Machine Learning

The Learning Problem

Kristóf Karacs

fides et ratio

Purpose of this course

How can machines learn?

Parts:

- Machine Learning *foundations* and *Supervised Learning*
- Brief theory on *Deep Learning*
- Glance at *Unsupervised Learning*
- An even briefer glance at *Reinforcement Learning*

Purpose of the "foundations" part

Based on Dr. Hsuan-Tien Lin's course at NTU

Main focus on *Supervised Learning*
(which we'll define in a couple of slides, hang on)

How can machines generalize to unseen data?

Today: introductory story
(Get comfortable, there is only *one* slide with maths.)

What is Machine Learning?

learning: acquiring *skill* from *observations*

$$\text{observations} \longrightarrow \boxed{\text{learning}} \longrightarrow \text{skill}$$

Compare this to procedural algorithms

What is Machine Learning?

machine learning: acquiring skill from observations

improving some performance measure

$$\text{observations / data} \longrightarrow \boxed{\text{ML}} \longrightarrow \text{skill}$$

Compare this to procedural algorithms

An example of learning

Tree recognition from images

- how do you define what is a tree?
- easier to show examples
- large number of images to learn from: large dataset

Tree

Not a tree

Source: pexels.com

Fun Time

Which of the following is best suited for ML?

1. predicting whether the next cry of the baby girl happens at an even-numbered minute or not
2. determining whether a given graph contains a cycle
3. deciding whether to approve credit card to some customer
4. guessing whether the earth will be destroyed by the misuse of nuclear power in the next ten years

Fun Time

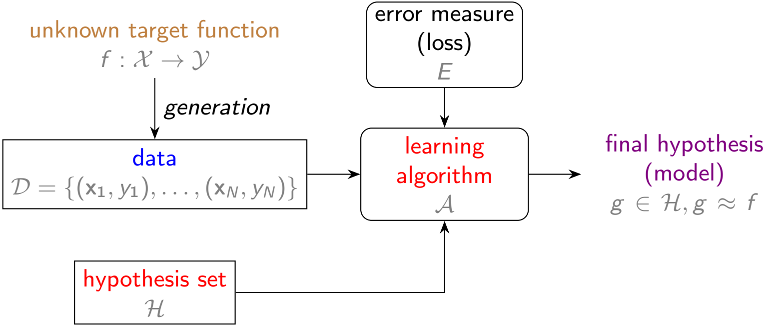
Which of the following is best suited for ML?

1. predicting whether the next cry of the baby girl happens at an even-numbered minute or not
2. determining whether a given graph contains a cycle
3. deciding whether to approve credit card to some customer
4. guessing whether the earth will be destroyed by the misuse of nuclear power in the next ten years

Reference answer: 3

1. no pattern to learn
2. programmable definition, easier to do procedurally
3. there is a pattern, there is data, alternatives are complex rules
4. arguable no (not enough) data

Formalization of learning problem



- input: $\mathbf{x} \in \mathcal{X}$
- output: $y \in \mathcal{Y}$
- target function: $f : \mathcal{X} \rightarrow \mathcal{Y}$
- data: $\mathcal{D} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)\}$
- hypothesis / model with good performance: $g = \arg\max_f E(f)$

Related fields

Machine Learning $\approx$ Data Mining

- ML: use data to compute hypothesis g that approximates f
- DM: use (huge) data to find interesting properties
- DM: more data cleaning, processing, storing, organising, and also feature engineering

Machine Learning $\subset$ Artificial Intelligence

- AI: compute something that shows intelligent behaviour
- ML is one possible way to AI

Machine Learning vs Statistics

- Stat: use data to make inference about an unknown process
- useful tools for ML (with nice mathematical proofs)

Related fields

Machine Learning $\approx$ Data Mining

- ML: use data to compute hypothesis g that approximates f
- DM: use (huge) data to find interesting properties
- DM: more data cleaning, processing, storing, organising, and also feature engineering

Machine Learning $\subset$ Artificial Intelligence

- AI: compute something that shows intelligent behaviour
- ML is one possible way to AI

Machine Learning vs Statistics

All the impressive achievements of deep learning amount to just curve fitting.

