Machine Learning

Dr. Peter Arndt & Dr. Konrad Völkel,
Wintersemester 2022/23


Dr. Peter Arndt,
Dr. Konrad Völkel

peter.arndt@hhu.de

konrad.voelkel@hhu.de #ml-2022-23

Gebäude 25.12

Raum 01.24 office hours by appointment

special arrangement

First half:
Dr. Peter Arndt,
Second half:
Dr. Konrad Völkel

For the whole time, you are supported by our tutors:

Liz Leutner, Christopher Orlowicz, Leon Erbrich, Maïwenn Fleig and Anthony Nascente.

We all speak English and German

Today's topics:

Heinrich Heine
Universität Düsseldorf

Machine Learning,
what is this?

What do we do in the lectures?

Where will we end up?

Which skills are we obtaining along the way?

What should students do?
(Exercise sheets,
Exercise classes,
Exam)

What do you need?

What do you wish for?

Why are you here?

Machine Learning

What is it?

“Steam engine machine learning by Magritte”

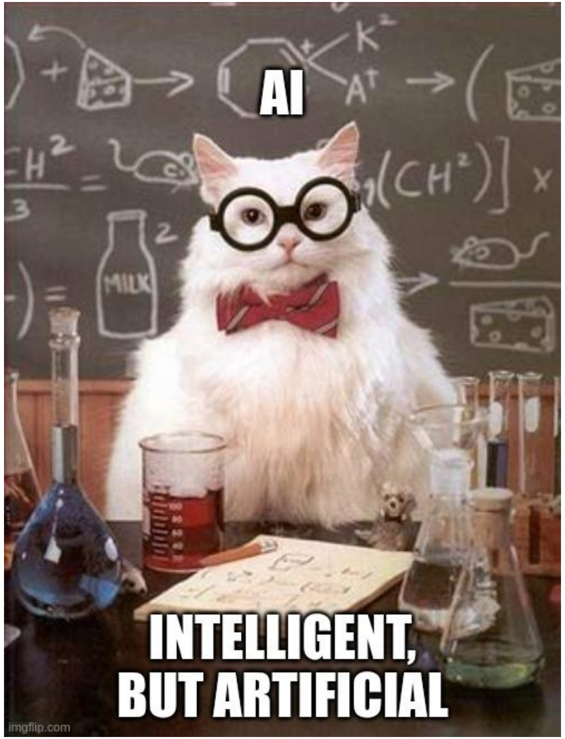
What is Machine Learning?

Heinrich Heine Universität Düsseldorf

Definition “Artificial Intelligence”

Intelligence usually means “the ability to solve hard problems” (Minsky, 1985).

“Intelligence is the capacity of an information-processing system to adapt to its environment while operating with insufficient knowledge and resources.” (Wang, 1995)

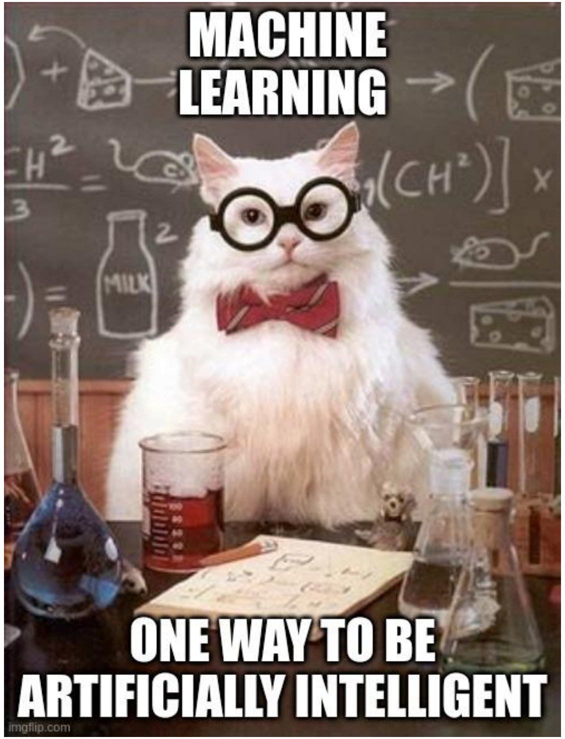


"Artificial intelligence (AI) systems are software [...] systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected [...] data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. (EU AI HLEG, 2019).

What is Machine Learning?

Machine Learning in contrast to Artificial Intelligence

“Machine learning is an application of AI.
It’s the process of using mathematical models of data to help a computer learn without direct instruction. This enables a computer system to continue learning and improving on its own, based on experience.”
 - Microsoft Research



“In just the last five or 10 years, machine learning has become [...] the most important way, most parts of AI are done.

So that's why some people use the terms AI and machine learning almost as synonymous … most of the current advances in AI have involved machine learning.”

