Introduction to Machine Learning

Alessandro Leite

October 4th, 2019

Outline

What is learning?

Why learn is a crucial concept?

- Learning **goes beyond remembering** individual **experiences**, i.e., memorization.
- Children, for instance, can **generalize** from previous experiences very quickly.
- Children generalizing from their specific experiences manifest a **predictable phenomenon**.
- It is also referred as **inductive reasoning** or **inductive inference**.

The probably approximately correct (PAC) learning model

- Learning process is carried out by concrete computation that takes a limited number of steps.
- The computation also requires only a similarly limited number of interactions with the word during learning[a].
- Learning should enable organism to categorize new information with a small error rate.
- The induction process is not logically **fail-safe**. In other words, if the world suddenly changes, but not the knowledge, then, no one should expect or require good generalization in the future.

[a] L. G. Valiant. "A Theory of the Learnable". In: *Communication of the ACM* 27.11 (1984), pp. 1134–1142.

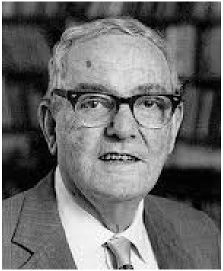
Why learn is an important concept?

- **Learning**: using past experience to guide future actions or to modify a behavior.
- **Machine learning**: programming computers to:
  1. **model phenomena**
  2. by means of optimizing an **objective function**
  3. use **data** and **examples**, instead of expert knowledge, to perform complex tasks automatically

What is machine learning?

"Learning is any process by which a system improves performance from experience".

"Machine learning is concerned with computer programs that automatically improve their performance through experience".



Herbert Simon
Turing Award (1975)
Nobel Prize in Economics (1978)

Definition

Machine learning can be defined as the process of an **algorithm extracts patterns from data** and to **make predictions without** being **explicitly programmed** to do so[a].

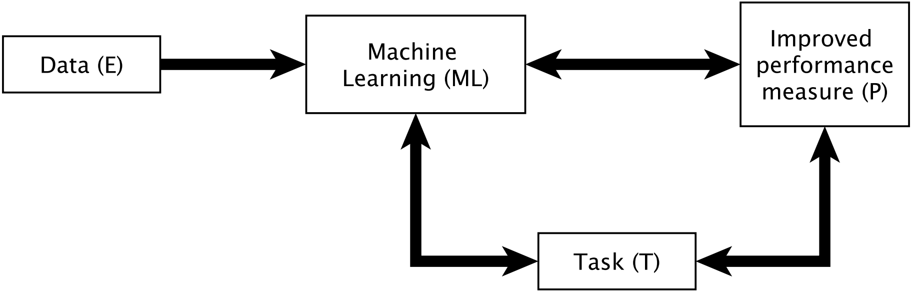
[a] A. L. Samuel. "Some Studies in Machine Learning Using the Game of Checkers". In: *IBM J. Res. Dev.* 3.3 (1959), pp. 210–229.

A more formal definition of machine learning

Definition

A computer program is considered to learn from **experience E** with respect to some **class of tasks T** and **performance measure P** if its performance tasks in **T**, as measured by **P**, improves with experience **E**[a].

[a] Tom M. Michell. *Machine Learning*. McGraw-Hill Education, 1997.



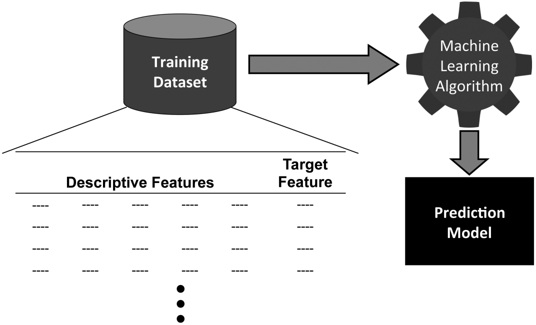


Figure 1: Using machine learning to induce a prediction model from a training dataset

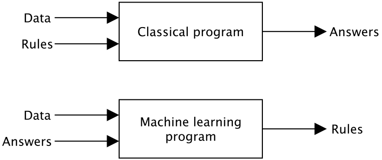
Figure 2: Using the model to make predictions for new query instances

Example

ID	Occupation	Age	Loan-Salary Ratio	Outcome
1	industrial	34	2.96	repaid
2	professional	41	4.64	default
3	professional	36	3.22	default
4	professional	41	3.11	default
5	industrial	48	3.80	default
6	industrial	61	2.52	repaid
7	professional	37	1.50	repaid
8	professional	40	1.93	repaid
9	industrial	33	5.25	default
10	industrial	32	4.15	default

- What is the relationship between the **descriptive features** (Occupation, Age, Loan-Salary Ratio) and the **target feature** (Outcome)?

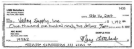
When do we need machine learning?



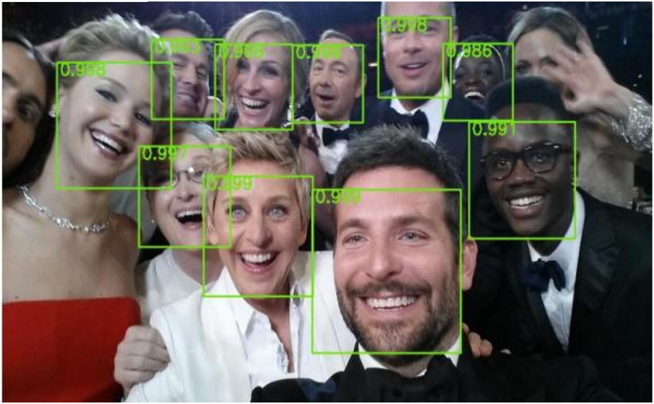
- There is not a requirement to learn to calculate a payroll, for instance.
- **Automatic learning** is used when:
  - **Text or document classification**: spam detection, automatically determines if the content of a web page is inappropriate or too explicit
  - **Humans are unable to explain their expertise** (e.g., speech recognition, computer vision)
  - **Lack of human expertise** (e.g., bioinformatics) 
  - **Tasks that are beyond human capabilities** (e.g., analysis of complex data sets: astronomical data, weather prediction, analysis of genomic data).

