VL Deep Learning for Natural Language Processing

1. Organization & Introduction

Prof. Dr. Ralf Krestel

AG Information Profiling and Retrieval

Lerning Goals for this Chapter

- Answering the question: am I right here?
- Explain and define deep learning
- Position deep learning with regards to machine learning
- Understand the historical development of the area of ML

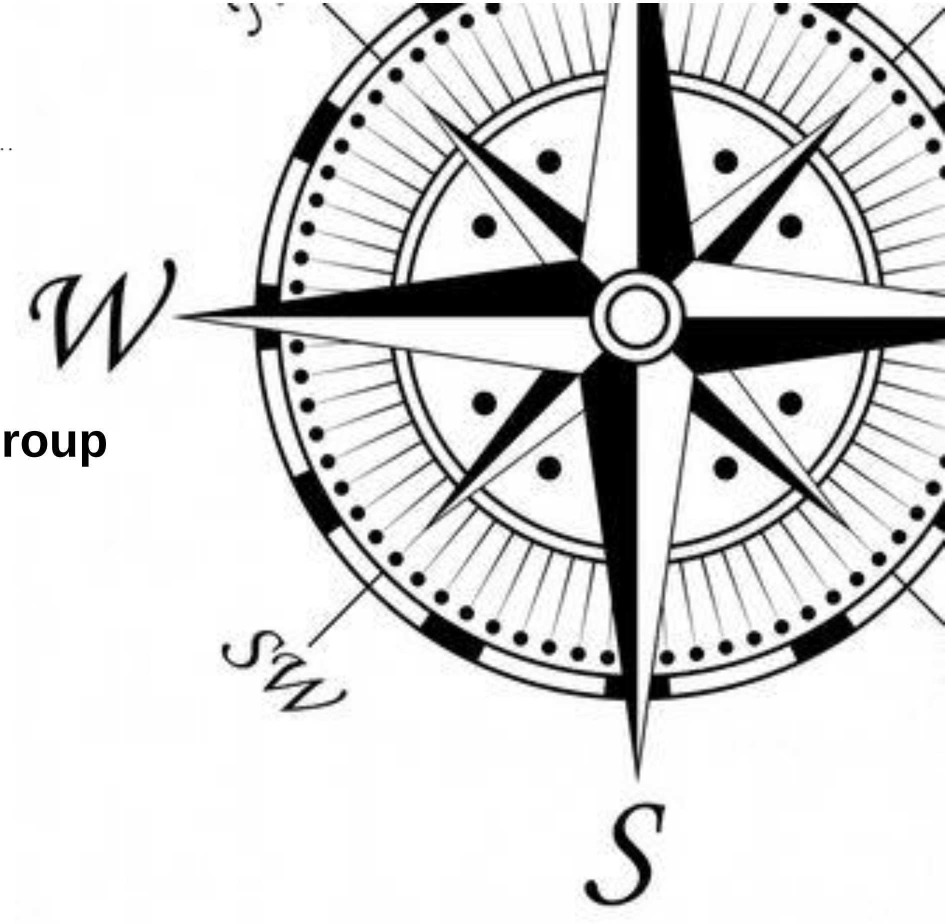
- Relevant chapters:
  - P2

15 Deep Learning Applications You Need to Know

- Fraud detection
- Customer relationship management systems
- Computer vision
- Vocal AI
- Natural language processing
- Data refining
- Autonomous vehicles
- Supercomputers
- Investment modeling
- E-commerce
- Emotional intelligence
- Entertainment
- Advertising
- Manufacturing
- Healthcare

https://builtin.com/artificial-intelligence/deep-learning-applications

Topics Today



1. **Organization**
   a. **Information Profiling & Retrieval Group**
   b. Course Organization
   c. Schedule for the Semester
2. Introduction to Deep Learning
3. Brief History of Machine Learning
4. Summary

Information Profiling & Retreival (IPR) Group

Who are we?

- Prof. Dr. Ralf Krestel

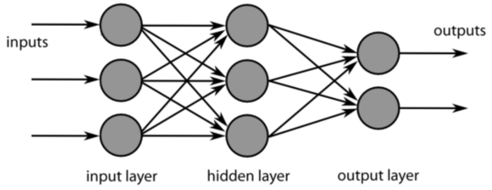
- In Kiel:
  - Anke Koslowski (Office Assistant)
  - Supryiol Mandal, PhD (Postdoc)
  - Aftab Anjum, M.Sc. (PhD student)

- In Potsdam:
  - Nitisha Jain, M.Sc. (PhD student)
  - Alejandro Sierra, M.Sc. (PhD student)

IPR – Research I

- Text Mining
- Information Retrieval
- Natural Language Processing
- Recommender Systems
- Knowledge Graphs

- Probabilistic Graphical Models
- Deep Learning



IPR – Research II

- Former Research Projects
  - CADL: Comment Analysis with Deep Learning
  - Mimir: Corpus Exploration and Knowledge Management