Malone 2021 (MIT Center for Collective Intelligence).

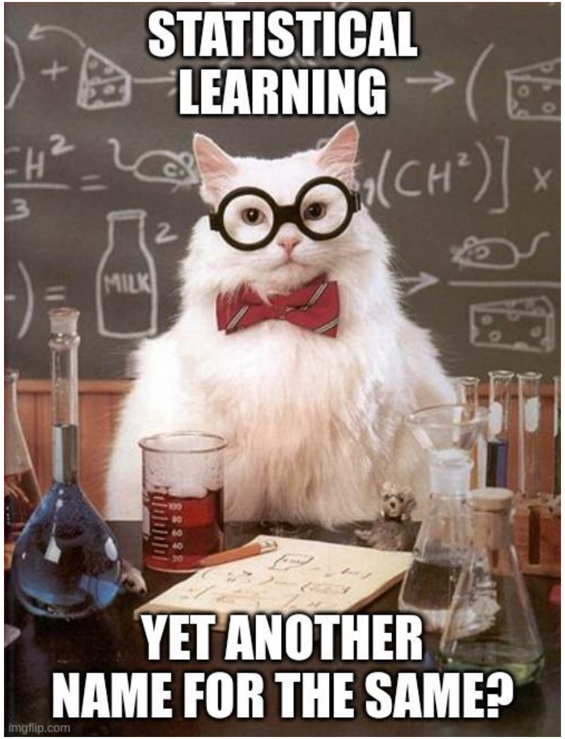
What is Machine Learning?

Definition “Statistical Learning”

A statistical model defines a relationship from a dependent to an independent variable.

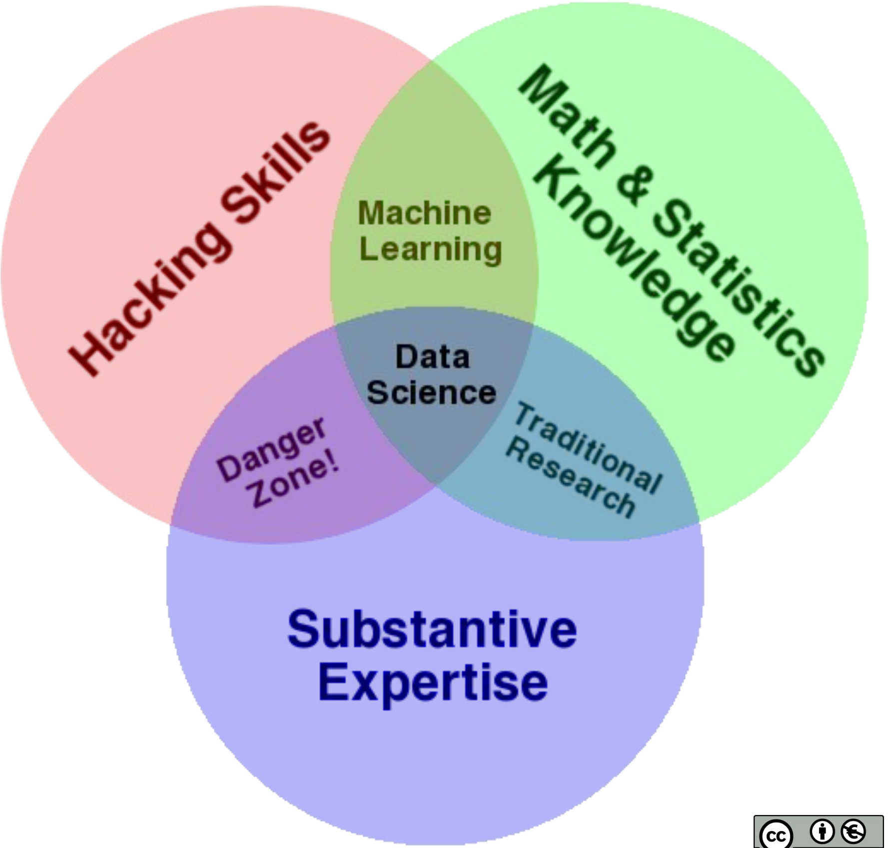
The relationship and the variables are stochastic, not deterministic.

Usually, such a model has one or several parameters - real numbers that can be changed to improve (or worsen) the model.



A key approach is ERM: empirical risk minimization. The model parameters are chosen to minimize the risk of being empirically wrong (so you need data = empirical evidence)

Another key concept is PAC: Probably approximately correct. We don’t want to have a perfect, precise always correct model.



A part of Data Science?

Heinrich Heine Universität Düsseldorf

(one way to see it)

"Venn-Diagramm" by Drew Conway, 2013

Many people say you can start using ML without any maths background.

To make it useful, most applications need a proper maths and stats background.

Many self-taught ML practitioners say they wish they had more maths education.