→ “b”

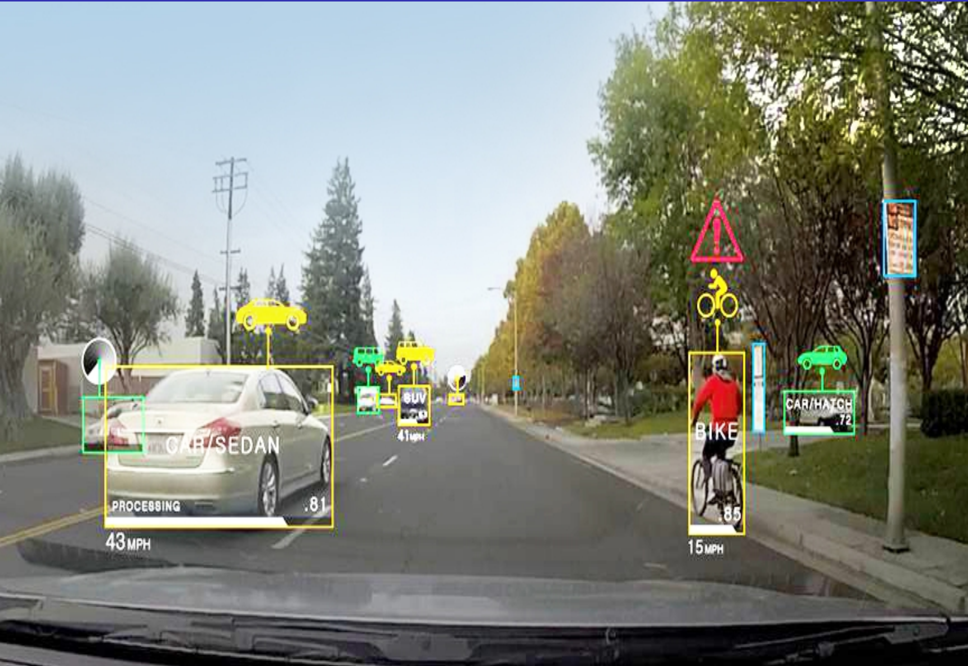
- Less programming code
- Robust and easily adaptive
- Less dependent on expert knowledge
- Used in many applications due to its good performance



Face detection



Object recognition



Autonomous Helicopter Control

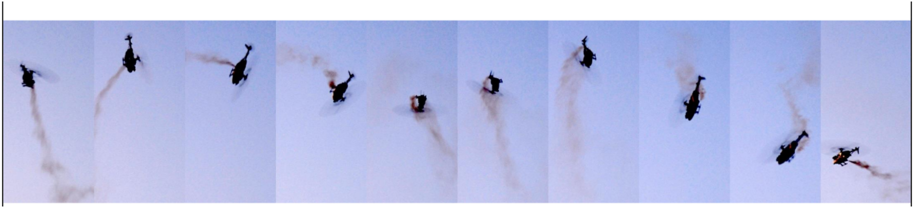


Figure 3: Autonomous helicopter control

The ability to learn grammars is **hard-wired** into the brain. It is not possible to "learn" linguistic ability—rather, we are born with a brain apparatus specific to language representation.

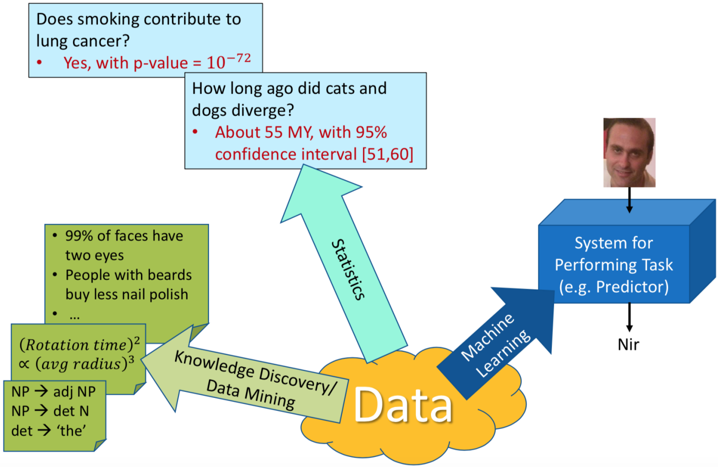
There exists some "universal" learning algorithm that can learn ***anything***: language, vision, speech, etc. The brain is based on it, and we're working on uncovering it. (Hint: the brain uses neural networks)

There is no "free lunch": no learning is possible without *some* prior assumption about the structure of the problem (prior knowledge)

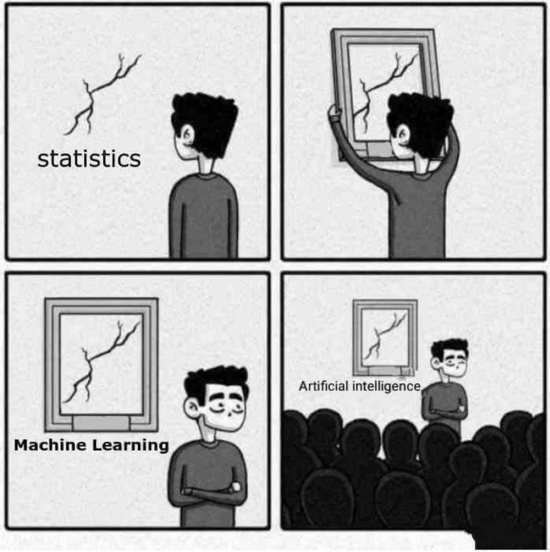
Data is the combustible of machine learning

Machine learning

Use data and examples, instead of expert knowledge, to automatically create systems that perform complex tasks



Is machine learning only statistics and probability?



