Universität Stuttgart
KI – Institute for Artificial Intelligence

Analytic Computing

Machine Learning
1 Introduction

Prof. Dr. Steffen Staab

Nadeen Fatallah
Daniel Frank
Akram Sadat Hosseini
Jiaxin Pan
Osama Mohamed
Arvindh Arunbabu
Tim Schneider
Yi Wang

https://www.ki.uni-stuttgart.de/

Today‘s objectives (Monday, April 7, 2025)

Completing this slide deck you should

- Know about basic organizational requirements
- Know basic terminology
  - Machine learning, data mining, knowledge discovery, supervised ML, unsupervised ML, regression, classification, reinforcement learning, clustering, dimensionality reduction

Foundations of Machine Learning

vs (deprecated) “Machine Learning”

- Foundations of Machine Learning is a Bachelor’s course and a Bachelor’s course only.
- It is suggested for attendance after “Foundations of Artificial Intelligence”
- It requires programming and mathematical expertise
  - Python
  - Calculus, Linear algebra, Probability theory
- Its exam is not yet scheduled
  - it will happen in August (likely) or September and again in March 26
  - **alternative exam dates will not be offered**

What if

I am a master student who attend “**Machine Learning**” in 2024 (or earlier)

- If you have earned admittance to exam in 2024 or before
  - Register for the exam in August/September
  - Students who attended the master course in 2024 or before and earned admittance to exam will receive a separate exam – in parallel to the exam for the bachelor students, but longer (120 minutes)
  - Correct registration is key!

What if

I am not a student of computer science (software engineering, data science, media informatics, etc.)

But: an engineer, architect, AISA student, … .

- **You are welcome**!
- But the expectation is that you have the knowledge that we expect from our computer science programs
- There are two exercise groups for engineers with special hints for self-learning

How to successfully pass the course?

- Attend lectures
  - Lectures will be recorded
- Submit excercises in groups of three
  - Pen-and-paper
  - Programming
  - Check out: the details of submitting and handing in excercises
- Acquire admission to exam by meeting criteria specified for the excercises (next)
- Pass written exam
- **Watch out for lecture material and exercise material on Ilias**

Monday, April 28: Recorded video lecture

Exercises

Schedule of Exercises

- For this week (10.04 and 11.04) introduction,

How we handle 3 Thursday public holidays

- May 1

Weekly assignments

- 11 Assignments in total
- Submission in groups of three via ILIAS
  - Submission groups do not have to be in the same exercise group
  - Use the ILIAS forum to find submission group members
  - Exam admission granted on individual basis (see next slide)
- Assignments published: Tuesdays at 12:00 (noon)
- Submissions due: Mondays at 12:00 (noon)
- First assignment announced:
  - Tuesday, April 15 (likely earlier – we will let you know)
- First assignment due: Monday, April 21

How to get the exam admission? (1)

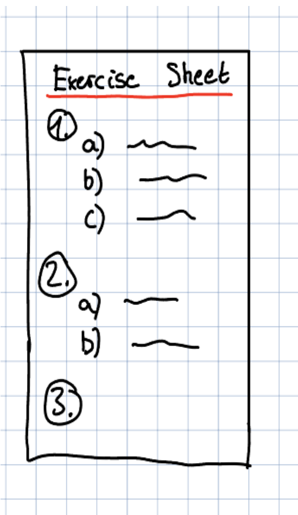
- Attend exercise sessions
- Presenters are randomly selected from audience ***who voted***
- You vote for tasks that you can present in the exercise group
- You vote in an ILIAS poll
  - Start poll: publication of assignment sheet
  - End poll: Thursdays 7:00 before exercise session
- **Exam admission requirement >= 80% voted tasks**

How to get the exam admission? (2)

1.

Students submit in groups of 3

To ILIAS on submission deadline

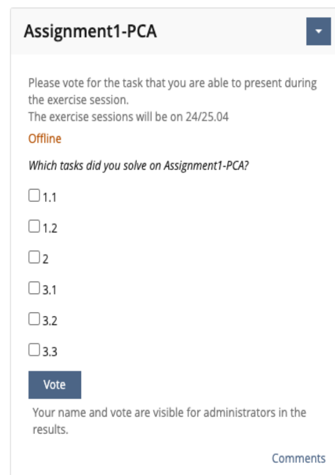


2.

Before the exercise session

Students vote for tasks

„I am able and willing to present the solution to this task“ in ILIAS poll


Assignment1-PCA

Please vote for the task that you are able to present during the exercise session.
The exercise sessions will be on 24/25.04

Offline

Which tasks did you solve on Assignment1-PCA?

- [ ] 1.1
- [ ] 1.2
- [ ] 2
- [ ] 3.1
- [ ] 3.2
- [ ] 3.3

Vote

Your name and vote are visible for administrators in the results.

Comments

3.

