Institute of Automotive Technology
Department of Mechanical Engineering
Technical University of Munich

MDA – Part 07: Machine Learning Project

Sebastian Krapf and David Ziegler

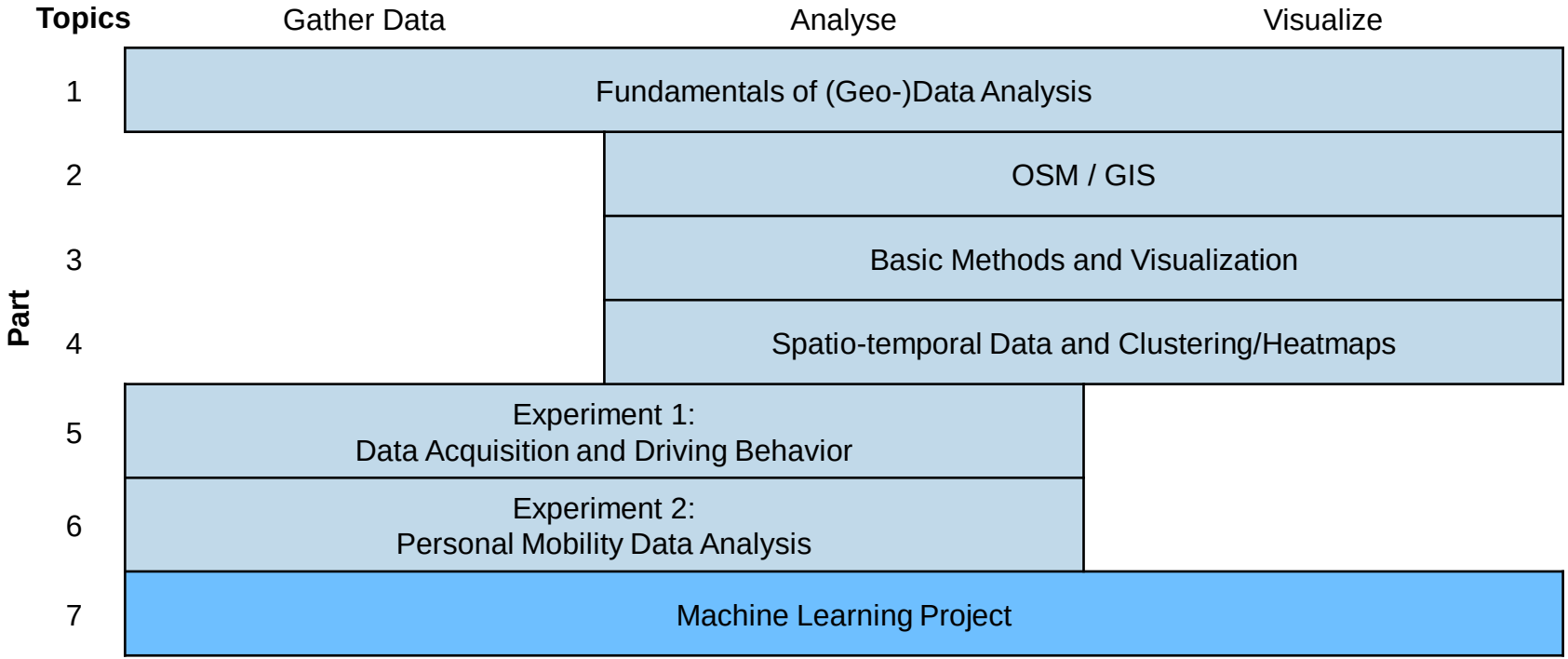
Technical University of Munich

Department of Mechanical Engineering

Institute of Automotive Technology

Institute of Automotive Technology
Department of Mechanical Engineering
Technical University of Munich