Comparing statistics techniques with machine learning methods

- Statistical methods is used to analyze the data
- Machine learning is used to make prediction
- When we do with machine learning, you have to understand statistics

- When data is wide (e.g., over 100 features) – it's machine learning
- Variables are correlated – it's machine learning
- Simple models are associated with statistics (e.g., regression), while fancy methods are associated with machine learning (e.g., random forests)

What is the relationship of machine learning with other fields?

Artificial Intelligence (AI)

Artificial Intelligence (AI)

- Machine learning is a subfield of artificial intelligence (AI):
  - Systems living in an **evolving environment** must have the ability to **learn** in order to **adapt** themselves
  - Machine learning algorithms are building blocks that make computers behave intelligently by **generalizing** rather than merely storing and retrieving data, as database systems usually do

Machine learning and AI

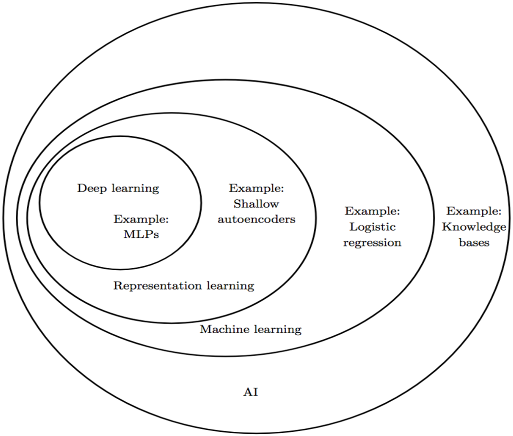
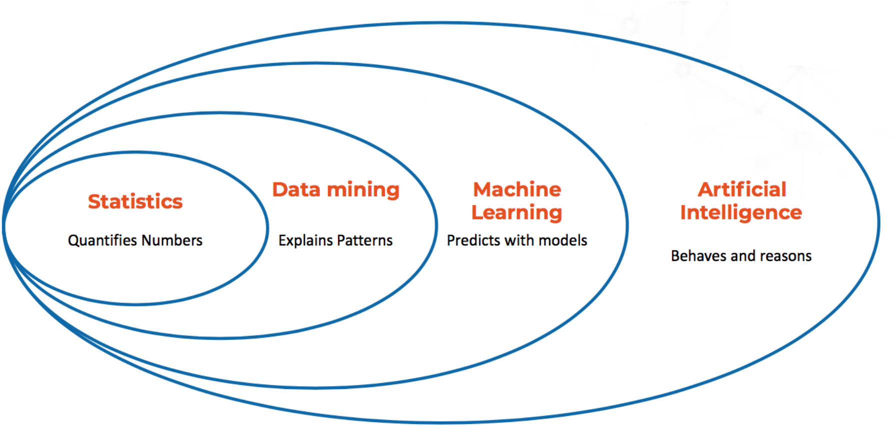


Figure 4: Relationships deep learning, machine learning, and AI

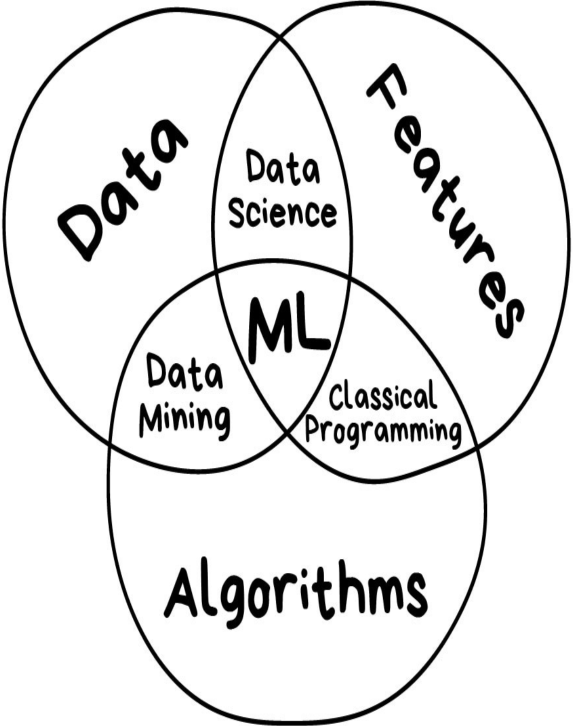
Machine learning and other fields



Role of statistics, and Computer Science

- **Statistics**:
  - build mathematical models to make inference from a sample
- **Computer Science**: develop efficient algorithms to:
  - solve the optimization problem
  - represent and evaluate the model for inference

How does machine learning work?