Random selection

Select a presenter from the voting sheet

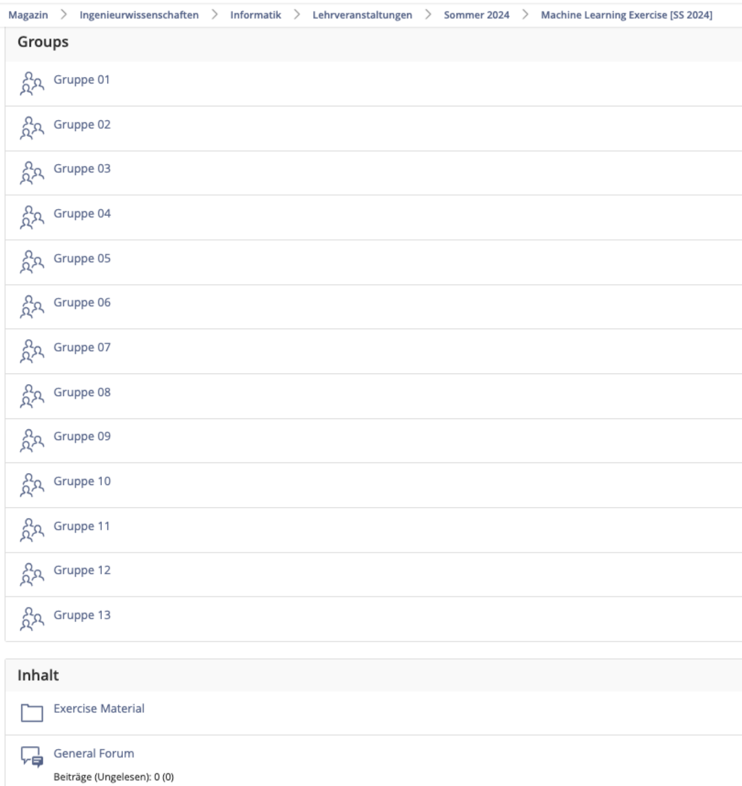
Anna

We check for: $V \subseteq S$

Thursday holidays

- Three weeks with public holidays on Thursday
  - 01.05 1.Mai
  - 29.05 Christi Himmelfahrt
  - 19.06 Fronleichnam
- No exercise sessions during these weeks
  - Neither on Thursdays nor on Fridays of these three weeks
  - We upload a recording to ILIAS
- Same voting system, voted tasks will be checked

Communication: Ilias Forums



Emails will be ignored unless the question requires knowledge about private data, such as grades – there are just **too many emails** already now

Back to the Lecture

Lecture slides are not a script

You must take notes

I record the lecture, however going back to lecture recordings is too time consuming in general

TAKE NOTES

What do we want to achieve? (1)

That you know the basics of machine learning

- core assumptions
- core methods
- how these are interwoven
  - it is a fabric, not a tree!

That you can master the theory (mathematical formulas!) underlying machine learning such that you can read and understand current machine learning papers

What do we want to achieve? (2)

That you have gained – a little – experience in implementing machine learning solutions

What do we want to achieve? (3)

From a student who was successful applying to an internationally competitive PhD program:

[Your lectures] gave me the knowledge and skills I needed to pass all technical interviews during the application process.

What is not the case

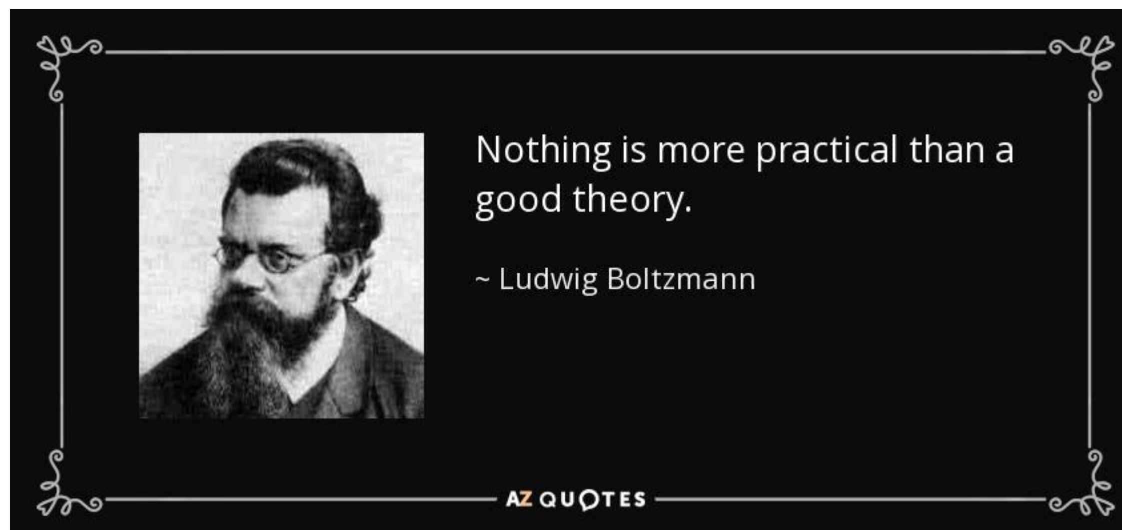
Lecture and exercise material are not completely aligned with the exam.

Why?

- We need an exam,
  but what matters more is what you learn
- Not everything you should learn
  can be reasonably tested in an exam

Lecture focus: formalization of machine learning
Exercise focus: practicing theory and engineering

- Why:
  1. Good theory as a basis for proper engineering
  2. Allows you to understand research papers
     - long formulas have repetitive structures that you need to learn to read!



Literature (cf. Ilias)

Most Recommended Book:

- Probabilistic Machine Learning, Book series by Kevin Murphy (free download!)