Course Overview



Contents

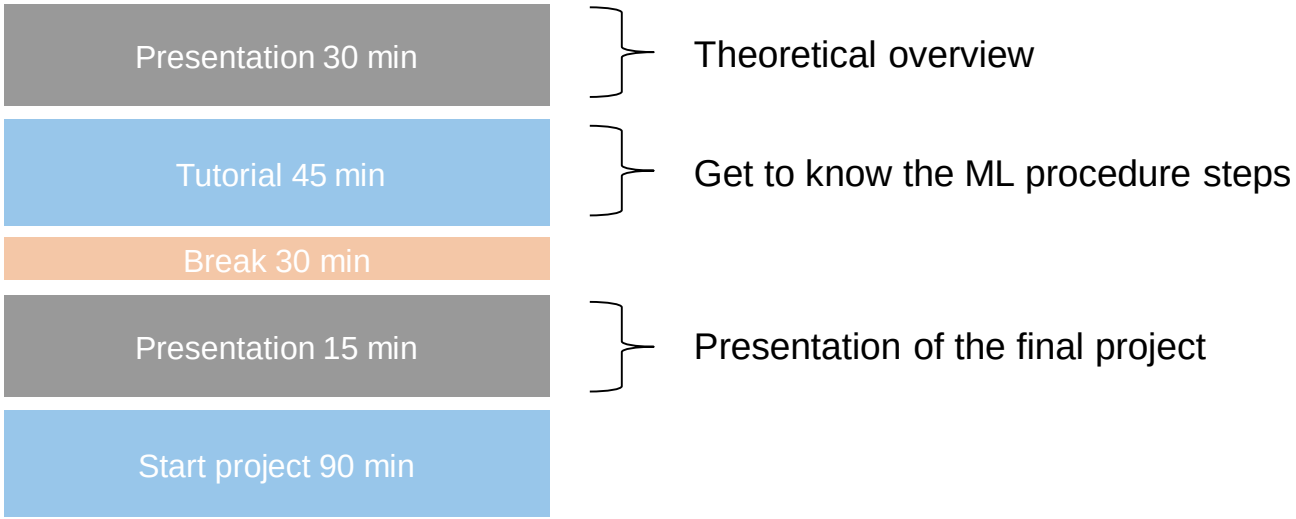
1. Overview / Definition of ML terms
2. Machine Learning methods
3. AI-Frameworks

4. ML Tutorial
5. Introduction Project

Institute of Automotive Technology
Department of Mechanical Engineering
Technical University of Munich

Today's Timetable

2 Blocks



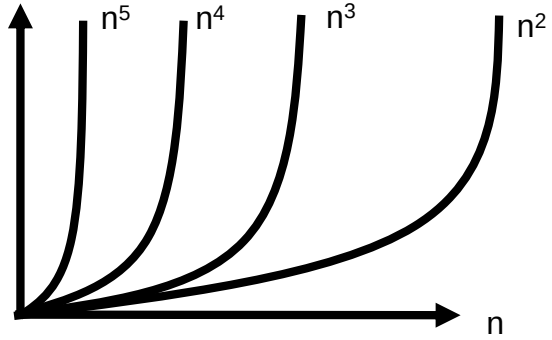
Institute of Automotive Technology
Department of Mechanical Engineering
Technical University of Munich

Depth of Comprehension

Objective	Remember	Understand	Use	Analyse	Evaluate	Create
Deal with common analysis tools	■	■	■	■		
Development of an open source data analysis toolchain	■	■	■			
Capturing data with smartphone and datalogger	■	■	■			
Datatypes, Formating and Conversation	■	■	■	■		
Generation of metadata and basic Analysis	■	■	■	■	■	
Clustering / Heatmap	■	■	■	■		
GIS and digital maps	■	■	■			
Data visualisation	■	■	■	■		
Machine-Learning with spatio-temporal data	■	■	■			

Theory

Why AI - Whats the problem?



Computational Complexity

Source: https://jalopnik.com/for-48-000-could-this-2008-factory-five-gtm-super-car-1786246861

Information Complexity

Unsupervised Learning

AI – Artificial Intelligence

Deep Neural Networks

KI – Künstliche Intelligenz

Perception

Supervised Learning

Clustering

Classification

Reinforcement Learning

Regression

Neural Networks

Convolutional Neural Networks

Machine Learning

What is Artificial Intelligence? A proposal for categories

Breaking down the general problem of creating AI into 9 sub-problems:

1. **Reasoning & Problem Solving:** A machine gets the ability for step-by-step reasoning by making logical deductions with uncertainty