Data
Features
Data Science
ML
Data Mining
Classical Programming
Algorithms

A machine learning model is the combination of different components

$$\text{Model} = \text{Data} + \text{Features} + \text{Algorithms}$$

- **Features** are chosen based on some objective and domain knowledge
- **Data** are selected based on some objective and on descriptive features
- **Algorithms** are selected base on objectives + features + data

- **Data**
    - What are its characteristics?
    - How complete are they?
    - How meaningful are them for the problem?
- **Features**
    - Which one are useful?
    - Which ones can improve the final results (i.e., accuracy)?
- **Algorithms**
    - Which one is more adequate considering the data and their characteristics?
    - How complex are them?
    - What are the parameters that must be tuned to improve the performance?

- Learning general models from particular examples
  - **data** are mostly cheap and abundant
  - **knowledge** is expensive and scarce
- However, because a training dataset is only a sample ML is an ill-posed problem.
- Example in retail:
  - From customers' transactions to consumers behaviors
  - People who watched *Lords of the Rings* also watched Games of Thrones
- **Goal**: build a model that is a good and a useful approximation of the data
  - An obvious criterion to drive this process is to look for models that are consistent with the data
  - Therefore, this may be an ill-posed problem

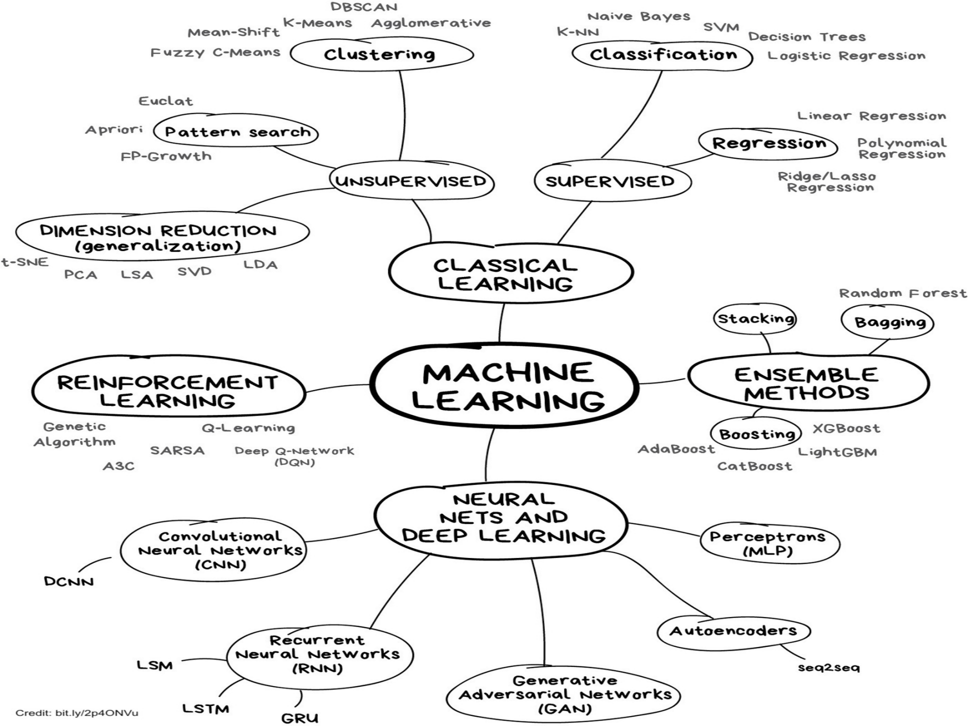
Machine learning's goal

- Optimizing a performance criteria using example past experiences (i.e., data)
- What criteria should we use for choosing a model?
  - Inductive bias

    set of assumptions that define the model selection criteria of a machine learning algorithm.
  - There are two types of bias that we can use:
    1. restriction bias
    2. preference bias
  - Inductive bias is necessary for learning beyond previous experiences

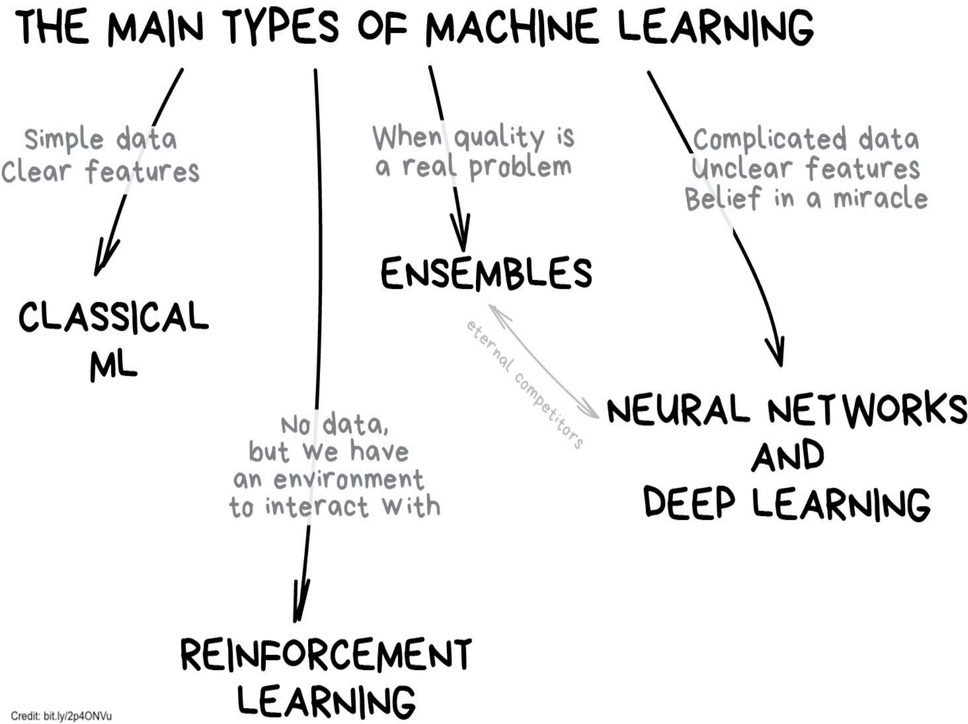
Summary of how machine learning algorithms work