In contrast to Data Science

How a Data Scientist spends their time

"Nearly all [...] working data scientists make their daily bread and butter through **data collection** and data **cleaning**; building **dashboards** and **reports**; data **visualization**; statistical **inference**; **communicating** results to key stakeholders; and **convincing** decision makers of their results."

Hugo Bowne-Anderson, Harvard Business Review, 2018

In these lectures, we focus on algorithms for machine learning

What is Machine Learning?

An example of statistical learning: Regression

Regression models a relationship between two variables X and Y as a functional relation Y = f(X) + error terms

Example: Y = income and X = age

What is Machine Learning?

An example of statistical learning: Regression

Regression models a relationship between two variables X and Y as a functional relation Y = f(X) + error terms

Example: Y = income and X = age

Usage example:

```
from sklearn.linear_model import LinearRegression
regression = LinearRegression()
regression.fit(X_train, y_train)
y_prediction = regression.predict(X_test)
```

What is Machine Learning?

Heinrich Heine Universität Düsseldorf

An example of statistical learning: Regression

We will learn the mathematical description of regression to **understand** variants, problems and solution approaches.

We want to be able to **implement ourselves**: `fit` and `predict`

Usage example:

```
from sklearn.linear_model import LinearRegression
regression = LinearRegression()
regression.fit(X_train, y_train)
y_prediction = regression.predict(X_test)
```

What is Machine Learning?

An example of statistical learning: Regression

After some more overview and administration stuff about the course, we take a look inside the black box!

Overview of the lectures

Where are we going?

DALL·E 2 outpainting of a campus photo

What we don’t do in this lecture series

Heinrich Heine Universität Düsseldorf

A list of excluded topics

Deep learning

Big data, parallelization, streams

Databases, SQL

Programming, Python (or R, Julia, etc.)

Advanced Statistics

Cleaning data with Pandas

Data visualization, Plotting

Productizing existing ML systems

Running MLOps pipelines

Overview of the lectures

Skills (learning outcomes)

After finishing this module, you should be able to

1. understand the theoretical foundations of machine learning
2. explain the theoretical foundations of machine learning
3. explain the foundations in mathematical terms
4. and do proofs about it
5. implement algorithms of machine learning
6. apply algorithms of machine learning

Overview of the lectures

Why Python?

- "Most used programming languages among developers worldwide, as of 2021: JavaScript, HTML/CSS, Python" - statista
  Rank 1 in TIOBE Index, March 2022 - tiobe
- Working ecosystem
  Scientific Python:
  - Numpy
  - SciPy, Scikit-Learn
  - Matplotlib, Seaborn
  - Pandas
  - Jupyter
- We use Python >= 3.8
  (see also NEP 29)

Overview of the lectures

Heinrich Heine
Universität Düsseldorf

Which topics do we learn to reach our learning outcomes?

- Background from Statistics:
  - Probability, frequentist statistics, **Bayesian statistics**
- Categories of machine learning algorithms:
  - Supervised and unsupervised learning
  - Generative vs. discriminative models
- Specific algorithms:
  - Linear regression
  - linear discriminant analysis
  - Gaussian processes
  - Support vector machines
- Kernel trick, kernel PCA
- Further topics: Graphical models, neural networks

Overview of the lectures

(tentative and preliminary plan)

Overview and introduction

Recalling the basics of probability theory:
Density functions, Gaussians

The basics of Scientific Python, Numpy

Maximum likelihood

Exponential families

Linear regression, Logistic regression

Bayesian linear regression

k-Means

Support Vector Machines (SVM)

Principal Component Analysis (PCA)

Nonlinear dimensionality reduction (ISOMAP)

K-nearest neighbours

Gaussian mixed models, Expectation maximization

Gaussian Processes

Gradient descent

Trees

Boosting

Statistical Learning Theory

No free lunch theorem

VC-dimension

Sampling, Monte Carlo

What do you have to do?

To be admitted to the exam and obtain a good grade

“Chewbacca sitting in the classroom drinking a cafe latte”

What do you have to do?

1) Exercise classes

We recommend weekly participation in the exercise classes, but this is not technically required.

Working on the weekly exercise is considered the most important part of this module.

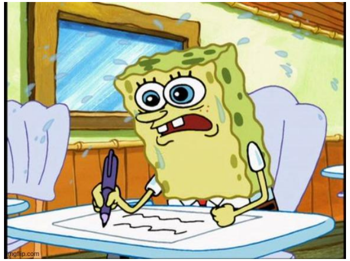
We intend to have 50% theory exercises and 50% practical exercises.

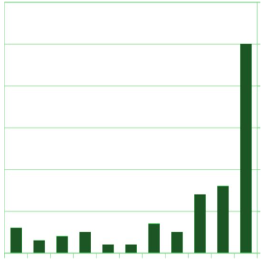
You have to hand in your exercise solutions in pairs - so go and find a partner (right now??)!

