Introductory Course: Machine Learning (WWI15B4)

Introduction

Fabio Ferreira, David Bethge

DHBW Karlsruhe

Outline

About us

David Bethge

- **background**: Industrial Engineering with focus on data science
- **interests**: Machine Learning, intelligent driving
- **experience**: various ML applications e.g.:
  - analytics of car logging data
  - life-cycle risk measures for power plants
  - ML for automatic building of virtual cars
  - predictive maintenance for robots
- **research interests**: robust statistical learning, perception, learning embeddings

About us

Fábio Ferreira

- **background**: CS with focus on Machine Learning (ML)
- **interests**: ML/ML for robotics, computer vision, reinforcement learning (RL)
- **experience**: various ML applications e.g.:
  - deep auto encoders for video processing
  - DNN architecture evaluation
  - DNN vs. kernel methods in finance
- **research interests**: fundamental ML without specific applications, cognitive robotics (e.g. advance artificial reasoning/world understanding through

Lecture Overview

[see Syllabus]

Course Material

- course material on Moodle
  / Meine Kurse / DHBW Karlsruhe / KA-Wirtschaft / KA-W-Wirtschaftsinformatik / WWI15 / Kurs WWI15 B4 / WWI15B4: ML
- obtain feedback about enrollment

Lecture Overview

- May 8: Introduction
- May 16: Statistical Learning (1)
- May 24: Concept Learning
- May 29: Statistical Learning (2)
- June 6: Classification (1)
- June 12: Classification (2) (3)
- June 20: Clustering
- July 3: Outlook

No lecture in week June 25-29

Lecture Overview

- May 8: Introduction
- **May 16: Statistical Learning (1) (3h15)** (1pm-4:15)
- May 24: Concept Learning
- May 29: Statistical Learning (2)
- June 6: Classification (1)
- **June 12: Classification 2(+3) (2h30)** (10:45am-1:15pm)
- June 20: Clustering
- July 3: Outlook

Exam / Exam Allocation

Vorlesung	VL-Stunden	Klausurzeit
Neue Konzepte I (Industrie 4.0)	20 (19+K)	35
Neue Konzepte II (Usability)	10 (9+K)	20
Neue Konzepte I (IT Security)	20 (19+K)	35
Neue Konzepte II (machine learning)	30 (28 + K)	60

Exam date: July 23, 10am - 12:30pm

Recommended Literature

or: `https://goo.gl/3ogHU4`

contact

email:

- Fabio Ferreira: ferreira.fabio@edu.dhbw-karlsruhe.de
- David Bethge: bethge.david@edu.dhbw-karlsruhe.de

Applications

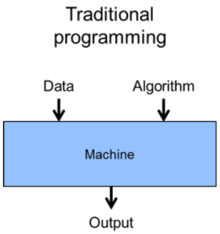
- Automotive
- Bioinformatics
- Computer Vision
- Games
- Financial markets
- Linguistics and Speech
- Insurance
- Marketing
- Object recognition
- Optimization
- Robotics
- Search engines
- ...

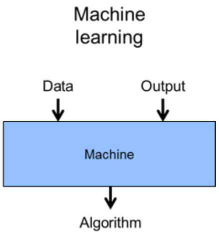
What is learning actually?

- *"Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the same task or tasks drawn from the the* ***same population more efficiently and more effectively*** *the next time."* - Herbert A. Simon
- Knowledge: "Interpretation of the information contained in data" (Kononekno, Matjazkukar, 2007)

Machine Learning - Definition

- Tom M. Mitchell: "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."
- T: Playing checkers game, E: Played games, P: Percentage of won games





Machine Learning - Overview

- Learning from data and making predictions or decisions on data
- Not explicitly, static programmed
- Computer programs, that automatically improve with experience
- Goal: Ability to perform accurately on new, unseen data, after learnt from a training data set.
- Knowledge base is needed
- Problem: How well does the training experience represents the distribution of examples?
- Problem: Why did the computer decided this way?