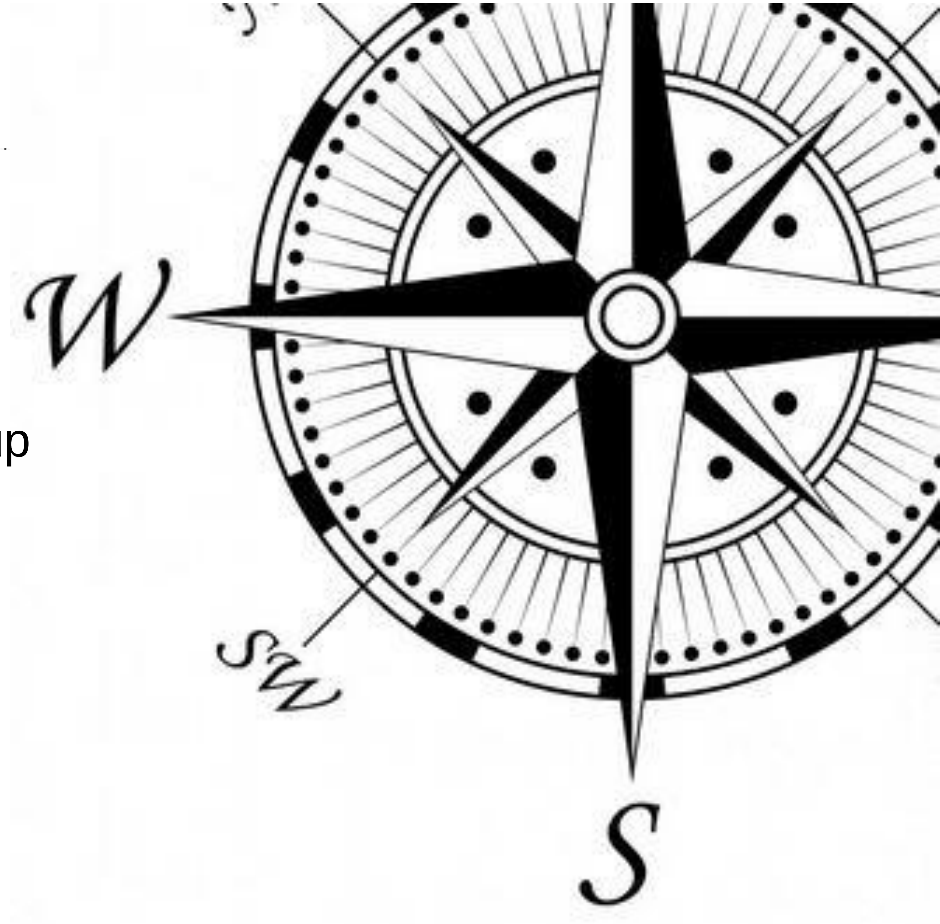
- Ongoing Research Projects
  - AI4art: Cognitive Analysis of art resources and texts
    - Named Entity Recognition
    - Knowledge Base Construction
  - CoCo: Connect & Collect – AI-powered Infrastructure for Labour Science
  - Patent Analysis with Deep Learning

IPR - Teaching

- IPR regularly offers the following courses (approximately every three semesters):
  - Text Mining (in German) BS W/INF V2 U2
  - Information Retrieval (in German) BS & MS W/INF V4 U1 PU1
  - Deep Learning for Natural Language Processing (in English) MS W/INF V2 U2
- In addition, seminars and projects are offered irregularly on the topics:
  - Knowledge Graphs
  - Recommender Systems
  - Topic Modeling

Topics Today

1. **Organization**
   a. Information Profiling & Retrieval Group
   b. **Course Organization**
   c. Schedule for the Semester
2. Introduction to Deep Learning
3. Brief History of Machine Learning
4. Summary



Communication & Resources

- We will use Moodle as our main platform:
  - **https://elearn.informatik.uni-kiel.de/course/view.php?id=51**
  - Enrolment key: Ernie&Bert
- Suggestions for improvment of
  - Use of slides/errors on slides
  - Assignments
- Questions any time!
  - During/After lecture
  - In person room CAP 901
    - Email first!
  - Email: **rkr@informatik.uni-kiel.de**

Audience

Who are you?

- Which semester?
- Which prior knowledge?
- Programming language skills?

Dates

- Lecture: Tuesday, 16:15 - 17:45
  - LMS8 - R.EG.017
- Exercises: Thursday 10:15 – 11:45
  - LMS8 - R.EG.017
- No lectures/exercises
  - 26.05.22 Holiday

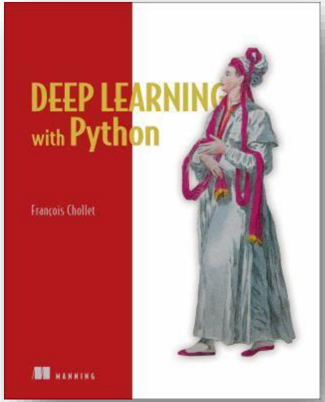
Grading

- Exam
  - Mostly „theoretical“ questions
  - Some „practical/applied“ questions
  - No programming questions

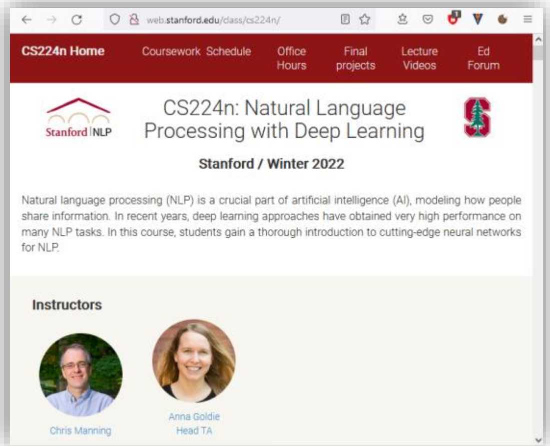
- Three homework assignments
  - **At least X points to be eligible for exam**
  - Hand-in through Moodle

Literature

- Deep Learning with **P**ython
  - Francois Chollet
  - Manning Publications Company