2) Written exam

First Date:

14.02.2023 from 11:30 to 14:30
in 5B and 3H





Second Date:

27.03.2023 von 11:30 bis 14:30
in 5D

What do you have to do?

Exercise groups

Th 8:30 -- 10:00 online via WebEx (Liz)

Th 12:30 -- 14:00 in 2C (Leon)

Fr 12:30 -- 14:00 in 2E (Chris)

Attention: in LSF, there is another group on Fridays 10:30, which overlaps with the AI Master math lectures, so we don't use that one.

Choose as you wish; you can change groups without notifying anyone.

This is going to happen there

- You can discuss any questions at length
- The solutions to all exercises can be discussed (however you wish)
- Live programming help, if needed
- Repeating material from the lectures that was difficult
- Direct feedback on your solutions, if the commentary via Rocket.Chat is not enough

What do you have to do?

Exercise sheets
(German: Übungsblätter)

Hand in your solutions before the deadline every Tuesday at 10:00 -- later won’t be accepted (upload not even possible, hand in by email is not allowed).

The first exercise sheet comes on Tuesday 10:00 and you have exactly one week to solve it.

What do you have to do?

Work together

Solve the exercises alone **and** in groups. You will have to hand in the solutions in pairs, but larger collaborations are often also very fruitful. Make sure to have a good study group early on.

You are free to use the Rocket.Chat channel to coordinate with others to form study groups and exercise hand-in pairs.

File format

Always put your name on the solution.

The filenames should always be coded like this:

username1_username2_sheet7.pdf

Instead of .pdf you can also have .ipynb files, but please hand them in also as a pdf printout (so we won’t have problems with non-compiling notebooks etc.)

What do you have to do?

Admission to the exam

To be admitted to the exam,
you need to be successful in the exercises in the sense that you have to obtain half of the points in each single exercise.

You are only allowed to deviate from this twice (use this for sickness).

Do not discuss points with us until the very end of the module.

Again: it is not enough to have 50% of the obtainable points overall!

What do you need?

We want to know you and help you reach your potential

“Solution to the Voynich manuscript”

Questions for you - 1

Raise your hands!

A) Who is in the first year of the AI Master?

B) Who is in the AI Master, but no longer in the first year?

C) Who is in the computer science bachelor and took the class "Data Science" before?

D) Who is in the computer science bachelor and took the class "Machine Learning" before?

E) Who is in the computer science bachelor and a complete newbie to the topic of machine learning?

Questions for you - 2

Raise your hands!

A) Do you know how to program?

B) Did you program in Python before?

C) Did you work with Numpy before?

Questions for you - 3

What motivates you to take this class?

A) I want to become rich (or at least make a decent living)

B) I am very interested in artificial intelligence

C) I want to mix ML with my other interests

D) Other reason (it’s complicated)

Questions for you - 4

What is most interesting to you (choose only one!)?

A) Abstract theory (might as well pursue PhD later), i.e.

Why does it work, how does it work inside?

B) Implementing/programming stuff in practice, i.e.

How do I make it work, how can I use this / apply this?

What do you need?

What did you find difficult in the past?

Raise your hands, multiple choices are allowed!

During my studies, I found challenging:

A. Weekly exercise sheets
B. Written exams
C. Programming
D. Mathematical proofs
E. Not losing track over the entire term
F. Preparing for exams
G. When it goes too fast in the lecture
H. When I get bored in the lecture

What do you need?

Heinrich Heine
Universität Düsseldorf

Which expectations and wishes do you have?

Raise your hand!

- Can you come up with a topic you want to learn which we did not mention yet?

- Do you know any specific teaching or learning method that you would like to use in this class?

What do you need?

Heinrich Heine
Universität Düsseldorf

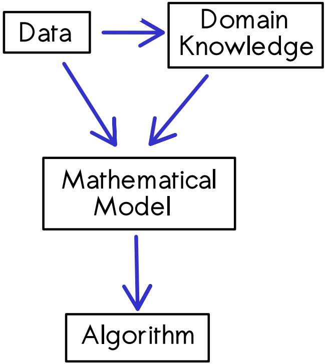
Keep in mind: we want you to succeed

In any doubt, please

- Ask once more rather than not. There may be stupid questions, but for every stupid question asked, there are 10 students happy that someone asked their question. Help them. Ask.
  And then ask again, if you still didn’t get it.
- After the lecture:
  Reach out to your classmates and tutors and lecturers via Rocket.Chat. We will all try to answer as soon as possible (don’t expect tutors and lecturers on the weekend).

