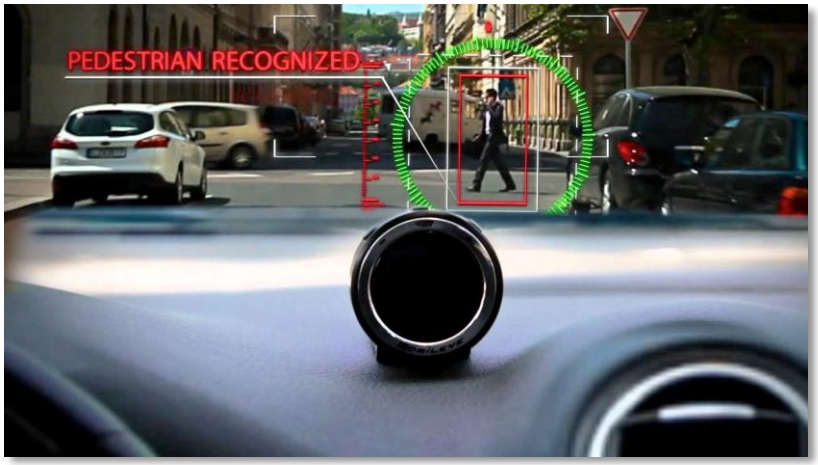


The Difficulty of Face Recognition

Example:
Fingerprint Recognition

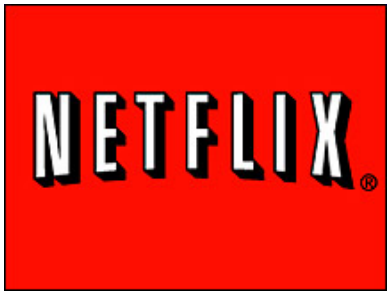
?

Assiting Car Drivers and Autonomous Driving



Assisting Visually Impaired People

Recommender Systems



The Road Ahead

Supervised learning

- Neural networks (from perceptrons to deep networks)
- Elements of statistical learning theory
- Support vector machines

Unsupervised learning (a.k.a. clustering)

- Spectral clustering
- Game-theoretic methods

Semi-supervised learning and context-aware methods

- Graph transduction games
- Consistent labeling problems