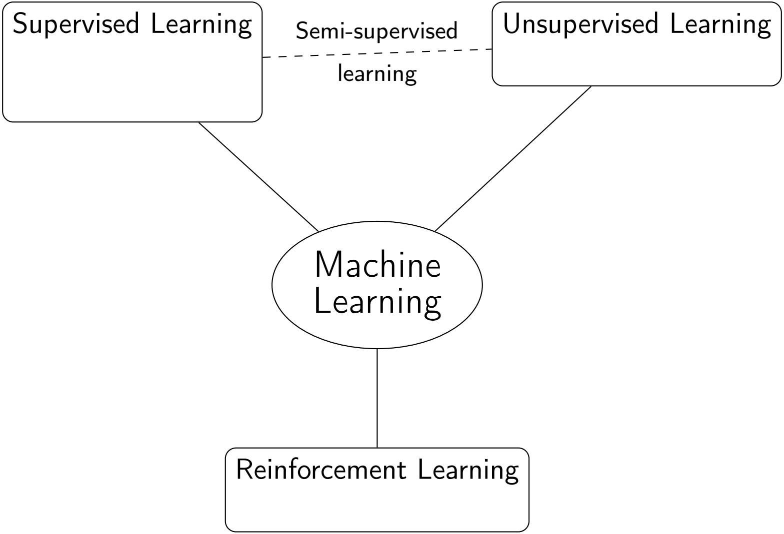
Judea Pearl

Categorization by spaces

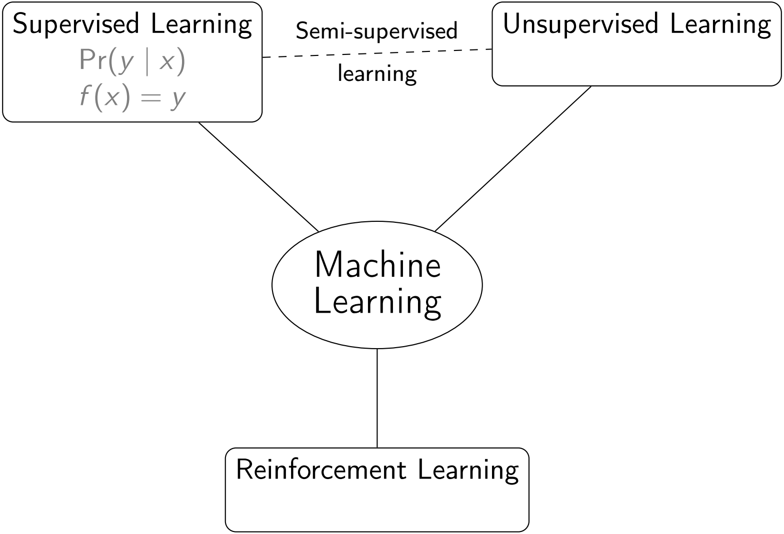
Input/output can be

- discrete (e.g. $\{\text{blue}, \text{black}, \text{brown}\}$)
- ordered (e.g. $\{\text{healthy} < \text{ill} < \text{extremely ill} < \text{horrendously(?) ill} < \text{dead}\}$)
- real (e.g. $\mathbb{R}^d$)
- sequence (e.g. "I love ML", or (healthy, ill, healthy, extremely ill, dead))