Now we look into the machine learning black box

Usual programming/ engineering versus Machine Learning



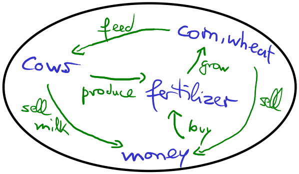
Usual programming/engineering versus Machine Learning

Data → Domain Knowledge

Data, Domain Knowledge → Mathematical Model

Mathematical Model → Algorithm

Example: Farm



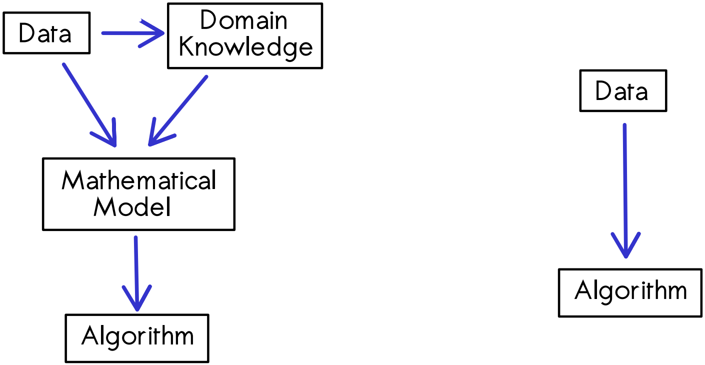
↓

$Cow + \sqrt{Fert^{17}} - sin(Wheat)...$

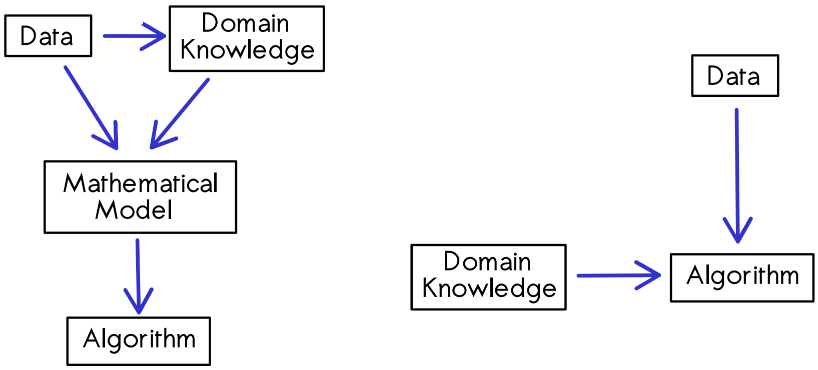
↓

need to sell corn for y €/t
buy x tons of fertilizer...

Usual programming/engineering versus Machine Learning



Usual programming/engineering versus Machine Learning



Example：Distinguishing Hamsters and Lions

Example: Distinguishing Hamsters and Lions

Want a map sending jpg–images to $\{0,1\}$

training set

Suppose we have a collection of examples
where we know what the outome should be:

0

1

Now we can learn a map $\{\text{jpegs}\} \longrightarrow \{0,1\}$ that
(approximately) behaves like this on our training set

Example: Distinguishing Hamsters and Lions

Want a map sending jpg-images to $\{0,1\}$

Suppose we have a collection of examples where we know what the outome should be:

$\}\ 0$

$\}\ 1$

Now we can learn a map $\{\text{jpegs}\} \longrightarrow \{0,1\}$ that (approximately) behaves like this on our training set,

... and hope that it also performs well on new examples.

How to learn that map:

$$\mathbb{R}^{512\times 512\times 3} \cong \{\text{.jpeg}_s\} \longrightarrow \{0,1\}$$

We need to combine the n:= $512 \times 512 \times 3$ numbers and produce either 0 or 1.

One way to build such a map:

- Fix $z_1, \ldots, z_n \in \mathbb{R}$ and define $g_{z_1,\ldots,z_n} : \mathbb{R}^n \to \mathbb{R}$

$$(x_1, \ldots, x_n) \mapsto z_1 \cdot x_1 + \ldots + z_n \cdot x_n$$

How to learn that map:

$$\mathbb{R}^{512\times 512\times 3} \cong \{\text{.jpegs}\} \longrightarrow \{0,1\}$$

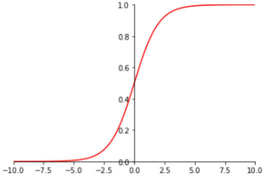
We need to combine the n:= $512 \times 512 \times 3$ numbers and produce either 0 or 1.

One way to build such a map:

- Fix $z_1, \ldots, z_n \in \mathbb{R}$ and define $g_{z_1,\ldots,z_n} : \mathbb{R}^n \to \mathbb{R}$

$$(x_1, \ldots, x_n) \mapsto z_1 \cdot x_1 + \ldots + z_n \cdot x_n$$

- Then apply the function $h(x) := \dfrac{1}{1+e^{-x}}$



How to learn that map:

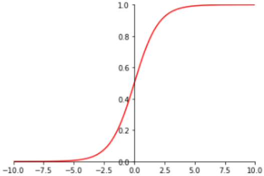
$$\mathbb{R}^{512\times 512\times 3} \cong \{\text{.jpegs}\} \longrightarrow \{0,1\}$$

We need to combine the n:= $512 \times 512 \times 3$ numbers and produce either 0 or 1.