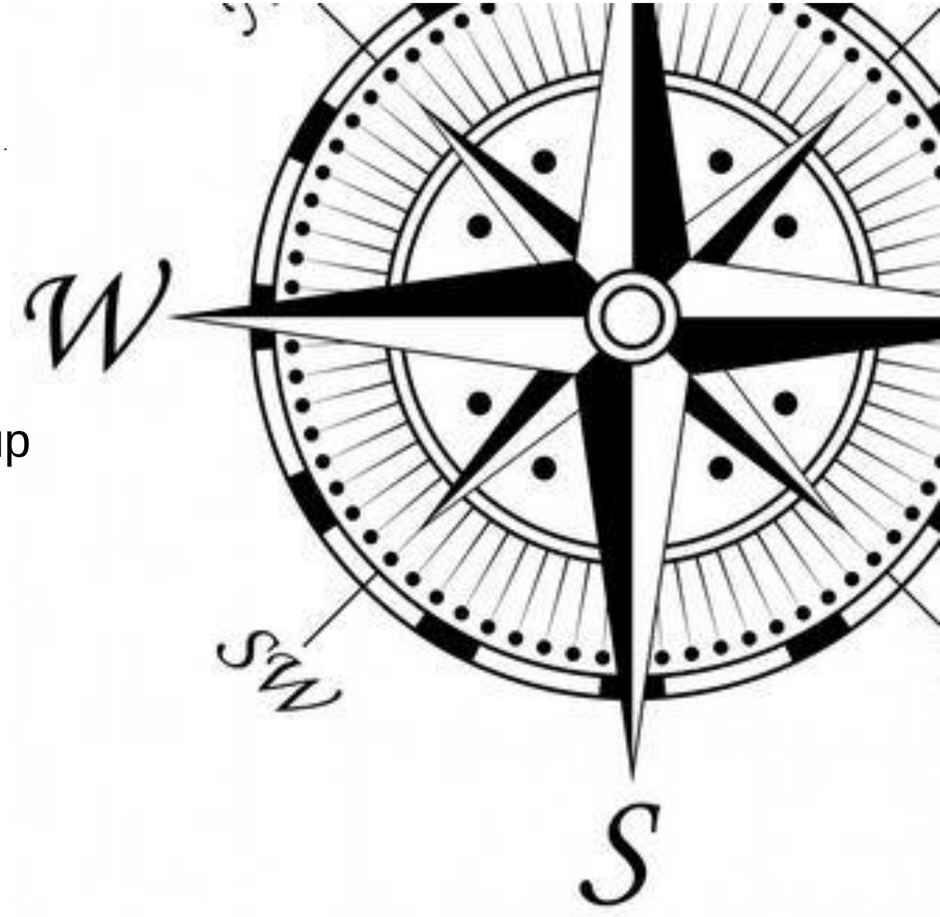


- **S**tanford Course
  - Chris Manning
  - http://web.stanford.edu/class/cs224n



Topics Today

1. **Organization**
   a. Information Profiling & Retrieval Group
   b. Course Organization
   c. **Schedule for the Semester**
2. Introduction to Deep Learning
3. Brief History of Machine Learning
4. Summary

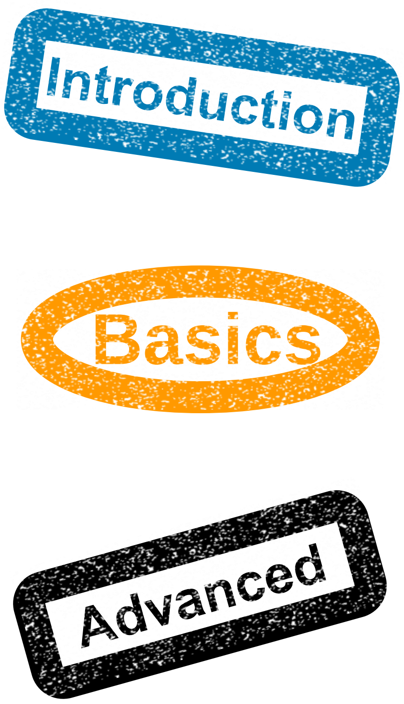


Overview of the Semester: Lecture

1. Organization & Introduction
2. NLP
3. Text Mining
4. Word Embeddings I
5. Word Embeddings II
6. Convolutional Models
7. Recurrent Models I
8. Recurrent Models II
9. Contextual Word Embeddings
10. Sequence-to-Sequence Models
11. Transformer Models
12. Neural Topic Models
13. Deep Generative Models

Overview of the Semester: Exercises

- Intro to Python/Keras/Colab
- Recap Neural Networks
- Praxis I
- Word2vec
- Text Classification
- Named Entity Recognition
- Praxis II
- More than Words
- Machine Translation
- Question Answering
- Recommender Systems
- Praxis III



Your Turn!

- After each topic there will be a small task which should be discussed with your neighbor for 5 minutes.
- You can use Moodle to further discuss these tasks, suggest solutions, comment, etc.

- What do you expect from this class?
  – Make a ranking of three learning goals.
  – Discuss with your neighbor.