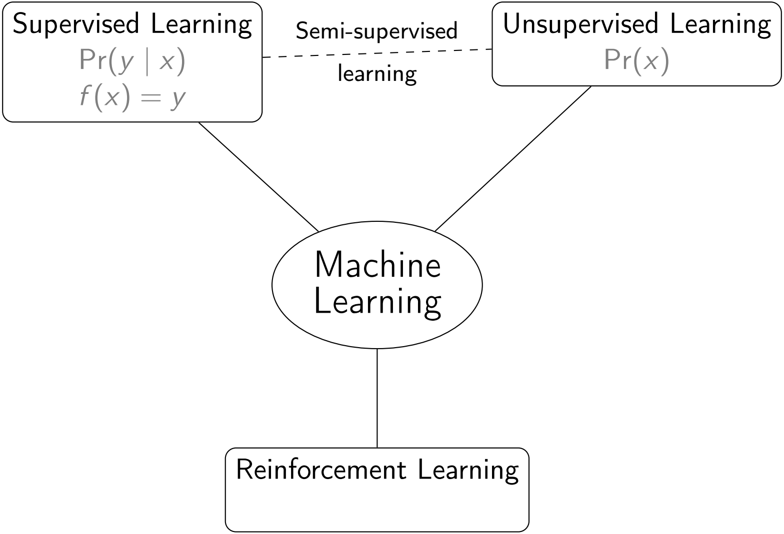
Categorization by problem type



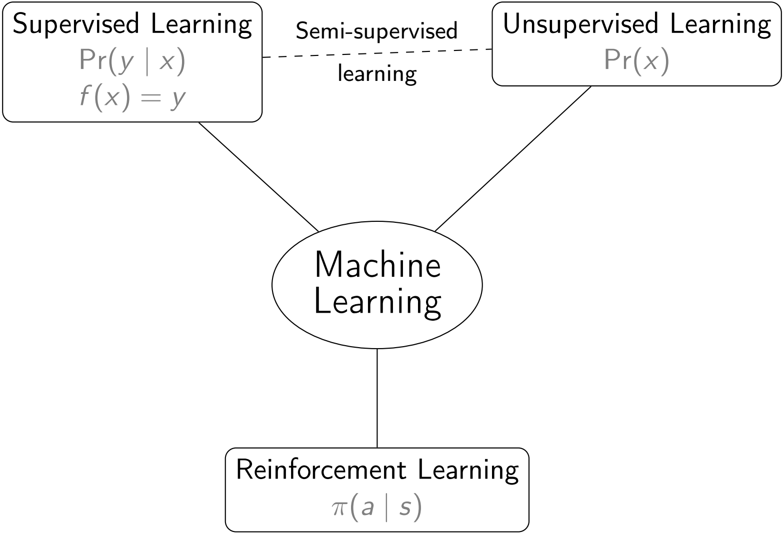
Categorization by problem type



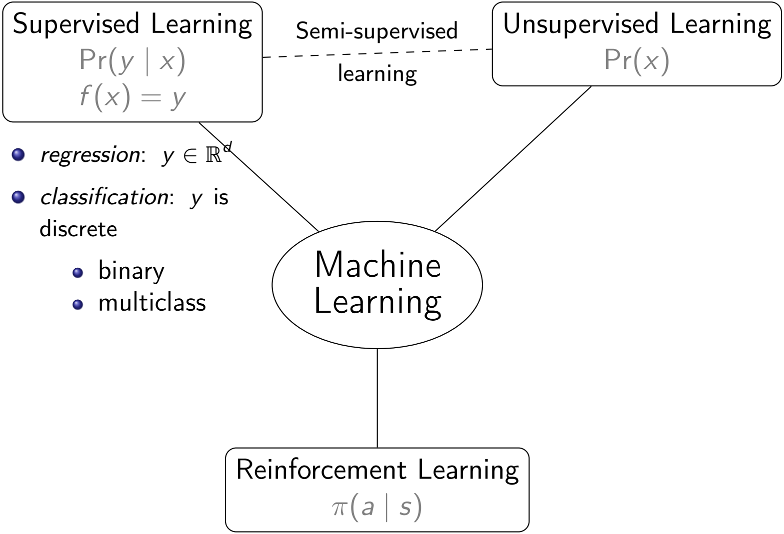
Categorization by problem type



Categorization by problem type

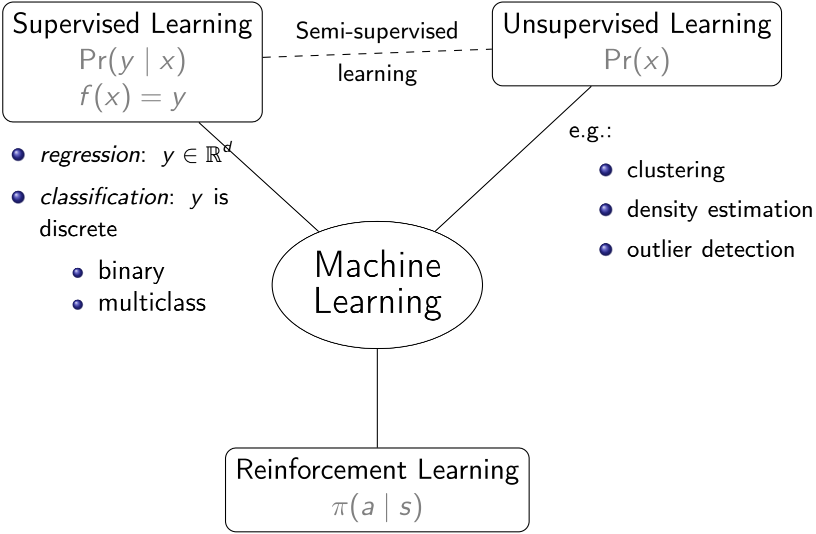


Categorization by problem type



- *regression*: $y \in \mathbb{R}^d$
- *classification*: y is discrete
    - binary
    - multiclass

Categorization by problem type



Categorization by context

- Batch / offline learning
  - learn from all known data
  - dataset does not change during learning
- Online learning
  - new data is received during learning (and potentially at inference time)
  - continuously improve hypothesis
  - e.g. spam filtering
- Active learning
  - the model (agent) asks for new data to learn from
  - strategically observed data