One way to build such a map:

- Fix $z_1, \ldots, z_n \in \mathbb{R}$ and define $g_{z_1,\ldots,z_n} : \mathbb{R}^n \to \mathbb{R}$
$$(x_1, \ldots, x_n) \mapsto z_1 \cdot x_1 + \ldots + z_n \cdot x_n$$

- Then apply the function $h(x) := \dfrac{1}{1+e^{-x}}$



- Then round to 0 or 1.

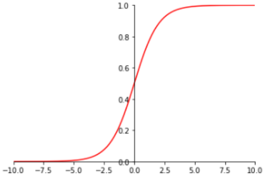
How to learn that map:

$$\mathbb{R}^{512\times 512\times 3} \cong \{.\text{jpeg}_s\} \longrightarrow \{0,1\}$$

We need to combine the n:= $512 \times 512 \times 3$ numbers and produce either 0 or 1.

One way to build such a map:

- Fix $z_1, \ldots, z_n \in \mathbb{R}$ and define $g_{z_1,\ldots,z_n} : \mathbb{R}^n \to \mathbb{R}$
$$(x_1, \ldots, x_n) \mapsto z_1 \cdot x_1 + \ldots + z_n \cdot x_n$$
- Then apply the function $h(x) := \dfrac{1}{1+e^{-x}}$



- Then round to 0 or 1.

Hope: The true function is close to one of those.

Try to find the best $z_1, \ldots, z_n \in \mathbb{R}$

How to learn that map:

$$\mathbb{R}^{512\times 512\times 3} \cong \{\text{jpegs}\} \longrightarrow \{0,1\}$$

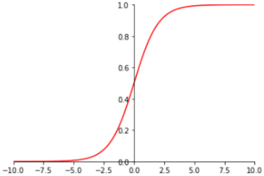
We need to combine the n:= $512 \times 512 \times 3$ numbers and produce either 0 or 1.

One way to build such a map:

- Fix $z_1, \ldots, z_n \in \mathbb{R}$ and define
$$g_{z_1,\ldots,z_n} : \mathbb{R}^n \to \mathbb{R}$$
$$(x_1, \ldots, x_n) \mapsto z_1 \cdot x_1 + \ldots + z_n \cdot x_n$$

- Then apply the function $h(x) := \dfrac{1}{1+e^{-x}}$

- ~~Then round to 0 or 1.~~



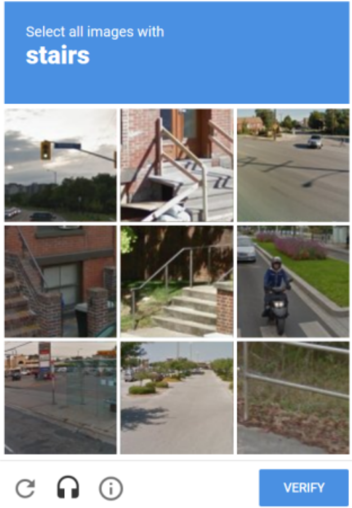
Interpret $f(x_1, \ldots, x_n \mid z_1, \ldots, z_n) := h(g_{z_1, \ldots, z_n}(x_1, \ldots, x_n))$
as $\boxed{\text{probability}}$, or degree of confidence, that $x_1, \ldots, x_n$ encodes a lion.

Interpret our training set as results of independent random draws of hamster/lion labels, according to this probability.

Why probability?

Data is uncertain.

Where does Data come from?



Interpret $f(x_1, \ldots, x_n \mid z_1, \ldots, z_n) := h(g_{z_1,\ldots,z_n}(x_1, \ldots, x_n))$
as $\boxed{\text{probability}}$, or degree of confidence, that $x_1, \ldots, x_n$
encodes a lion.

Interpret our training set as results of independent random draws of hamster/lion labels, according to this probability.

Why probability?

Data is uncertain.

E.g. the same picture could have been classified as lion

by 3 persons and as hamster by 1 person.

$\rightsquigarrow$ declare it a lion with probability $\frac{3}{4}$

Interpret $f(x_1, \ldots, x_n \mid z_1, \ldots, z_n) := h(g_{z_1,\ldots,z_n}(x_1, \ldots, x_n))$
as probability, or degree of confidence, that $x_1, \ldots, x_n$ encodes a lion.

Interpret our training set as results of independent random draws of hamster/lion labels, according to this probability.