Machine Learning - Overview

- Often uses statistical and mathematical optimization techniques
- Learning problems often formulated as minimization problems, using a loss function, expressing a discrepancy between prediction and training data.
- Direct vs. indirect training feedback

Machine Learning - Data Mining

- Data Mining: Overlap with Machine Learning but only focusing on data analysis
- Machine Learning: Prediction, based on known properties
- Data Mining: Discovery of unknown properties
- Uncover hidden insights
- Dimensional reduction

Why now?

- Most ideas from the second half of the 20th century
- Today: **Availability of data and possibility to do large computation cheaply**

A little bit of history

It is virtually impossible to find an 'original' idea that was not present earlier in the scientific literature. When the need for an idea is in the air, the idea is usually discovered by several groups simultaneously" - James A. Anderson

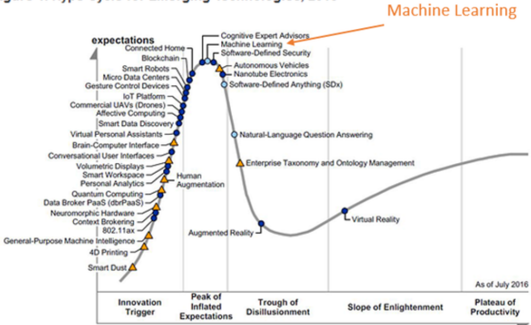
A little bit of history

- 1941: Zuse Z3
- Arthur Samual, 1959 (at IBM): Program, that played checkers better than himself (based on rules and probabilities)
- Modeling of Neurones (Rosenblatt 1958)

A little bit of history

- 1955-1968: Begin
- 1969-1979: "Depression" (XOR-Problem)
- 1979-1986: "Renaissance" (Various methods)
- since 1986: Extensions, more complex systems...
- today: Attempts to explain why a system does what it does.

Figure 1. Hype Cycle for Emerging Technologies, 2016



Classification criteria: Machine Learning

Type of Inference	inductive	deductive
Learning approach	symbolic	non-symbolic
Types of learning	supervised	unsupervised
Example usage	incremental	non-incremental
Extent of examples	extensive	little
Background knowledge	empirical	axiomatic

Type of Inference

- Inductive: Specific to general: Making generalizations from specific, available observations.
- Deductive: General to specific: Based on a theory or formulas we can make a prediction: "Syllogism"

- Example inductive: "Jim is a man", "Jim is mortal" $\rightarrow$ "All men are mortal"
- Example deductive: "All men are mortal", "Jim is a man" $\rightarrow$ "Jim is mortal"

Learning approach

- Symbolic: Using human-readable information related to what you think is necessary. The reasoning process can be easily understood. Rules and knowledge has to be hand coded
- Non-symbolic: Feeding in raw information that can be analyzed and then implicit knowledge can be constructed. It is often difficult to understand how the system came to a conclusion

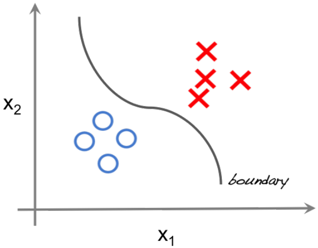
- Example symbolic: ontology, taxonomy, rule-based knowledge representations
- Example non-symbolic: matrices, scalars or vectors

Types of learning

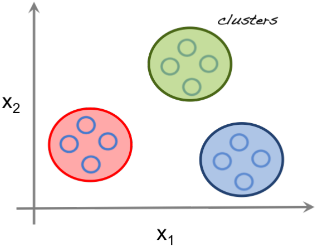
- **Supervised**: With labeled data, with input and output variables
- **Unsupervised**: Without labeled data, without an output variable