- By searching through a set of potential models
- There are two sources of information that guide this search:
  1. the training data
  2. the inductive bias of the algorithm

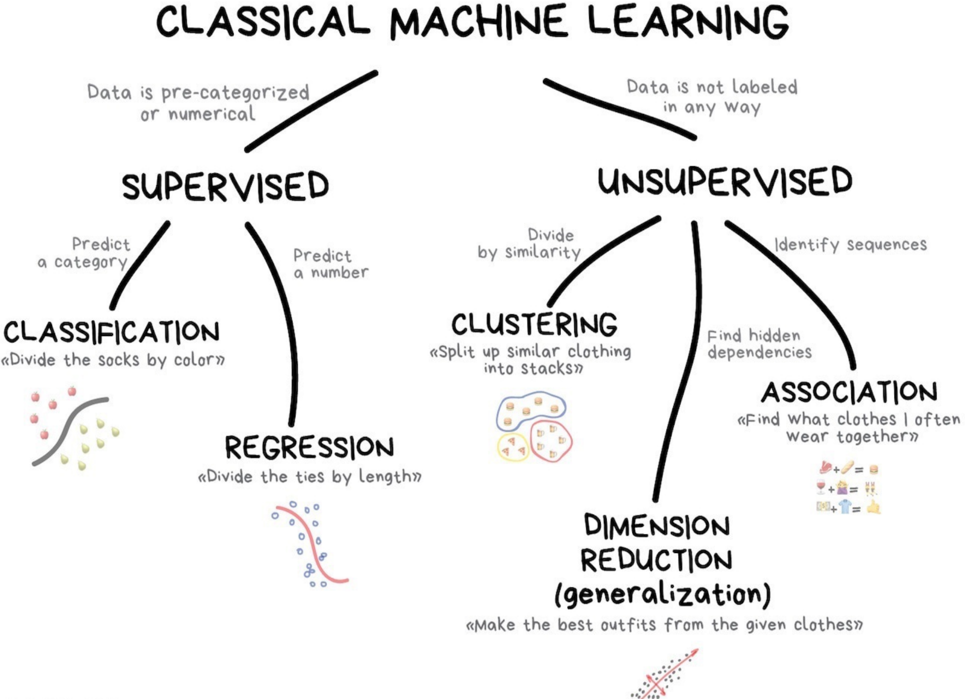


Credit: bit.ly/2p4ONVu

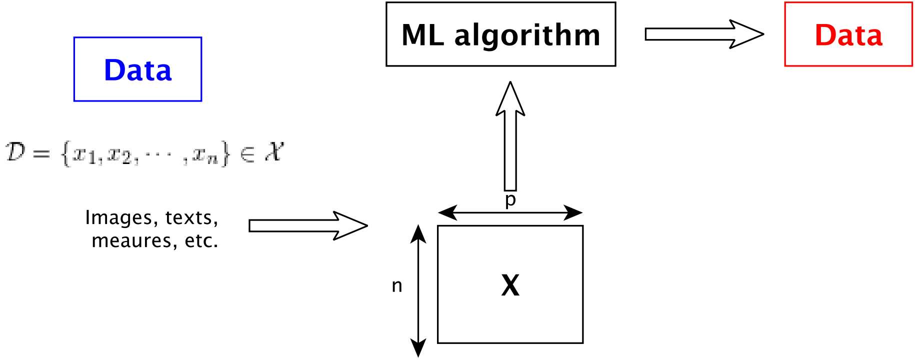
Types of Learning



Credit: bit.ly/2p4ONVu



Credit: bit.ly/2p4ONVu



Goal: learn a new **representation** of the data

$\mathcal{D} = \{x_1, x_2, \cdots, x_n\} \in \mathcal{X}$

Goal: group similar data points together

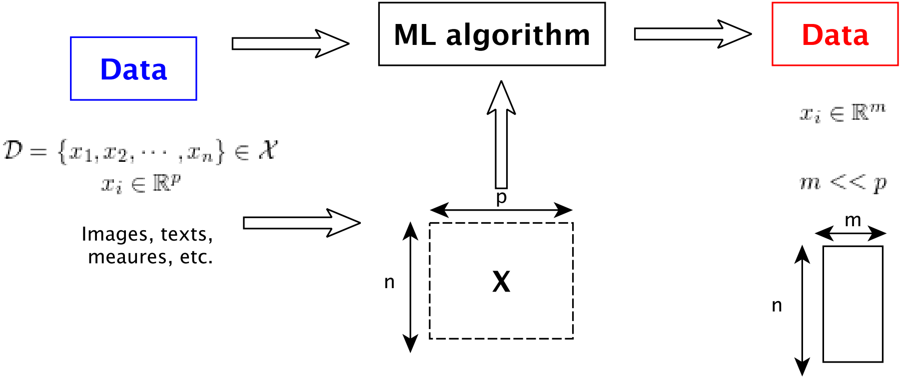
Useful to:

- Understand general characteristics of the data
- Infer some properties of an object based on how it relates to other objects

Clustering applications

- **Customer segmentation**: find groups of customer with similar buying behavior
- **Image compressions**: find groups of similar pixels that can be easily summarized
- **Topic modeling**: group document based on the words they contain to identify shared topics
- **Disease sub-typing**: find groups of similar patients with closed pathologies (e.g., symptoms level)