Start 5 4 3 2 1 Ende

Learning Goals

- Students will be able to…
  - Explain different neural network architectures
  - Identify application areas and tasks for deep learning
  - Select suitable network architectures for a given task
  - Explain the functions of different components of NN
  - Apply DL in Python
  - Design and implement their own applications and evaluate their performance
  - Understand the theoretical background, in particular, they will be able to run the backpropagation algorithm manually
  - Realize the limits of deep learning and get an overview of current research in the field
  - Assess societal consequences of DL and discuss them

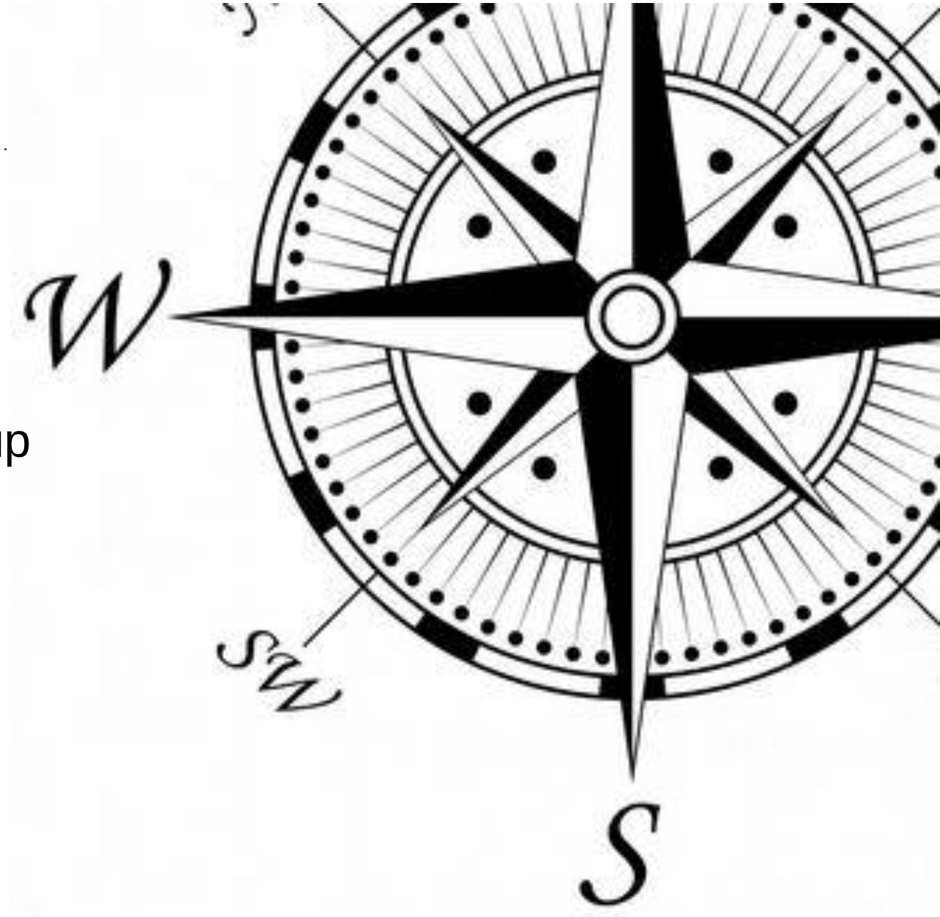
Your Learning Goals

- Which learning goals are most important for you?
  – Choose up to three!
  – Moodle

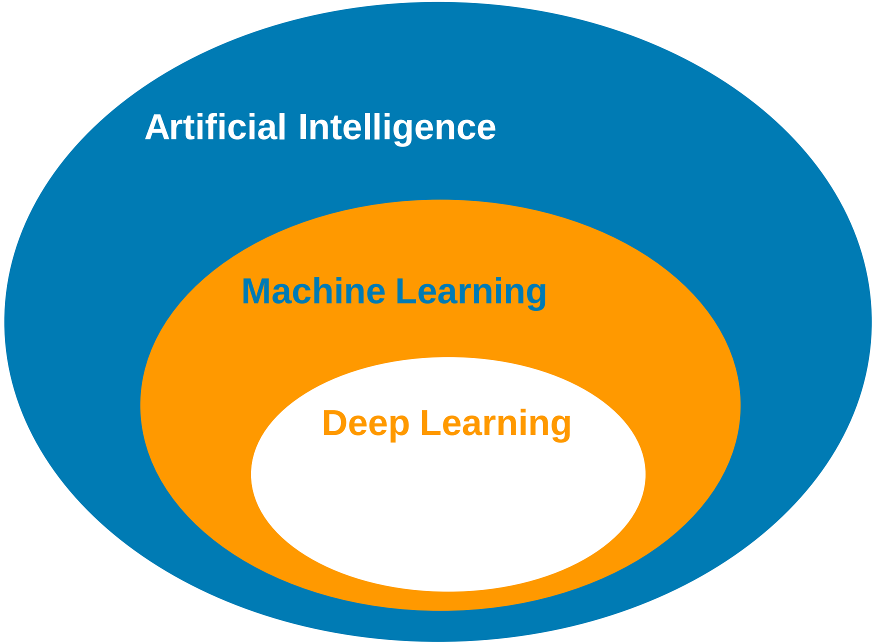
Start 3 2 1 Ende

Topics Today

1. Organization
   a. Information Profiling & Retrieval Group
   b. Course Organization
   c. Schedule for the Semester
2. **Introduction to Deep Learning**
3. Brief History of Machine Learning
4. Summary



What is Deep Learning



Artificial Intelligence

- AI: „The effort to automate intellectual tasks normally performed by humans“
  - Non only machine learning but also rule-based approaches
    - **Symbolic AI**
    - Expert systems in the 80ies
    - E.g. chess computer
  - Only works for certain problems
    - Clearly defined
    - Logical problems
  - Does not work for complex, diffuse problems
    - Image classification
    - Speech recognition
    - Translation

http://illumin.usc.edu/188/deep-blue-the-history-and-engineering-behind-computer-chess/

Machine Learning

- Instead of humans coming up with rules (classical programming),

Rules → **Traditional Programming** → Output
Data →

- Let computers **learn** the rules.