Find those $z_1, \ldots, z_n \in \mathbb{R}$ that make the observations in the training set most likely – i.e. that maximize the likelihood function:

$$\ell(z_1, \ldots, z_n) := \Pi_{(x,1)\in D} f(x \mid z_1, \ldots, z_n) \cdot \Pi_{(x,0)\in D} (1 - f(x \mid z_1, \ldots, z_n))$$

Interpret $f(x_1, \ldots, x_n \mid z_1, \ldots, z_n) := h(g_{z_1,\ldots,z_n}(x_1, \ldots, x_n))$ as $\boxed{\text{probability}}$, or degree of confidence, that $x_1, \ldots, x_n$ encodes a lion.

Interpret our training set as results of independent random draws of hamster/lion labels, according to this probability.

Find those $z_1, \ldots, z_n \in \mathbb{R}$ that make the observations in the training set most likely – i.e. that maximize the likelihood function:

$$\ell(z_1, \ldots, z_n) := \Pi_{(x,1)\in D} f(x \mid z_1, \ldots, z_n) \cdot \Pi_{(x,0)\in D}(1 - f(x \mid z_1, \ldots, z_n))$$

Then we choose our hamster/lion prediction function to be

$$\mathbb{R}^n \to \mathbb{R}, \quad x \mapsto f(x \mid z_1, \ldots, z_n)$$

Did we really only use the data to arrive at this function?

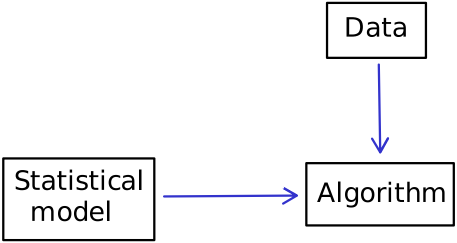
Did we really only use the data to arrive at this function?

No: We chose a family of functions, and then searched for the function in that family that explained the data best.

Did we really only use the data to arrive at this function?

No: We chose a family of functions, and then searched for the function in that family that explained the data best.

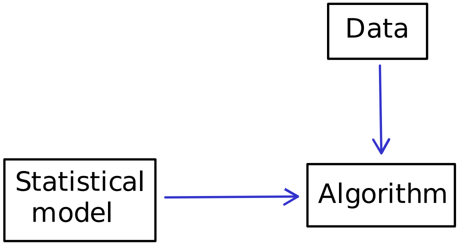
A statistical model is a family of probability distributions



Did we really only use the data to arrive at this function?

No: We chose a family of functions, and then searched for the function in that family that explained the data best.

A statistical model is a family of probability distributions



We need to know probability theory to do Machine Learning

Machine Learning

Lecture 1: Introduction

Stefan Harmeling

8. October 2018

Computer programming

Simple example

input

`[17, 42, 4, 1, 5]`

unsorted sequence

program →

output

`[1, 4, 5, 17, 42]`

sorted sequence

Hard example

input

image of handwritten digit

? →

output

7

digit

Machine learning

Hard example

input

image of handwritten digit

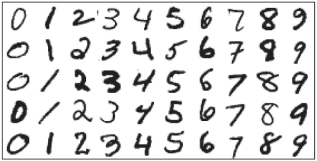
?

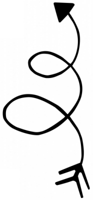
output

7

digit

Input/output examples





0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9
0 1 2 3 4 5 6 7 8 9

Fascinating! And also extremely useful today because...

Prof. Stefan Harmeling / HHU / 3

Big data

We are drowning in information and starving in knowing.
John Naisbitt

- 40 billion web pages
- 100 hours of video uploaded every minute on youtube
- 1000s of genomes sequences (each having about 3.8e9 base pairs)
- bigger telescopes generate each night terabytes of data
- ...

Everybody needs automated data analysis
(which is what machine learning does)

Types of machine learning

Supervised learning:

- discrete output: classification
- continuous output: regression
- ..., e.g. graphs or other structures

Unsupervised learning:

- discrete output: clustering
- continuous output: dimensionality reduction

Semi-supervised learning:

- ...

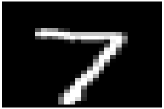
Reinforcement learning:

- (different setup)...

