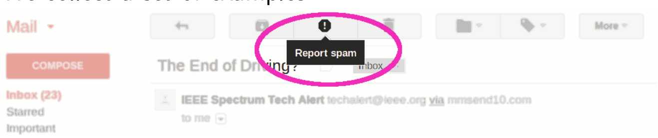
Algorithms VS Machine learning

- For some problems we don't have an algorithm
  - Example: Recognize a face
  - Example: Perform Optical Character Recognition (OCR)
  - Example: Tell spam from legitimate emails
  - Example: Group products frequently bought together
- Why?
  - Humans are often very good at performing a task unconsciously but are unable to explain how. If we are not able to explain our expertise, we cannot write a computer program!
  - The task may change with time or situation (e.g., subjective definition of spam); we need a flexible framework to perform the task.
  - Sometimes we don't even know what we are looking for but would like a machine to provide us with clever questions as well as answers.

Algorithms VS Machine learning

- We would like the computer (machine) to extract automatically the algorithm for these tasks.
- How?
  - We collect a set of examples



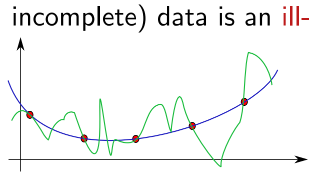
  - We build computer algorithms able to learn the task-related algorithm from the data

    What we lack in knowledge, we make up for using data.

  - On the other hand, data by itself is not enough: we still the expert's knowledge to design the ML system and provide the so-called prior knowledge.

Importance of prior knowledge in ML

- Prior knowledge: information about the problem available in addition to the training data
- Why? without prior knowledge, learning a ML model from finite (and incomplete) data is an ill-posed problem (no unique solution)



Without prior knowledge both solutions (and infinitely many more) are possible and equally "legitimate"

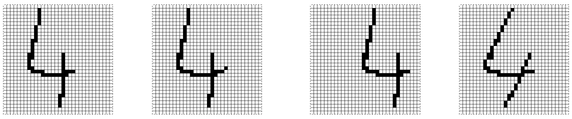
x_1	x_2	y
0	0	0
0	1	1
1	0	1

Should ML learn OR or XOR? Without prior knowledge, both are equally "legitimate" solutions

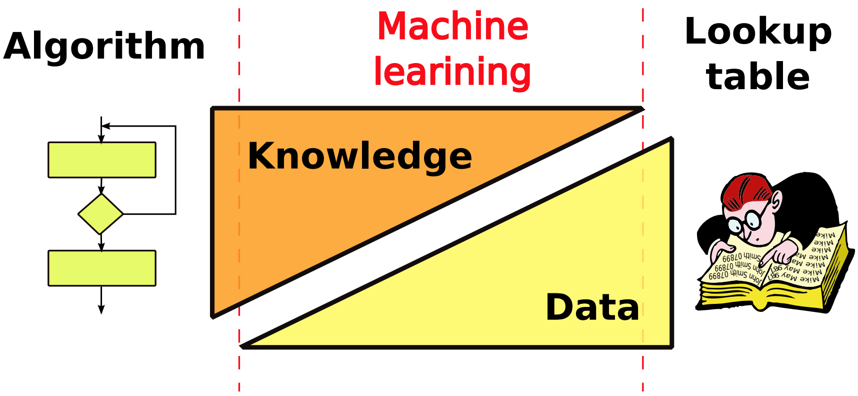
- The no free lunch theorem states that, if we average over all possible problems, no learning algorithm is better than another one (not even random guessing) on unseen data

Importance of prior knowledge in ML

- Listing all possible inputs is often impossible or unfeasible:
  - Even a simple $\mathbb{R} \to \mathbb{R}$ function would require a table with an (uncountable) infinite number of entries.
  - OCR on a small (32x32) B/W image would require $2^{32 \cdot 32} \sim 1.8 \cdot 10^{308}$ labelled sample images
- By exploiting prior information about a specific problem one can improve the performance by favouring solutions that are known to perform better in similar situations. For example:
  - General smoothness assumption, often small changes in the input are unlikely to cause big changes in the output
    (e.g., in OCR a single pixel is unlikely to change a 4 to a 6)
  - Transformation-invariance, often the output can be assumed to be invariant to some transformations of the input pattern
    (e.g., shear and translation in OCR, pitch and speed in speech recognition)



Machine Learning: data & prior knowledge



Assumptions for using Machine Learning

- We believe that there is a process that explains the data we observe.
- We try to learn a general model from limited amount of data and prior knowledge.
- Even if we may not be able to identify the process completely, we might be able to construct a good and useful approximation.
- The learned model should be able to make sufficiently-accurate predictions in previously-unseen cases

Learn a simple model with good GENERALIZATION abilities