Output → **Machine Learning** → Rules
Data →

- A system based on machine learning (a **model**) is being trained, not programmed.
- ML is nowadays an important subfield of AI and
  - Closely connected to **statistics,**
  - Which usually deal with much **smaller** and **less complex** data.

Machine Learning as Transformation

- For machine learning you need
  - **input data** and the
  - corresponding, expected **output data** and
  - a **measure** to check how well the learning works.
- **Learning** is the (stepwise) adaption of the model to **increase** the chosen **measure**.
- What is learned: a **transformation** from the input to the output data.
- The challenge of ML and DL is to find a good transformation.
  - This is easier if the input (and output) data is represented in a meaningful way.
  - Dependend on the task:
    - E.g. image pixel as RGB values or hsv values

Example: automatic speech recognition

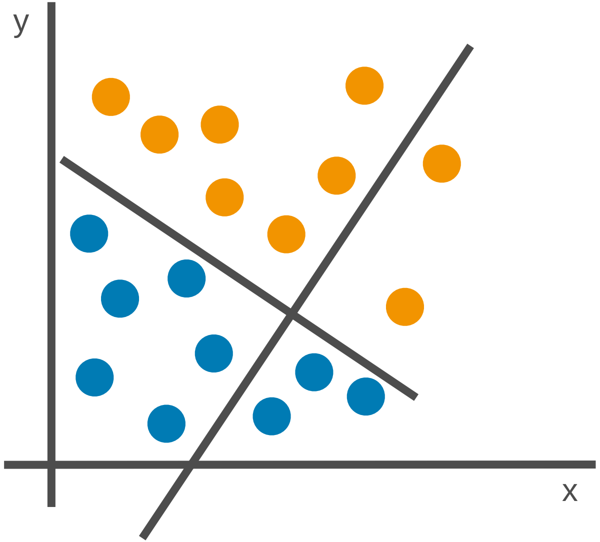
Input data: audiofiles

Output data: manually created transcripts

Measure: count of correctly recognized letters/words

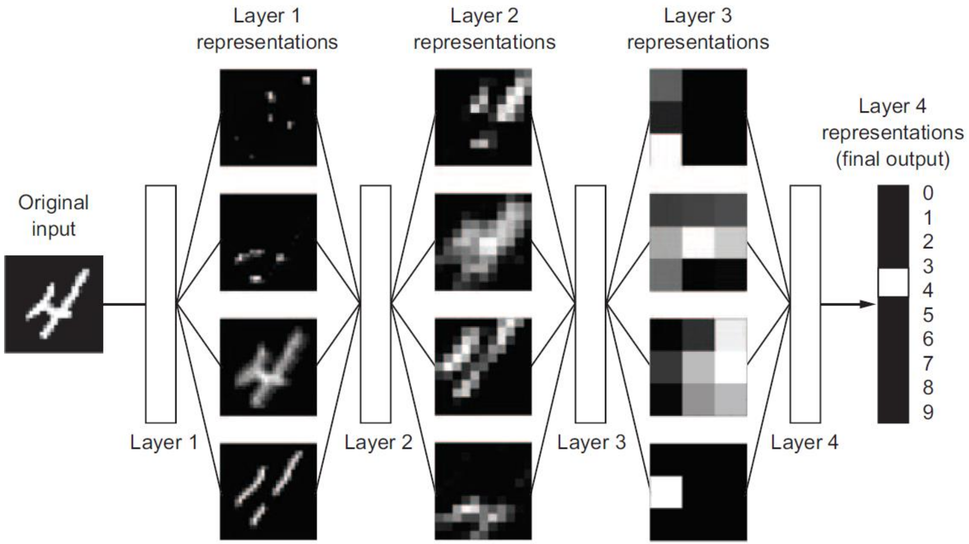
Representation Learning

- Example: classification of points into red and yellow
- Input: coordinates of the points
- Expected output: Color of points
- Measure: Proportion of correctly classified points
- New representation of the points
- Change of coordinates
- Now, classification very easy: x>0

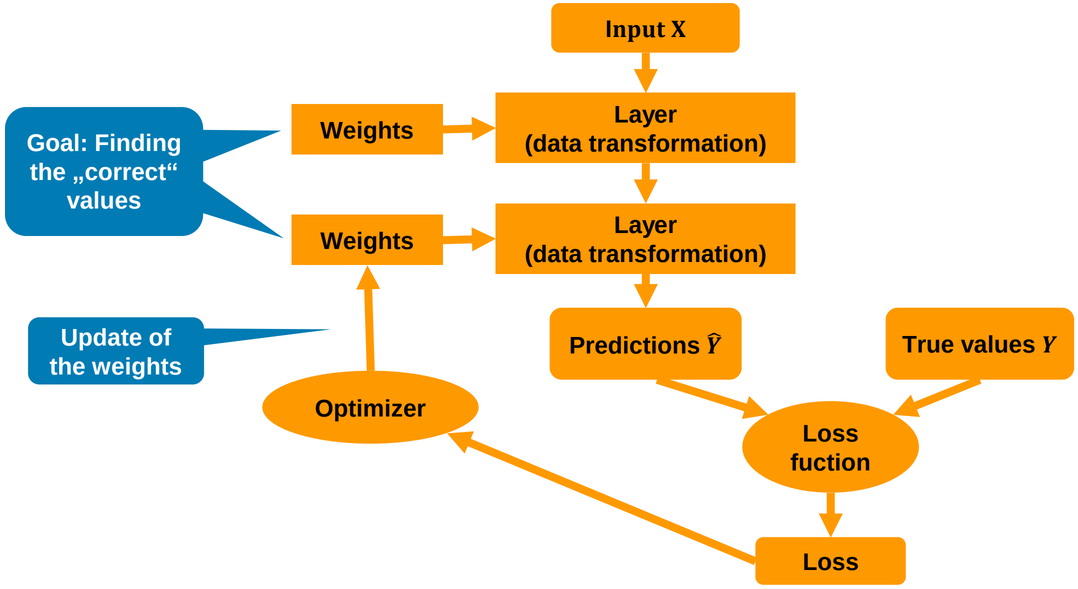


- ML: Systematic search for good representation
  - Learning=Search for **better representation**
    - Coordinate change, linear projection, non-linear operations, …
  - ML-Algorithms are not very creative
    - Simple search of the **hypothesis space**