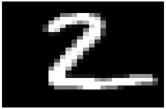
Examples: classification

- recognizing handwritten digits

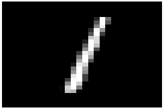
true class = 7



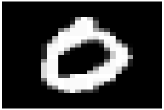
true class = 2



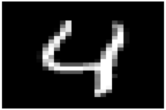
true class = 1



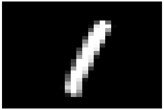
true class = 0



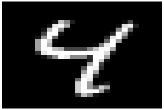
true class = 4



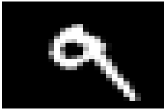
true class = 1



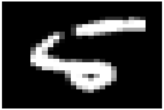
true class = 4



true class = 9

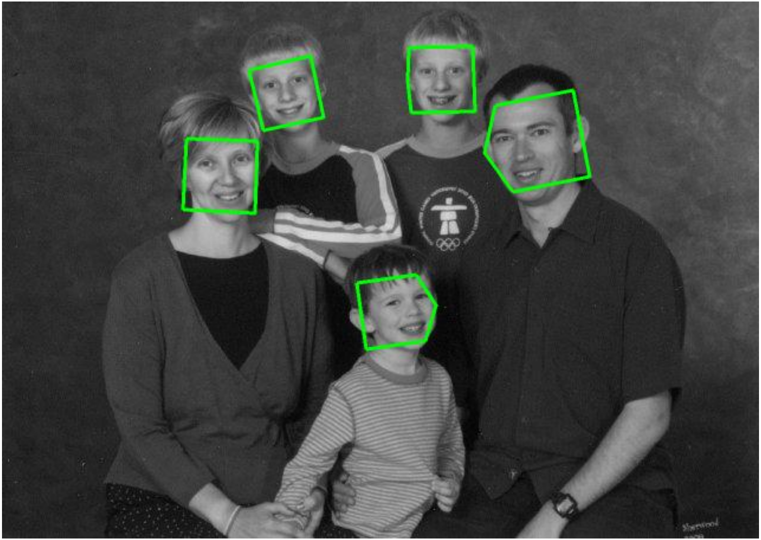


true class = 5



Examples: classification

- face detection and recognition

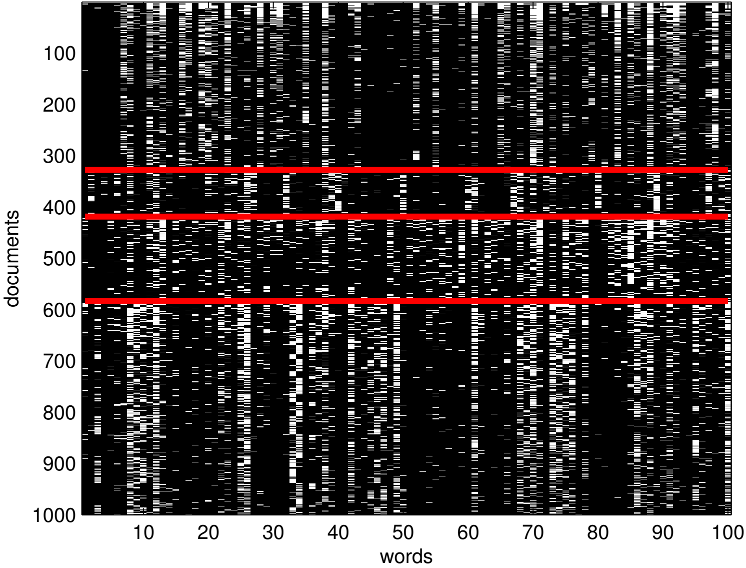


Examples: classification

- object detection

Examples: classification

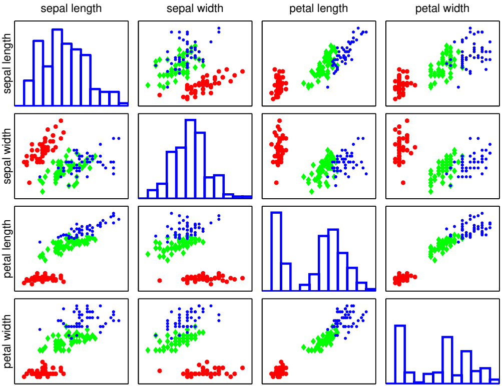
- document classification (e.g. email spam filtering)



Examples: classification

- Iris flower data set from Ronald Fischer (1936)





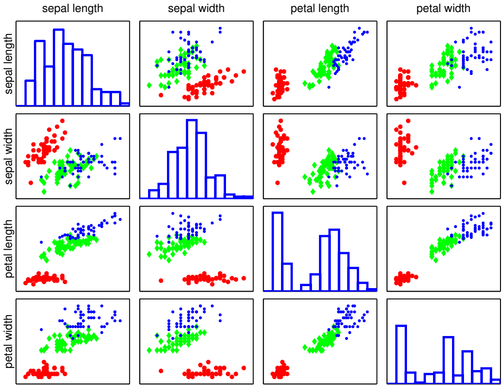
Examples: regression

- predict prices at the stock market given the past
- predict 3d location of a robot arm given sensor readings
- predict temperature at any location around the world given discrete set of sensor stations

Examples: clustering

- Iris flower data set from Ronald Fischer (1936), forget the labels



Examples: dimensionality reduction (PCA)

- Eigenfaces (Olivetti face database)