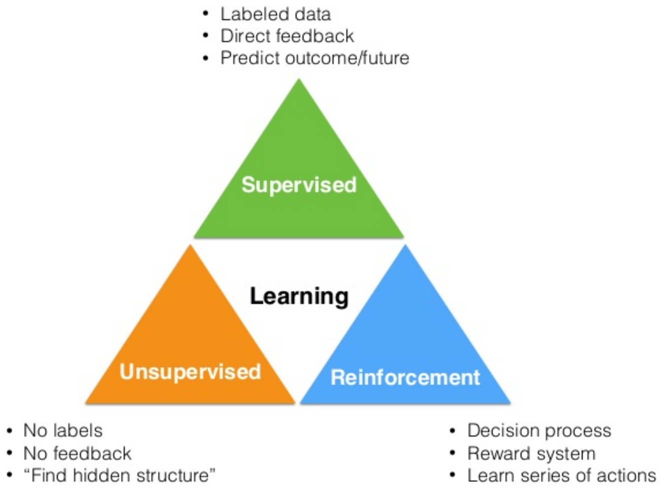
- Example supervised: Classification: Decide whether a picture shows a man or a woman
- Example unsupervised: Model underlying structure: Clustering: Grouping customers by purchasing behavior

Supervised learning



Unsupervised learning




• Labeled data
• Direct feedback
• Predict outcome/future
Supervised
Learning
Unsupervised
Reinforcement
• No labels
• No feedback
• "Find hidden structure"
• Decision process
• Reward system
• Learn series of actions

Example usage

- Incremental: Training data can be severally/successively inserted and the system improves with each example the hypotheses.
- Non-incremental: All data have to be inserted at once.

- Example incremental: Adding new rules based on new examples, which do not affect the old rules based on the old examples:
  {"All men are mortal", "Jim is a man"} $\rightarrow$ "Jim is mortal"
  {"All dogs are mortal", "Garfield is a dog"} $\rightarrow$ "Garfield is mortal"
- Example non-incremental: Clustering a group of people by their salary. If one person with a really different salary is added, all other persons have to be clustered again.

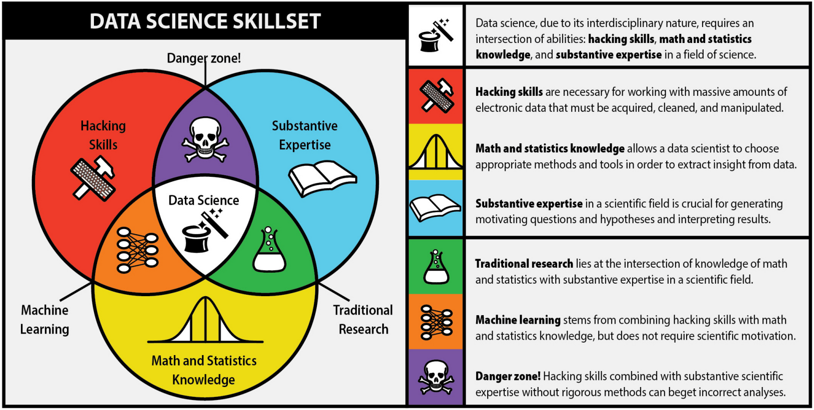
Background knowledge

- Empirical: Statistical analyzed know-how
- Axiomatic: Logic deductions

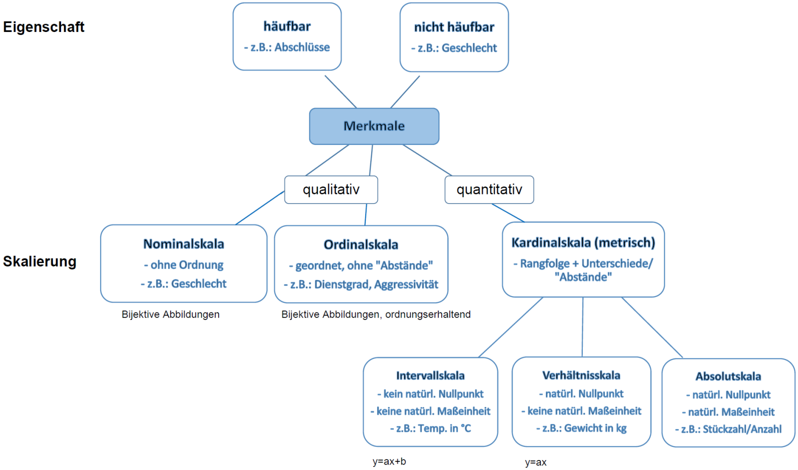
Classification criteria: Machine Learning