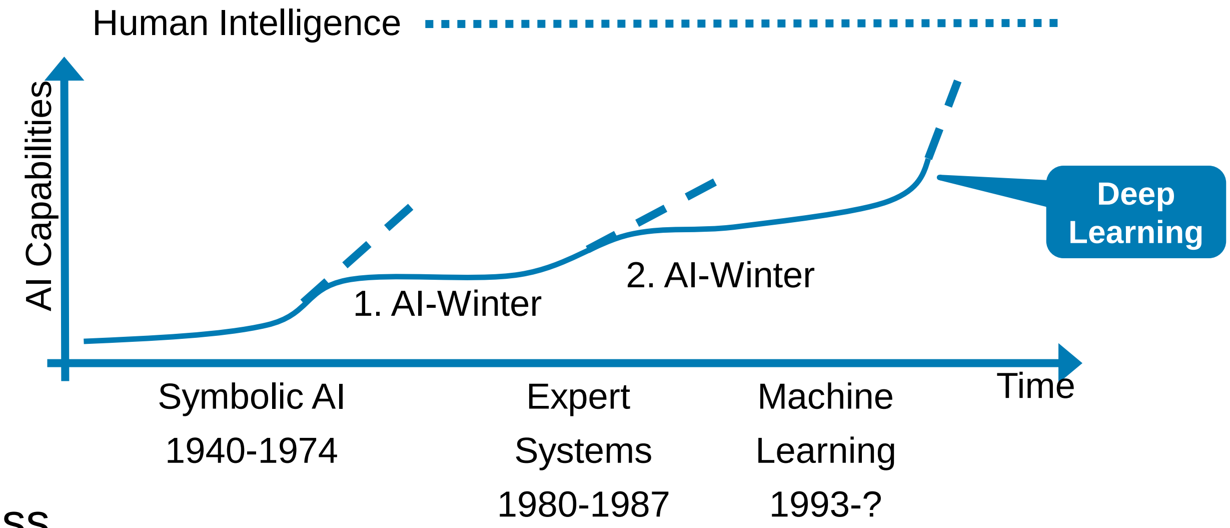
„Deep“ Learning



Deep Learning Architecture



Deep Learning Hype



1. First sucess
2. Hopes/Expectations (too high)
3. Disappointment
4. Less research funding

Deep Learning Tasks

- Which kind of problems can be solved with DL?
  - Name examples of task which can be solved with DL.

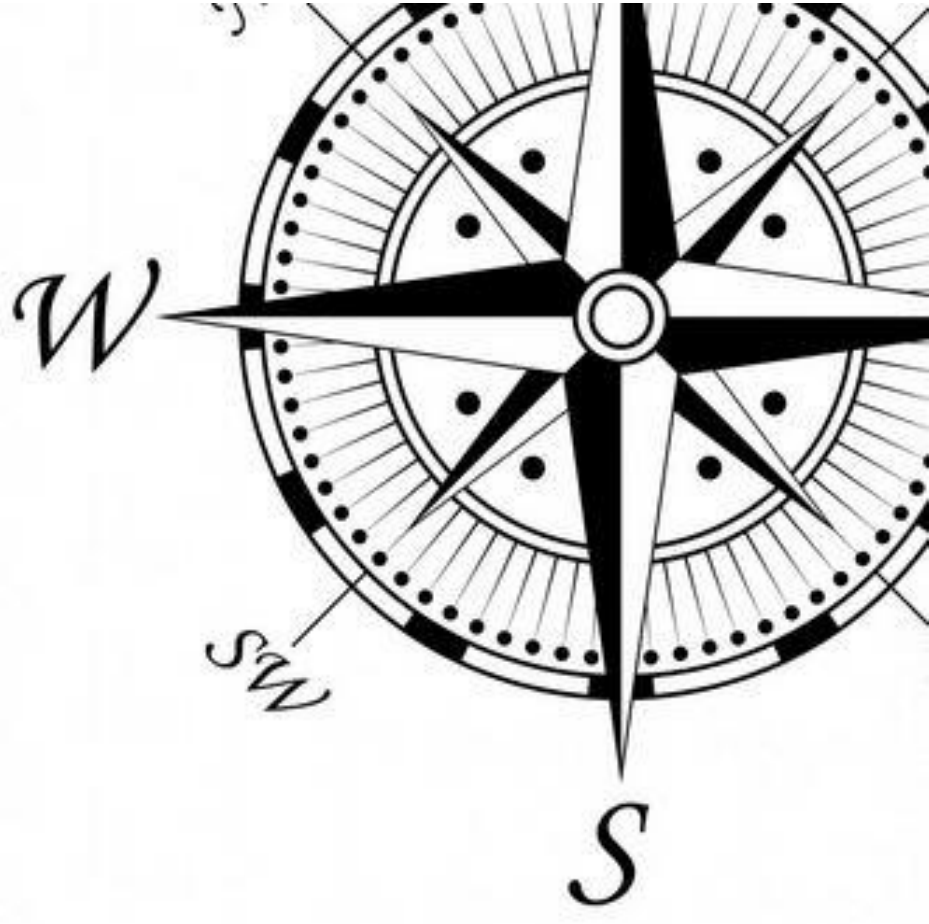
- Which type of problems DL cannot solve as well, if at all?
  - What are the reasons?