Dimensionality reduction

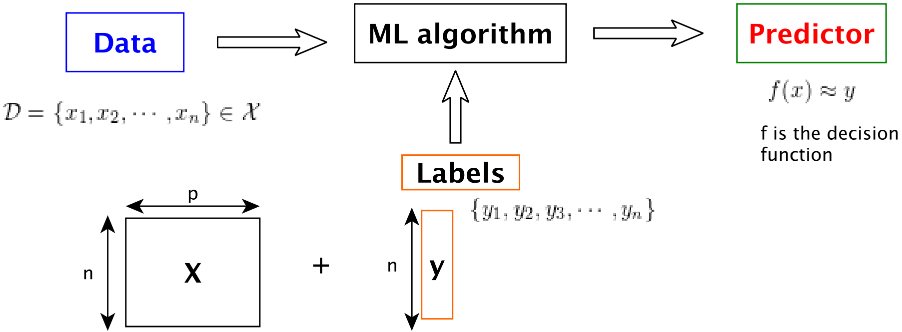


Goal: find a lower-dimensional representation of the data

Useful to:

- reduce storage and computing time
- remove redundancies
- Visualization of the data (e.g., in 2 or 3 dimensions) and interpretability

Supervised learning



Goal: make predictions

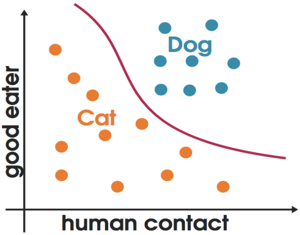
Data ⟹ **ML algorithm** ⟹ **Predictor**

$\mathcal{D} = \{x_1, x_2, \cdots, x_n\} \in \mathcal{X}$

Labels ⇑ (into ML algorithm)

$\{y_1, y_2, y_3, \cdots, y_n\}$

$f(x) \approx y$



Binary classification

$y_i \in \{0, 1\}$

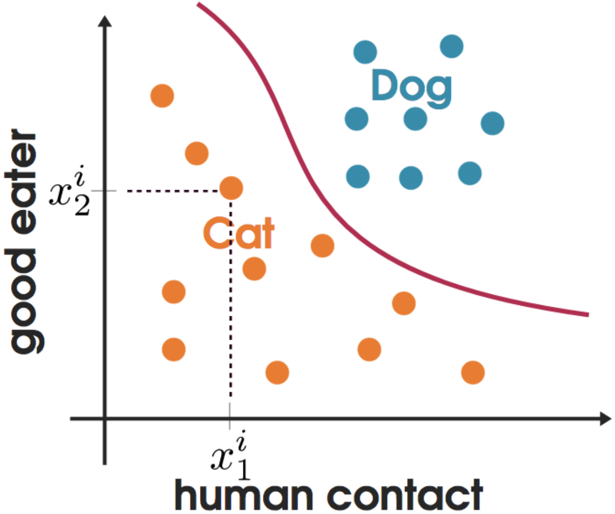
Multi-class classification

$y_i \in \{0, 1, \cdots k\}$

Goal: make discrete predictions

Classification

Given $\mathcal{D} = \{x_i, y_i\}_{i=1,\ldots}$, finds f such that $f(x) \approx y$



$$\mathcal{D} = \{x_i, y_i\}_{i=1,\ldots,n}$$

$$y_i = \begin{cases} 1 \text{ if } x_i \in \mathcal{P} \\ 0 \text{ if } x_i \in \mathcal{N} \end{cases}$$

$$x_i = \begin{pmatrix} x_{i,1} \\ x_{i,2} \end{pmatrix}.$$

Classification

- **Face recognition**: identify faces independently of pose, lighting, make-up, and hair style
- **Character recognition**: read letters or digits independently of handwriting styles
- **Spam detection**
- **Sound recognition**: which music is playing on? Who composed this music? Who is the singer?
- **Precision medicine**: does this sample come from a sick or from a healthy person?

Data ⟹ **ML algorithm** ⟹ **Predictor**

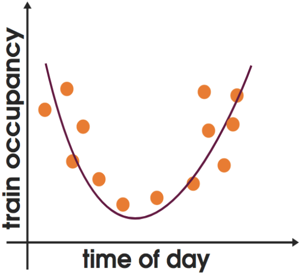
$\mathcal{D} = \{x_1, x_2, \cdots, x_n\} \in \mathcal{X}$

Labels

$\{y_1, y_2, y_3, \cdots, y_n\}$

$f(x) \approx y$

$y_i \in \mathbb{R}$



Goal: make continuous predictions

Regression's applications

- **Algorithmic trade**: what will be the price of this share?
- **Click prediction**: how many clicks will this ad receive? How many people will share this article on social media?
- **Electricity consumption**: Do I need to turn on this power plant?

Generalization

- Separate your data set into training and validation set
- It is usually easy to build a model that performs well on the training data
- But how well it performs on **new data**?
- Use **cross-validation** to assess that your model can generalize to independent data set
- The fundamental goal of machine learning is to generalize beyond the examples in the training set.

What Can Go Wrong With ML?