Books:

- An Introduction to Statistical Learning with Applications in R. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani.
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman.
- Deep Learning. Ian Goodfellow, Yoshua Bengio, Aaron Courville.
- Neural Networks and Deep Learning. Michael Nielsen.
- Data Mining. Charu C. Aggarwal.
- M. Deisenroth, A. A. Faisal, C. Soon Ong. Mathematics for Machine Learning. Cambridge University Press, 2020.
- Stanley Chan. Introduction to Probability for Data Science, Michigan Publishing 2021

Further resources

Web:

- A visual introduction to machine learning.
- KD Nuggets.
- KDD Video Lectures.
- Mathematics for Machine Learning. MIT Open Courseware

Conferences/Proceedings:

- NeurIPS - Neural Information Processing.
- Proceedings of Machine Learning (includes ICML).
- Int. Conference on Learning Representations
- ACM SigKDD Knowledge Discovery and Data Mining. Int. Conference.

On with it

What is Machine Learning?

- **Machine Learning** is like someone learning from past experiences to predict future outcomes.

[ChatGPT]

- **A computer program is said to learn**
  - from ***experience E***
  - with respect to some class of ***tasks T***,
  - and ***performance measure P***,

  if its performance at tasks in T, as measured by P, improves with experience E.

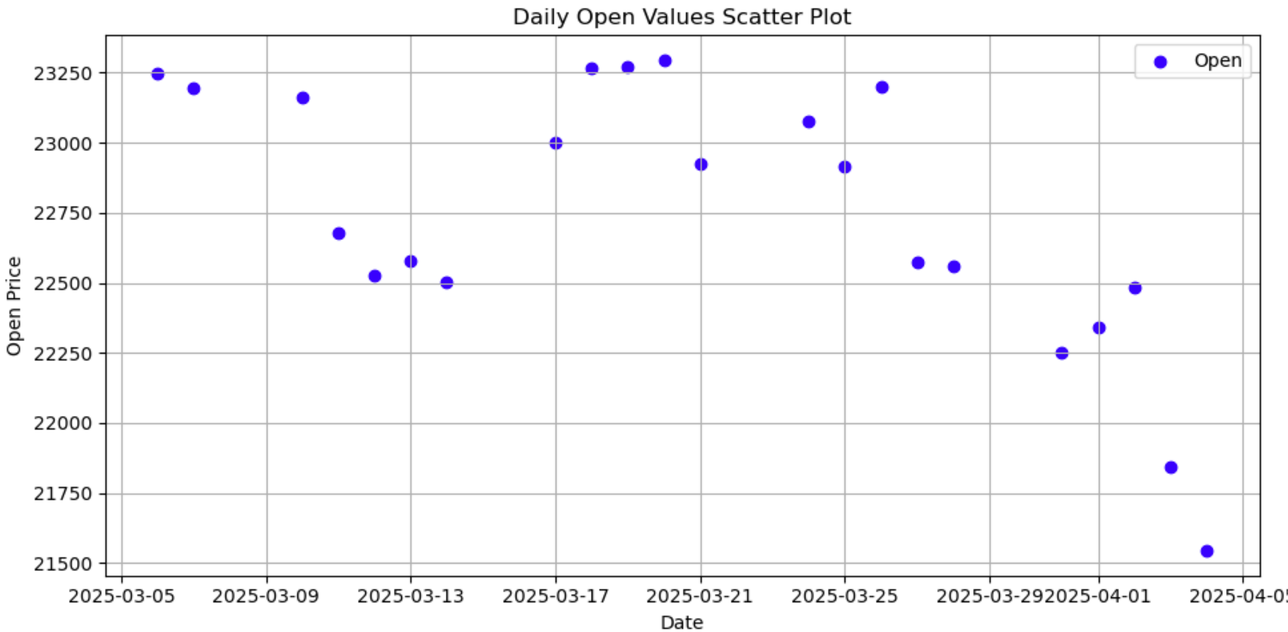
[Tom Mitchell]

What does Machine Learning learn?

- Regression

Regression as task *T*: Predict a value (infer)

Example: Predict DAX index over next days



What next?

- $\hat{f}(2025{-}04{-}05)=?$
- $\hat{f}(2025{-}04{-}06)=?$
- $\hat{f}(2025{-}04{-}07)=?$

Training Data (the experience ***E***):

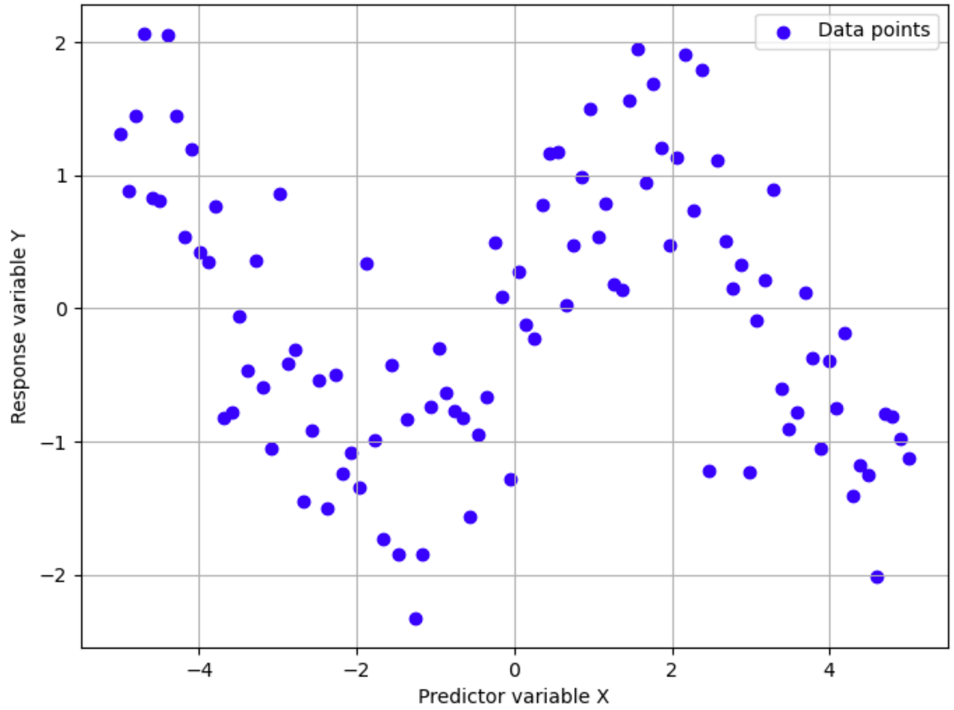
$$\mathcal{D} = \{(2025{-}04{-}04; 21{,}543.47), (2025{-}04{-}03; 21{,}842.08), \dots\}$$