- How likely do you think is a third AI winter?
  - How to avoid a third AI winter?

Start 5 4 3 2 1 Ende

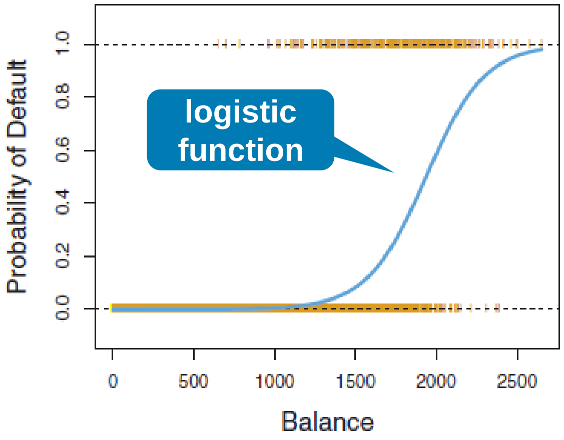
Topics Today

1. Organization
2. Introduction to Deep Learning
3. **Brief History of Machine Learning**
4. Summary



Probabilistic Modeling

- Originated from statistics
  – Older than computers



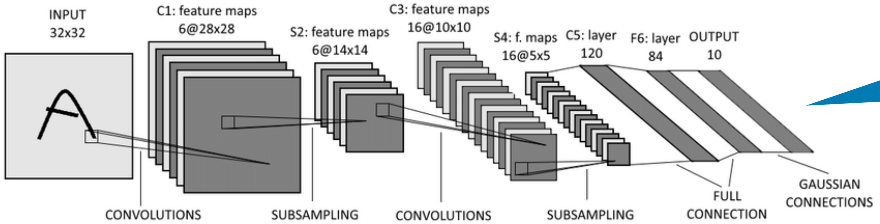
- Naive Bayes algorithm
  – Naive: input data independent
  – Bayes: $P(c|f) = \frac{P(f|c)\,P(c)}{P(f)}$

Bayes theorem

- Logistic regression
  – Classification, **not regression**!

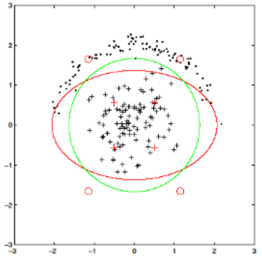
Early Neural Networks

- First idea in the 1950s
- Problem back then: Not possible to train large networks
  - Only toy examples, no useful applications
- Mid of the 1980s
  - Rediscovery of the backpropagation algorithm
    - Training of large nets using gradient descent
- In 1989, first practial application of neural Nets
  - LeCun (Bell Labs): LeNet, hand writing recognition
  - Convolutional neural network (CNN) & backpropagation



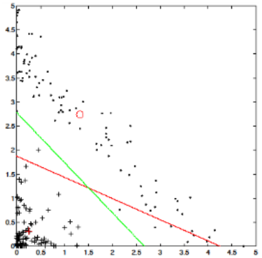
LeNet-5: In the 1990s used by USPS to recognize zipcodes automatically

Kernel Methods



- Group of classification algorithms
- Developed in the 1990s; still very popular
- Best-known algorithm: Support Vector Machines (SVM)[CV95]

1. Data is being projected into an high-dimensional space.
2. In this space, classes can be easily seperated by a hyperplane.
3. Hyperplane is chosen to maximize the margins of points and hyperplane.

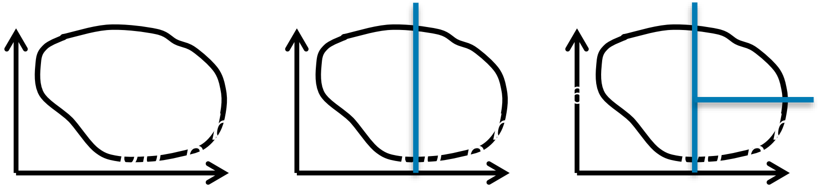


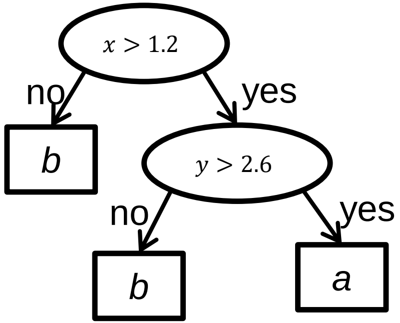
- Actual projection too costly, thus apply the **kernel trick**
  - Only calculate the distances in the high-dimensional space using a kernel function
    - Kernel function is not learnt but manually specified
    - Input data needs to be represented in a suitable way since this is a **shallow** (vs. deep) methods.

[CV95] Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine learning*, *20*(3), pp.273-297.