2. **Knowledge Representation:** Representing information about the world in a form that a computer system can utilize to solve complex tasks

3. **Planning:** A machine gets the ability for an optimized automated planning or scheduling that leads to action sequences

4. **Learning:** A machine gets the ability to “learn” based on algorithms that improve automatically through experience and data without being explicity programme **(Machine Learning)**

What is Artificial Intelligence? A proposal for categories

Breaking down the general problem of creating AI into 9 sub-problems:

5. **Natural Language Processing:** A machine gets the ability to read and understand human language

6. **Perception:** A machine gets the ability to use input from sensors for deducing aspects of the world and sensing the environment around the machine

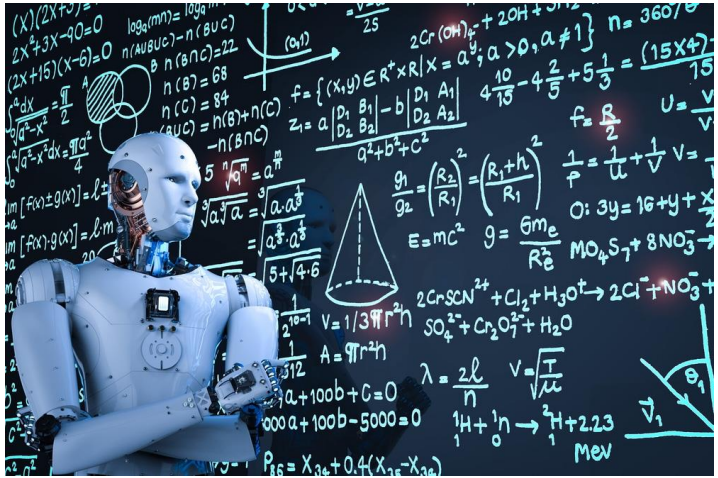
7. **Motion and Manipulation:** A machine gets the ability to learn how to plan their motion and move efficiently

8. **Social Intelligence:** A machine gets the ability to recognize, interpret, process, and simulate human affects

What is Artificial Intelligence? A proposal for categories

Breaking down the general problem of creating AI into 9 sub-problems:

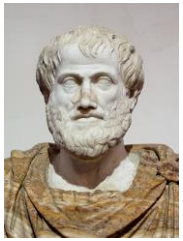
9. General Intelligence: Achieving the full range of human cognitive abilities
(= general / strong / full AI)



Source: https://inform.tmforum.org/catalyst/2018/01/smart-bpm-catalyst-makes-ai-explainable/

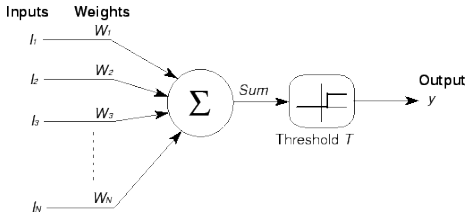
Institute of Automotive Technology
Department of Mechanical Engineering
Technical University of Munich

A brief history



300 BC: Aristoteles
Described syllogism

1641: Hobbes
Theory of cognition



1943: McCulloch & Pitts –
Foundations for artificial
neuronal network

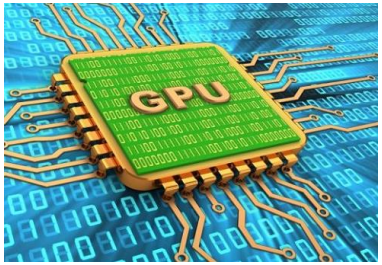
**2005: AI Big Bang –
GPUs and Data**

2009: Google – Self
Driving Car

2016: Google AlphaGo
Defeat Human in Go Game

2018: Google Duplex
Personal Assistant

AI Methods – Why now?