Regression (task *T*)

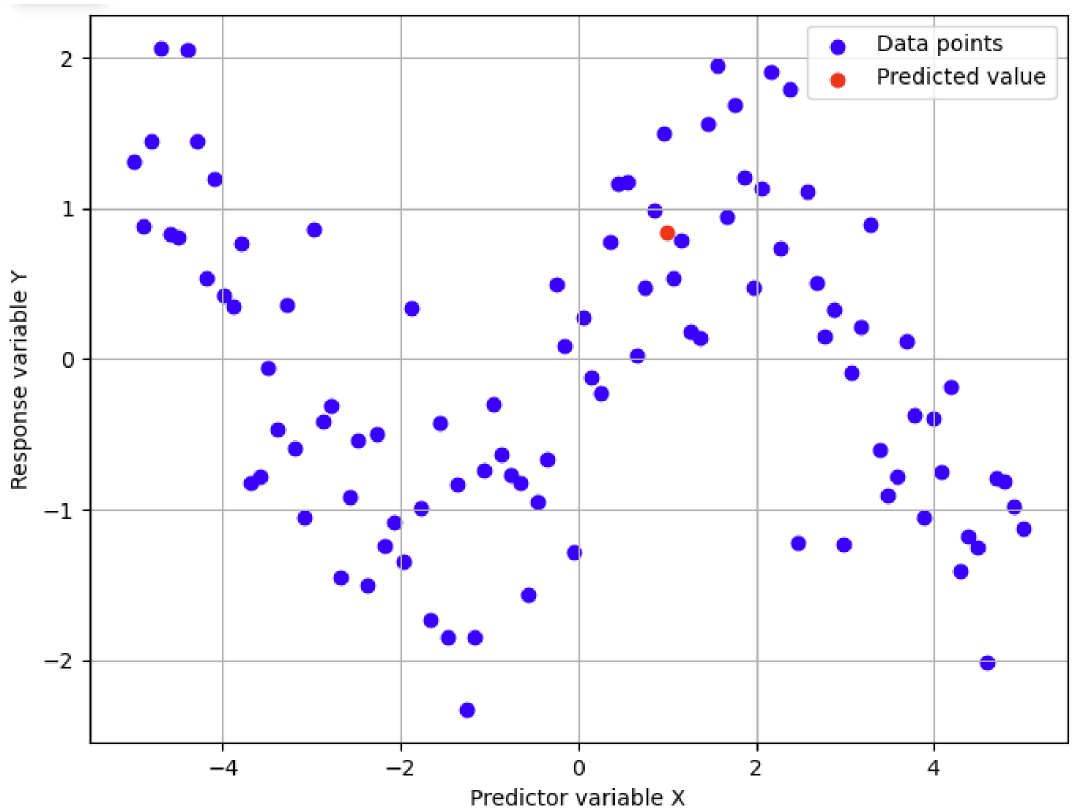
$$\widehat{f_\theta}(x) = y,$$
$$x \in \mathbb{R}^n, y \in \mathbb{R}^k$$

Training Data (experience *E*)

$$\mathcal{D} = \{(x_1, y_1), \dots, (x_n, y_n)\} \in \mathbb{R} \times \mathbb{R}$$