- No free lunch!
- What happens if we choose the wrong inductive bias:
  1. **underfitting**
  2. **overfitting**

Data alone is not enough

No Free Lunch theorem[1]

David Wolpert

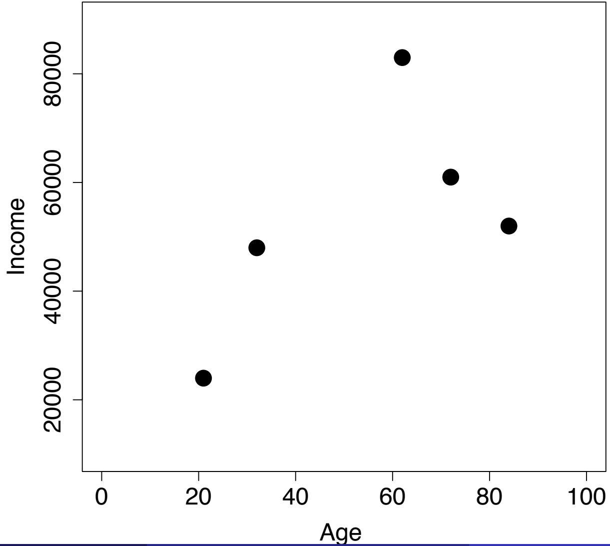
"For any two learning algorithms, there are just as many situations (appropriately weighted) in which algorithm one is superior to algorithm two as vice versa, accordingly to any of the measures of superiority"

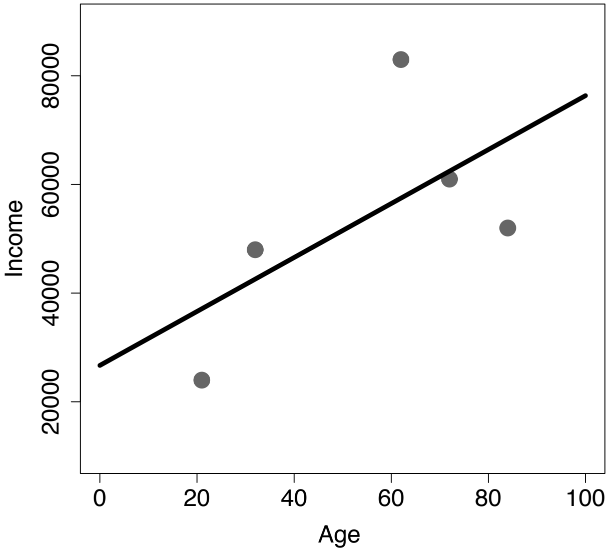
[1] David H. Wolpert. "The Lack of a Priori Distinctions Between Learning Algorithms". In: *Neural Computation* 8.7 (1996), pp. 1341–1390.

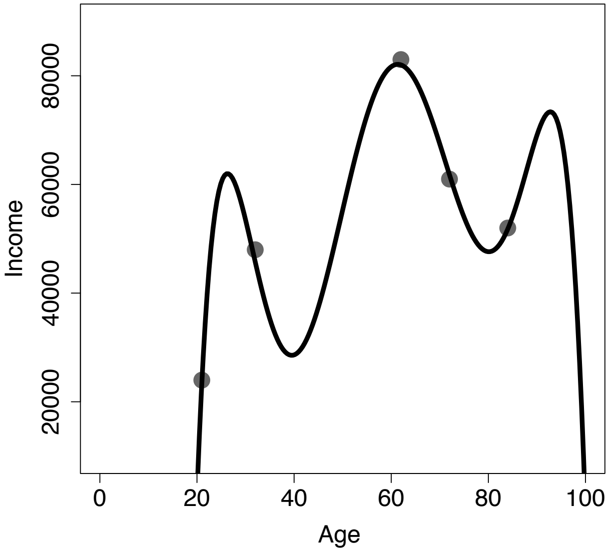
Table 1: An age-income dataset

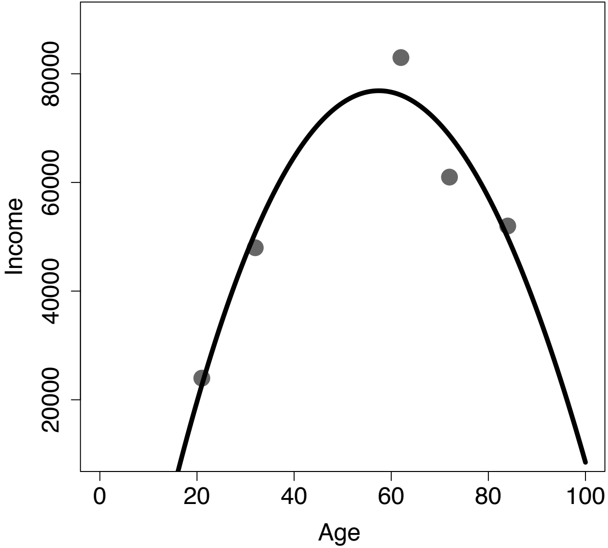
ID	AGE	INCOME
1	21	24,000
2	32	48,000
3	62	83,000
4	72	61,000
5	84	52,000