Type of Inference	inductive	deductive
Learning approach	symbolic	non-symbolic
Types of learning	supervised	unsupervised
Example usage	incremental	non-incremental
Extent of examples	extensive	little
Background knowledge	empirical	axiomatic

Skillset


DATA SCIENCE SKILLSET
Danger zone!
Hacking Skills
Substantive Expertise
Data Science
Machine Learning
Traditional Research
Math and Statistics Knowledge
Data science, due to its interdisciplinary nature, requires an intersection of abilities: **hacking skills**, **math and statistics knowledge**, and **substantive expertise** in a field of science.
Hacking skills are necessary for working with massive amounts of electronic data that must be acquired, cleaned, and manipulated.
Math and statistics knowledge allows a data scientist to choose appropriate methods and tools in order to extract insight from data.
Substantive expertise in a scientific field is crucial for generating motivating questions and hypotheses and interpreting results.
Traditional research lies at the intersection of knowledge of math and statistics with substantive expertise in a scientific field.
Machine learning stems from combining hacking skills with math and statistics knowledge, but does not require scientific motivation.
Danger zone! Hacking skills combined with substantive scientific expertise without rigorous methods can beget incorrect analyses.

Types of data



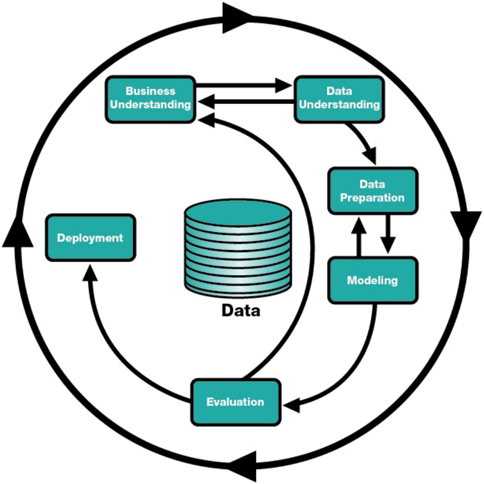
Diskret: „nicht teilbar“, z.B. Anzahl der Studenten in einem Raum
Stetig: „teilbar“, z.B. Volumen

Machine Learning Approach: CRISP-DM

How is a machine learning project done in practice?

- Cross Industry Standard Process for Data Mining
- 1997: Founded by the European Commission
- 1996: DaimlerChrysler, SPSS, NCR
- Today: IBM

CRISP-DM


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data

Business Understanding

What does the customer want to accomplish?

- Uncover important facts
- Record information about the business situation
- Describe primary and related business goals

Produce Project Plan

- Describe the plan to achieve the data mining goals
- Describe success criteria

Business Understanding

Go into detail

- List available resources
- List requirements, assumptions and constraints
- List risks and alternatives

Costs and benefits

- Do a cost-benefit analysis

Determine Data Mining Goals

Data Understanding

- Where is the needed data?
- What: restrictions, type, structure, date, . . .
- Quality: accuracy (outliers), completeness (missing values), consistency (errors)

Data Understanding

- Collect data
- Write data collection report
- Describe data
- **Explore Data: Analyze, visualize (scatter-plot, histogram etc.**
- Examine the quality of the data
- Data exploration and quality report

Data Preparation

- **Goal**: Dataset that can be used to model (apply machine learning algorithms)

Data Preparation: Selection

- Merge data if necessary
- Select attributes as well as observations
- Observations: Sampling with or without replacement
- Observations: Systematic sampling, stratified sampling, bootstrap sampling
- Attributes: Select subset, missing values,too much or too little variation, correlations, ...
- Attributes: Dimension reduction