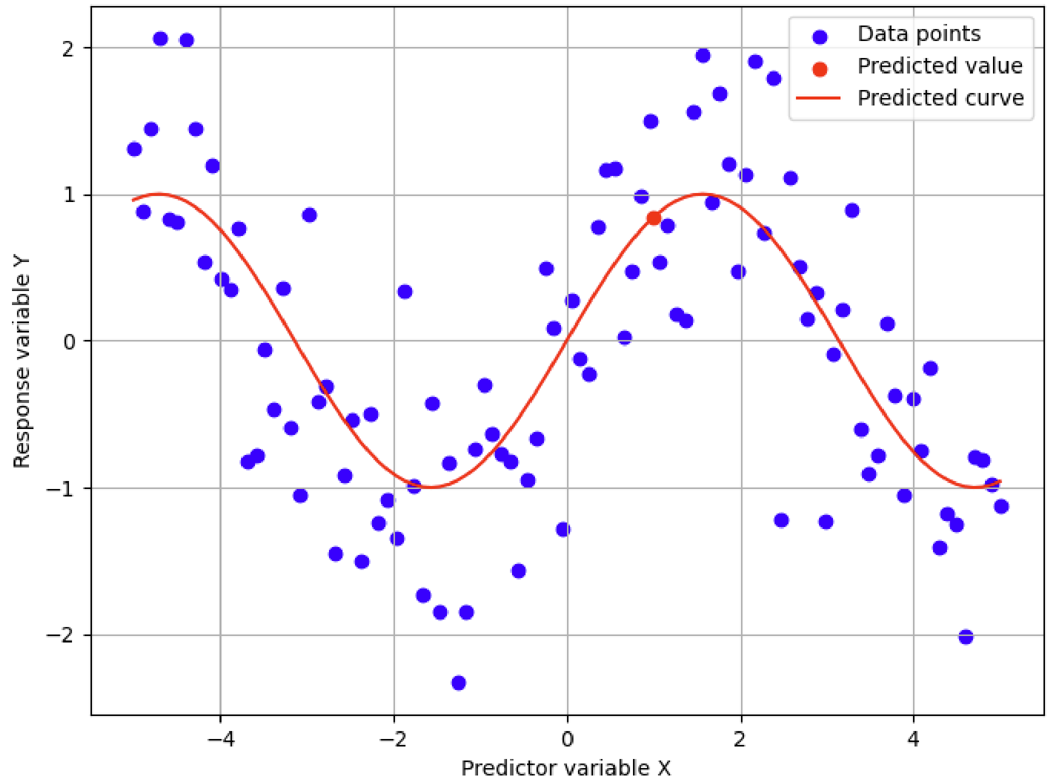
Inference: Predicting a value: $\hat{f}(1) = 0.841$



Predicting the function $\hat{f}(x)$

Values for in-distribution: $\hat{f}(x)$, for all $x \in [-5,5]$

Values for out-of-distribution: $\hat{f}(x)$, for all $x \notin [-5,5]$



What does Machine Learning learn?

- Regression
- Classification

Classification as task *T*

Example: which animals are depicted?

$$\hat{f}\left(\vphantom{\int}\right.$$  $$\left.\vphantom{\int}\right) = ?$$

Classification as task *T*

Example: Dangerous or not?

$$\hat{f}\left(\vphantom{\int} \right.$$  $$\left. \vphantom{\int} \right) = ?$$

Learning a Classifier from Experience *E*

Training data $\mathcal{D} =$

{( , 0), ( , 0), ( , 0),

( , 1), ( , 1), ( , 0), ... }

0 harmless

1 dangerous

Inference: Classification

$\hat{f}\Bigg($  $\Bigg) = 0$

$x \in \mathbb{R}^{200\times350}, y \in \{0,1\},$

0 for harmless

Learning a Classifier from Experience E

Training data $\mathcal{D} =$

$\Bigg\{\Bigg($  $, \begin{pmatrix}1\\0\\0\\0\\\ldots\end{pmatrix}\Bigg), \Bigg($  $, \begin{pmatrix}0\\1\\0\\0\\\ldots\end{pmatrix}\Bigg), \Bigg($  $, \begin{pmatrix}0\\0\\1\\0\\\ldots\end{pmatrix}\Bigg),$

$\Bigg($  $, \begin{pmatrix}0\\0\\0\\1\\\ldots\end{pmatrix}\Bigg), \Bigg($  $, \begin{pmatrix}0\\0\\0\\0\\\ldots\end{pmatrix}\Bigg), \Bigg($  $, \begin{pmatrix}0\\0\\1\\0\\\ldots\end{pmatrix}\Bigg), \ldots \Bigg\}$

with y_1 one−hot encoding of cat,
with y_2 one−hot encoding of dog,
with y_3 one−hot encoding of fox,
with y_4 …

Inference: Classification

$\hat{f}\Bigg($  $\Bigg) = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ \dots \end{pmatrix}$

$$x \in \mathbb{R}^{200\times 350}, y \in \{0,1\}^{10{,}000},$$

with y_1 one−hot encoding of cat,

with y_2 one−hot encoding of dog,

with y_3 one−hot encoding of fox,

with y_4 ...

Supervised Machine Learning: Common paradigm of regression and classification

In supervised machine learning, **we assume** that there is a ground truth, a function $f\colon \mathcal{X} \longrightarrow \mathcal{Y}$
that maps values from the input space $\mathcal{X}$ to values from an output space/target space $\mathcal{Y}$.

We are given training data $\mathcal{D} = \{(x_1, y_1), \ldots, (x_n, y_n)\}$
that **we assume** represents examples of the mapping f.

We aim to generate a function $\hat{f}\colon \mathcal{X} \longrightarrow \mathcal{Y}$ that approximated **as closely as possible (P)** the correct mapping f.

What else can be predicted?
Classification vs regression

Discrete classification

- When to send car to inspection: Now / soon / later
- What disease is it: Viral infection / bacterial infection
- Is it malware: Virus 1 / virus 2 / virus 3 /... / goodware
- Stock at stock market: Buy / hold / sell
- Word form in a text: Proper noun / noun / adjective...
- How good is a move in chess: Win / hold / loss

Continuous prediction (regression)

- How to steer the car: [-30,+30]
- When to buy a stock: [0,10000]

Endless list

What is Machine Learning?