Plotting the age-income data set




Income
80000
60000
40000
20000
0
20
40
60
80
100
Age


Income
80000
60000
40000
20000
0
20
40
60
80
100
Age



Income
80000
60000
40000
20000
0
20
40
60
80
100
Age

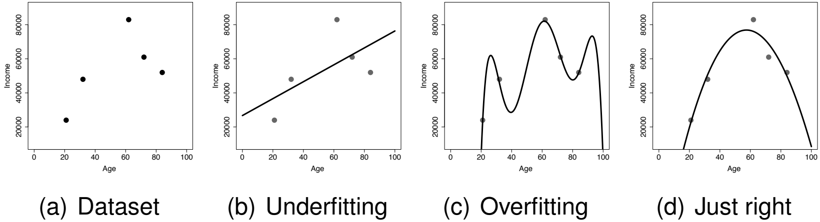
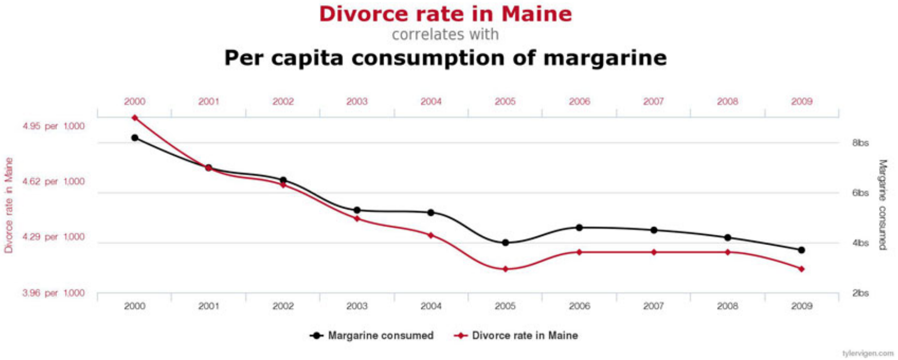
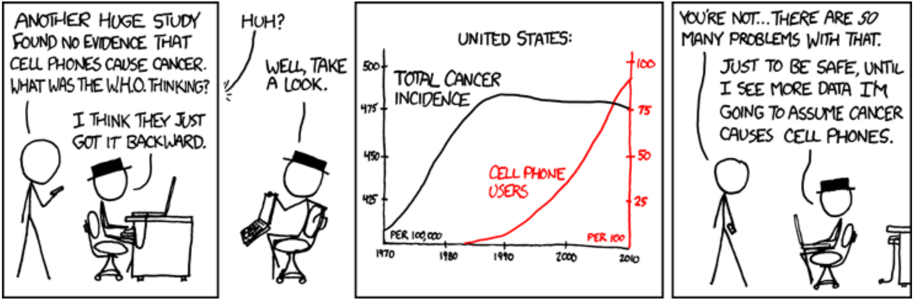


Figure 5: Striking a balance between overfitting and underfitting when trying to predict age from income

Correlation does not imply causation

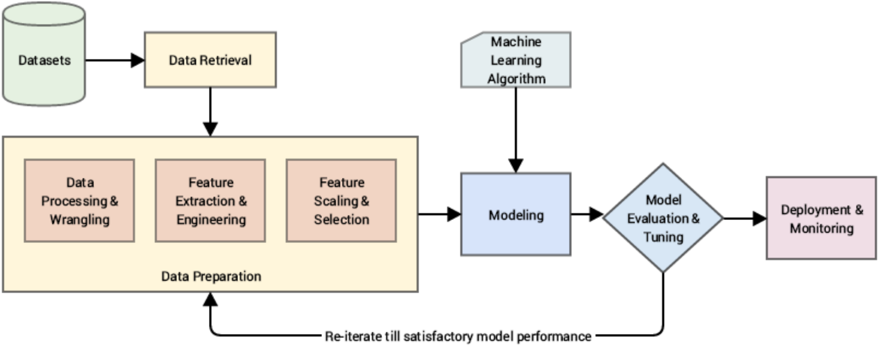


Correlation does not imply causation



Feature engineering is the key

- Raw data is rarely in a form that is amenable to learning
- Therefore, we can build features from it that are.



- Knowledge representation is a key concept in machine learning
- Programming, like all engineering, is a lot of work: everything need to be built from scratch
- Learning is more like farming
- Farmers combine seeds with nutrients to grow crops
- Learners combine knowledge with data to grow program

Summary

In summary

- Machine learning techniques automatically learn the relationship between a set of **descriptive features** and a **target feature** from a set of historical examples.
- Machine learning is an **ill-posed** problem:
  1. **generalize**
  2. **inductive bias**
  3. **underfitting**
  4. **overfitting**
- Striking the right balance between model complexity and simplicity (between underfitting and overfitting) is the hardest part of machine learning.

Learning objectives

Learning objectives

- Understanding what machine learning is; i.e., define machine learning
- Know the prominent methods used in contemporary machine learning
- Learn how to use machine learning correctly
- Given a problem:
  - **decide** whether it can be solved with machine learning
  - decide what type of machine learning technique you can use to formalize it (e.g., supervised — regression, classification, unsupervised — clustering, dimension reduction)
  - describe it formally in function of design matrix, features, samples, and target
- Define a loss function
- Define generalization

Bibliography

- Pedro Domingos. “A Few Useful Things to Know About Machine Learning”. In: *Communication of the ACM* 55.10 (2012), pp. 78–87
- D. Sculley et al. “Hidden Technical Debt in Machine Learning Systems”. In: *28th International Conference on Neural Information Processing Systems*. 2015, pp. 2503–2511
- Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2nd. Springer, 2016. URL: https://web.stanford.edu/~hastie/Papers/ESLII.pdf
- Leslie Valiant. *Probably Approximately Correct: Nature’s Algorithms for Learning and Prospering in a Complex World*. Basic Books, Inc., 2013
- Hal Daume III. *A Course in Machine Learning*. 2nd. Self-published, 2017. URL: http://ciml.info/dl/v0_99/ciml-v0_99-all.pdf
- Gareth James et al. *An introduction to statistical learning*. Springer, 2013. URL: http://www-bcf.usc.edu/~gareth/ISL/
- Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016. URL: deeplearningbook.org
- Dominique Cardon, Jean-Philippe Cointet, and Antoine Mazières. “La revanche des neurones”. In: *Reseaux* 5 (2018), pp. 173–220


WELCOME
ABOARD