Data Preparation: Selection

Filter vs. Wrapper vs. Embedded techniques

- Filter: Evaluation of attributes independent of modeling technique
- Wrapper: Using performance of learning method for evaluation
- Embedded techniques: Attribute selection part of the method

Data Preparation: Cleaning

- Check for missing values, inaccurate data, duplicates, outliers, ...
- Deal with errors by selecting a clean subset or replacing errors

Missing Values

- Ignore
- Replace with the most common attribute value
- Replace with the most common attribute value of the same class
- Try to model: Use for example probability distribution

Data Preparation: Construction and Transformation

- Transform data (e.g. year of birth to age)
- Construct derived attributes (e.g. PCA) or observations (sampling e.g. bootstrap)

Modeling

- Select concrete modeling technique
- Record model assumptions
- Describe how to and separate data into training, validation and test data
- **Run model**
- List parameter settings
- Describe produced model with interpretation and difficulties
- Run different models and compare them: Different parameters and/or different methods
- Summarize results

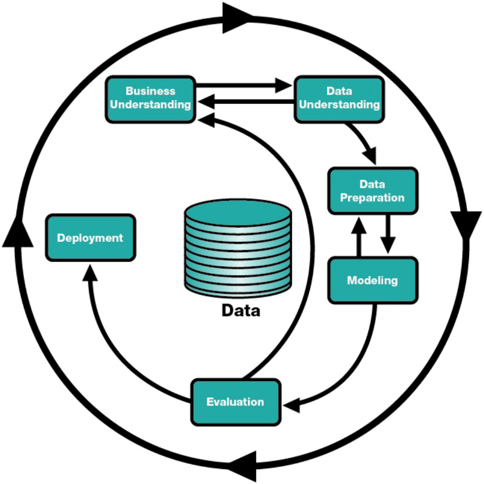
Evaluation

- Evaluate models with respect to the business objectives
- Review the whole process: Are the models correctly built? Were only attributes used that will also be available in the future?
- Decide how to proceed and document everything

Deployment

- Describe deployment plan
- Plan and summarize monitoring and maintenance
- Final report
- Review project, document and learn for future projects

CRISP-DM


Business Understanding
Data Understanding
Data Preparation
Modeling
Evaluation
Deployment
Data

Notation

Data matrix X
For supervised learning: Output matrix Y
n: number of observations: $i \in \{1, \ldots, n\}$
p: number of attributes: $j \in \{1, \ldots, p\}$

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{pmatrix} \qquad X = \begin{pmatrix} x_{11} & \ldots & x_{1j} & \ldots & x_{1p} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \ldots & x_{ij} & \ldots & x_{ip} \\ \vdots & & \vdots & & \vdots \\ x_{n1} & \ldots & x_{nj} & \ldots & x_{np} \end{pmatrix} = \begin{pmatrix} X_{(1)} \\ \vdots \\ X_{(i)} \\ \vdots \\ X_{(n)} \end{pmatrix}$$

For simplicity: $X_{(i)} = x_i$

Learning Theory

- Shortcut of Fabios and Davids 4th chapter.

Overview methods

Methods	Inference: inductive	Inference: deductive	Learning approach: symb.	Learning approach: non-symb.	Types of learning: sup.	Types of learning: unsup.
Regression	×		×		×	
Decision Tree	×		×		×	
K-nearest neighbors	×			×	×	
SVM	×			×	×	
K-means (clustering)	×			×		×
PCA	×		×			×
NN	×			×	×	

	Example usage: incr.	Example usage: non-incr.	Extent of examples: extensive	Extent of examples: little	Background knowledge: empirical	Background knowledge: axiomatic
Regression		×	×		×	
Decision Tree	×	×	×		×	
K-nearest neighbors		×	×		×	
SVM		×	×		×	
K-means (clustering)		×	×		×	
PCA		×	×		×	
NN		×	×		×	

Reading assignment

Next lecture we will need some basic matrix algebra.