- **Classification**
- **Regression**
- Reinforcement Learning

  [See course by Prof. Niepert in winter term]

Reinforcement Learning: Exploration and Exploitation

Start **without experience**, have some random ***policy*** (= strategy what to do)

1 ***Explore*** (pick random action) or ***exploit policy*** (pick action): which animal to stroke

2 Determine ***reward*** (**collected experience**): alive and happy vs bitten or dead

3 Adapt policy

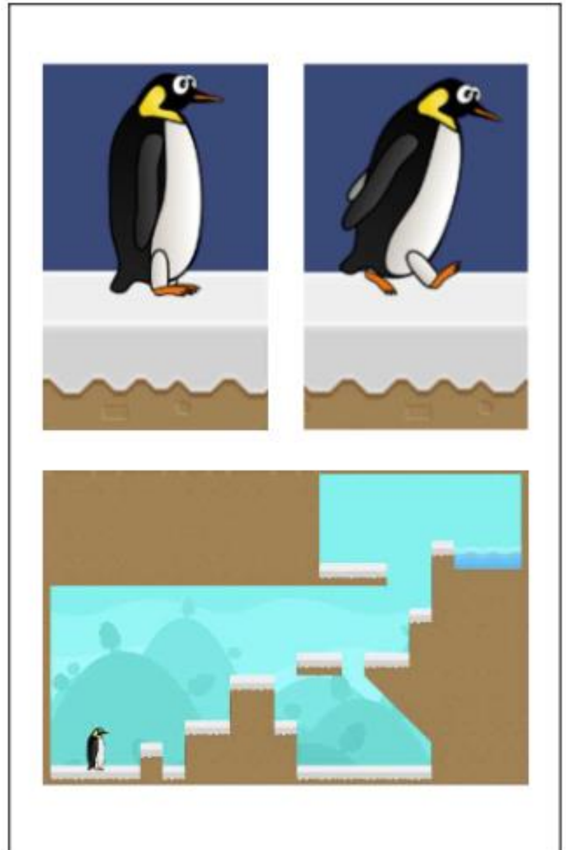
4 Goto 1

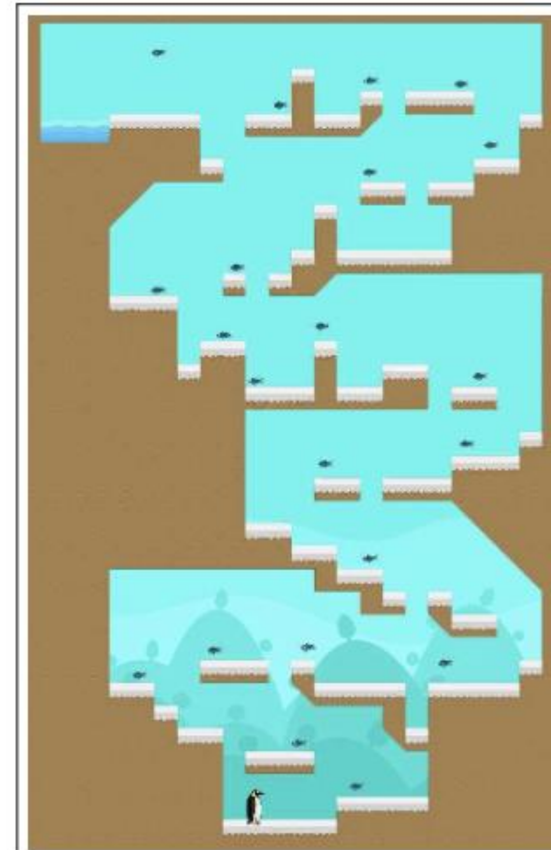
Reinforcement Learning
Another example

Learn policy how to solve a hit and run game.

Learning happens by **experiencing rewards** – over more than 1 step.

No example of how to solve the game.





R. Hedeshy et al. All Birds Must Fly: The Experience of Multimodal Hands-free Gaming with Gaze and Nonverbal Voice Synchronization. In: Proc. ICMI '22, November 7–11, 2022, Bengaluru, India

Unsupervised Machine Learning: Another paradigm

Unsupervised machine learning refers to machine learning **without known target values** and **without rewards** from the environment.

"Here is the raw data. Learn!"

Unsupervised machine learning
is only based on the data attributes itself

Coverage: Fur, skin, scales, feathers

Visible ears: yes, no

Larger than man: yes, no

Dominating color: white, grey, brown, golden/yellow, red

Unsupervised machine learning

Coverage: Fur, skin, scales, feathers

Visible ears: yes, no

Larger than man: yes, no

Dominating color: white, grey, brown, golden/yellow, red

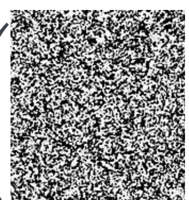
Dimensionality reduction according to fur – not fur:

Learning data distributions

$f\Big($  $\Big) = 1$

$f\Big($  $\Big) = 1$

observed

most likely $f\Big($  $\Big) = 0$

most likely $f\Big($  $\Big) = 0$

unobserved

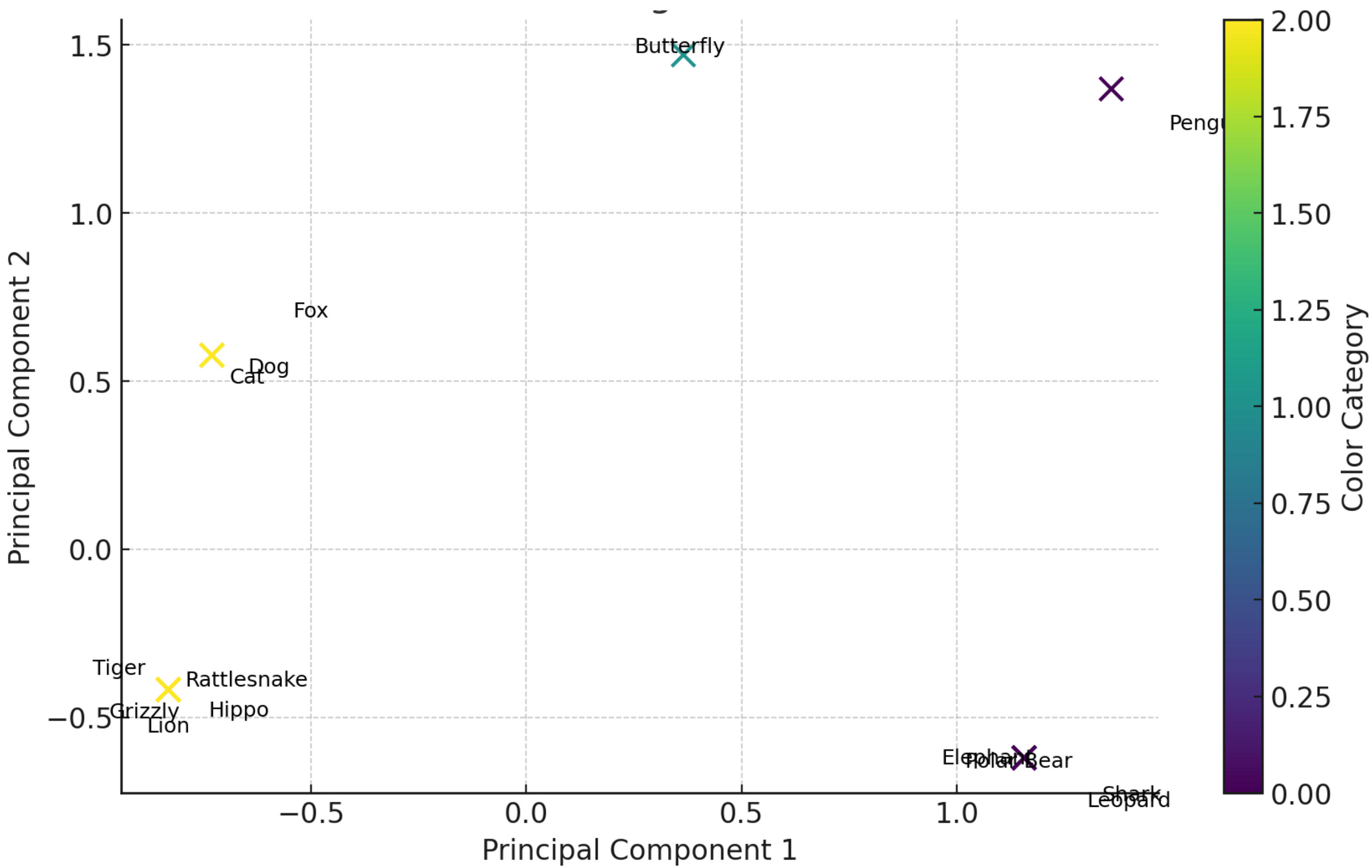
Learn the distribution

learn about the data generating process that occurrence of black and white color is as it is in a penguin – not as it is in white noise