Decision Trees, Random Forest, GBM

- Decision Trees:
  - Heyday between 2000 and 2010
  - Easy to understand/interpret/visualize





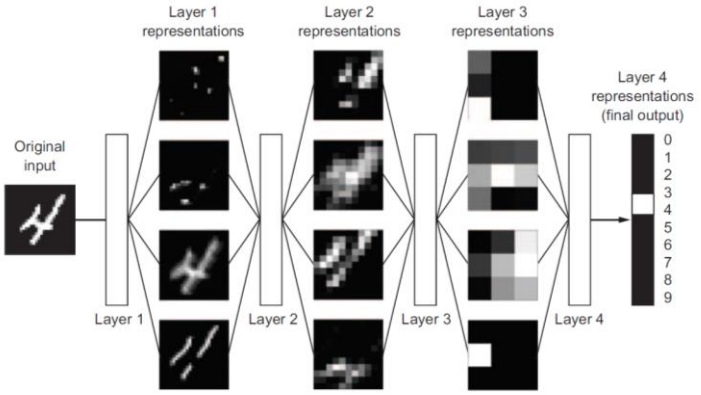
- Random Forest
  - Many small decision trees whose results are combined
  - Very successful in practice (http://kaggle.com)
- Gradient Boosting Maschines (GBM), since 2014 top choice at Kaggle
  - Similar to Random Forest
    - Improvement by focusing during training iterations on previously wrongly classified instances
  - Best algorithm for „non-perceptual“ problems

Deep Neural Nets

- 2010 important research results for Deep Neural Nets
  - Hinton, U Toronto; Bengio, U Montreal; LeCun, NYU
- 2011 first success at image classification using GPU-trained DNNs
- 2012 breakthrough at image classification challenge ImageNet
  - 1.4 millionen training images, 1000 classes
    - 2011 without DL: 74,3% accuracy
    - 2012 Hinton‘s group: 83,6%
    - 2015 problem solved (96,4%)
- Since then, CNNs (convnets) de facto standard in the area of computer vision.
- Many task also in the area of Natural Language Processing (NLP)
- In case of many available training examples, DL has superseded SVMs and decision trees for many tasks.
  - E.g. at CERN for analyzing ATLAS detector data of the LHC

ML Today

- Deep Learning is not only superior based on **better results** for certain tasks, but also because there is no need for **feature engineering** any more.



For shallow problems:
GBM: XGBoost library
For perceptual problems:
DL: Keras/Pytorch library

- Shallow methods (SVM, decision trees) transform input data at most one to two times; for many complex input data this is not enough.
- In deep learning, representations are not learned in isolation, layer-by-layer, but jointly.
  - This leads to more complex representations in each layer.

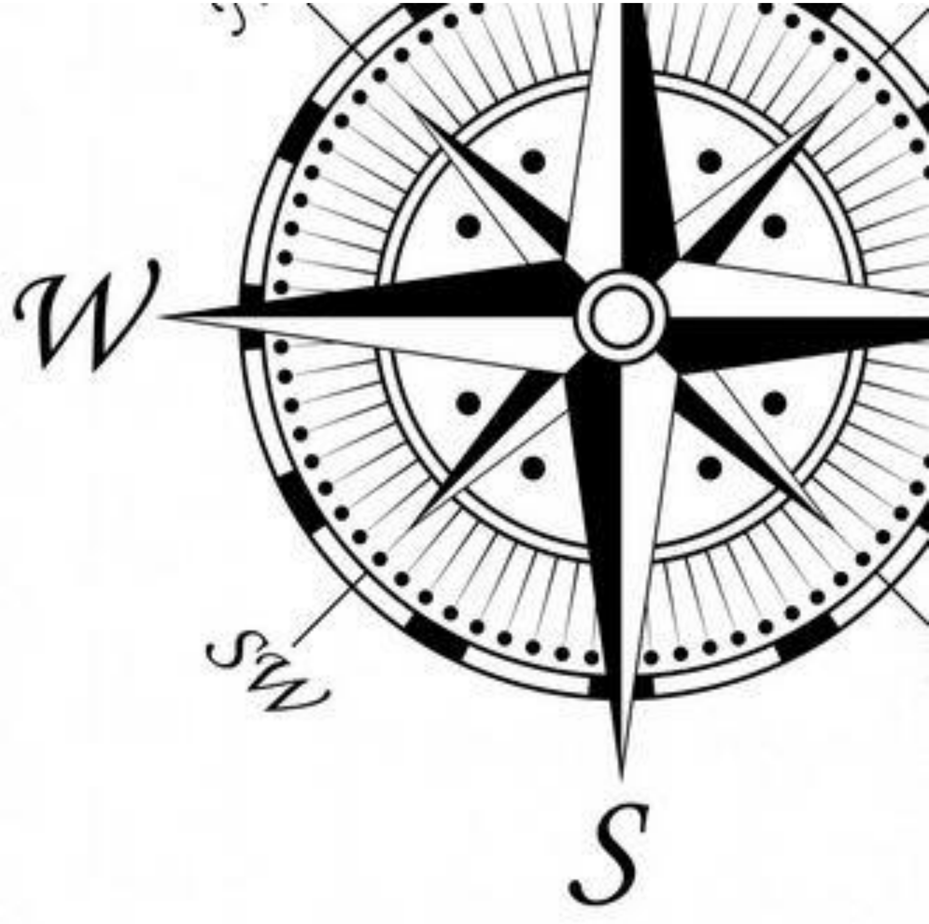
Machine Learning

- Why was DL not already successful during the 1950ies?
- How about chaining five SVMs together, one after the other? Would this be a deep SVM giving good results?
- What are the weaknesses/disadvantages of DL?
- Where and why are SVMs and decision trees still in work, although DL is so successful?

Start 5 4 3 2 1 Ende

Topics Today

1. Organization
2. Introduction to Deep Learning
3. Brief History of Machine Learning
4. **Summary**



Lerning Goals for this Chapter

- Answering the question: am I right here?
- Explain and define deep learning
- Position deep learning with regards to machine learning
- Understand the historical development of the area of ML

- Relevant chapters:
  – P2