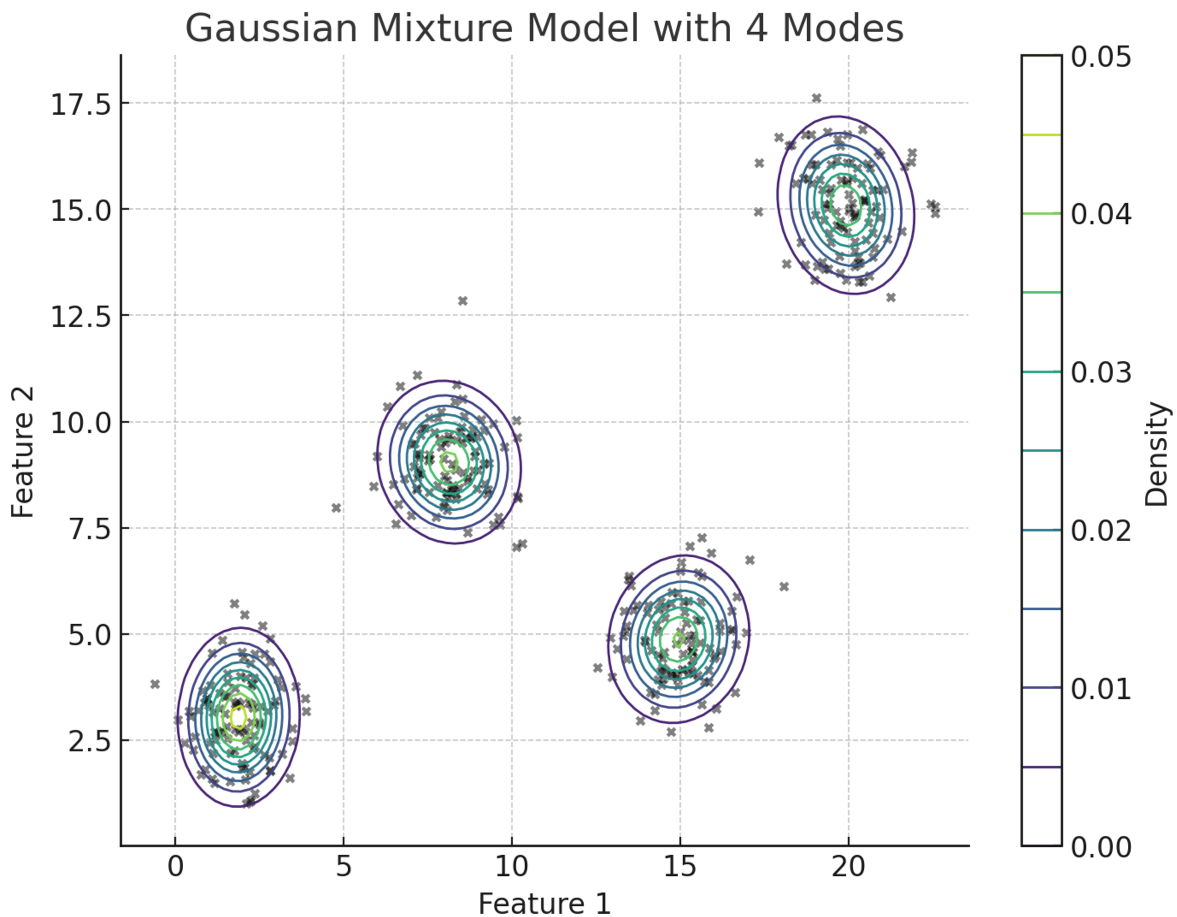
Example: Applying principal component analysis

```
# List of animals and attributes for clustering
animals = ['Lion', 'Shark', 'Hippo', 'Rattlesnake', 'Polar Bear', 'Grizzly',
'Tiger', 'Leopard', 'Cat', 'Dog', 'Elephant', 'Fox', 'Penguin', 'Butterfly']

# Color and size attributes for clustering
# Color categories: warm, cool, neutral/multicolored
colors = ['Warm', 'Cool', 'Warm', 'Warm', 'Cool', 'Warm', 'Warm', 'Cool',
'Warm', 'Warm', 'Cool', 'Warm', 'Cool', 'Neutral']

# Size categories: large, medium, small
sizes = ['Large', 'Large', 'Large', 'Large', 'Large', 'Large', 'Large', 'Large',
'Medium', 'Medium', 'Large', 'Medium', 'Small', 'Small']
```

PCA for dimensionality reduction/clustering



A more sophisticated ideal



Data Clustering

Data Clustering is a type of **unsupervised learning** technique in which data points are grouped into clusters (or groups) based on their similarities. The goal of clustering is to partition a dataset into distinct groups such that the data points within each group (cluster) are more similar to each other than to those in other groups.

Clustering is used to uncover patterns, structures, or relationships within data that are not explicitly labeled.

What is (related to) Machine Learning?

- **Classification**
- **Regression**
- Reinforcement mearning

  [See course by Prof. Niepert in winter term]
- …

- **Dimensionality reduction**
- **Clustering**
- Subgroup discovery
- Rule mining

- Self-supervised learning
- Semi-supervised Learning
- Anomaly detection
- Generative AI
-

Terminological chaos

Good amount of chaos in the use of presented terminology

- ”pattern recognition” is also used for supervised machine learning
- Gesellschaft für Klassifikation uses the term “classification” to refer to “clustering”
- “anomaly detection” exists in purely supervised or purely unsupervised versions
- “data mining” is sometimes a synonym for ”knowledge discovery”
- “generative AI” / ”self-supervised learning” is considered machine learning, but not data mining – but shares perspectives with data mining
- …

Is there a system in unsupervised machine learning?

Knowledge Discovery

„**Knowledge Discovery** ... is

the non-trivial ***process*** of identifying
valid, **novel**, potentially **useful**, and
ultimately **understandable patterns** in data"

[Fayyad et al 1996]

Problem of knowledge discovery

Input: Given **data** $\mathcal{D}$, such as

- set of images
- set of tuples from a relational database
- a set of edges forming a graph
- a set of text documents

Output: Find patterns, where a **pattern** is

- Expression S in a language L that describes correlations in $\boldsymbol{\mathcal{D}}$
  - Constraints on relational attributes
  - Correlations between attributes
  - Rules between values or words
- and S is simpler than listing data $\mathcal{D}$

Understandability: humans must understand the expression S

Validity: pattern described by S must apply to $\mathcal{D}$ and must likely apply to new data $\mathcal{D}^{new}$

Process: comprises multiple stages

Non-trivial: rules out simple operations like averaging

Stages of knowledge discovery from data

1. **Data Selection**: Choosing relevant data for analysis.
2. **Data Cleaning**: Handling missing values, noise, and outliers.
3. **Data Transformation**: Converting the data into a suitable format for mining.
4. **Data Mining**: Applying algorithms to discover patterns or models.
5. **Evaluation**: Assessing the quality and validity of the discovered knowledge.
6. **Knowledge Representation**: Presenting the results in a way that is understandable and useful to the user.

Universität Stuttgart
KI

Thank you!

Steffen Staab

E-Mail Steffen.staab@ki.uni-stuttgart.de
Telefon +49 (0) 711 685-88100
www.ki.uni-stuttgart.de/

Universität Stuttgart
Analytic Computing, Institut für Künstliche Intelligenz
Universitätsstraße 32, 50569 Stuttgart