1. Data, Labeled Data, Knowledge is available: **Big Data**

2. New AI Algorithms are available: **Deep Learning**

3. Computer power is available: **GPU**

Source: https://qlu.ac.pa/english/3-de-diciembre-conferencia-internacional-gratuita-competitividad-sustentable-utilizando-analytics-big-data/
https://ai.googleblog.com/2017/05/using-machine-learning-to-explore.html
https://www.hpcwire.com/2018/03/27/nvidia-riding-high-as-gpu-workloads-and-capabilities-soar/

ML Applications – Automotive Technology

Predictive Maintanence

Problem / Motivation

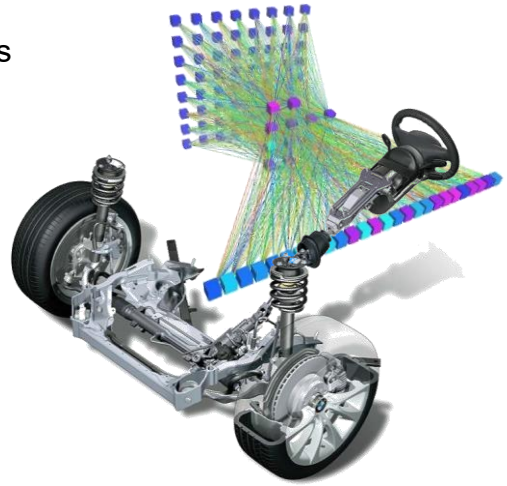
- Decreasing driver‘s perception for suspension wear and change of vehicle dynamics

Goals

- Development of an automated diagnosis system
- Detection of chassis system defects based on different sensors

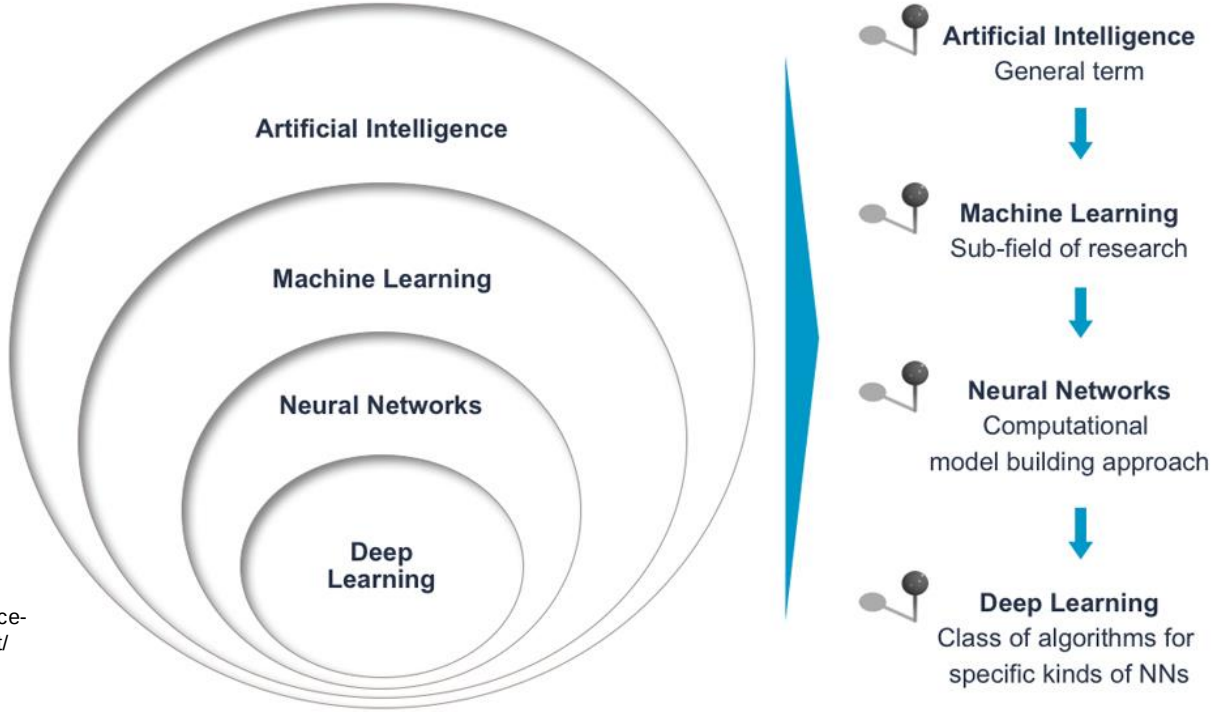
Approach

- Generation of measurement data with different component defects
- Classification of measurement data by machine learning algorithm
- Anomaly detection algorithms to use only healthy data for training



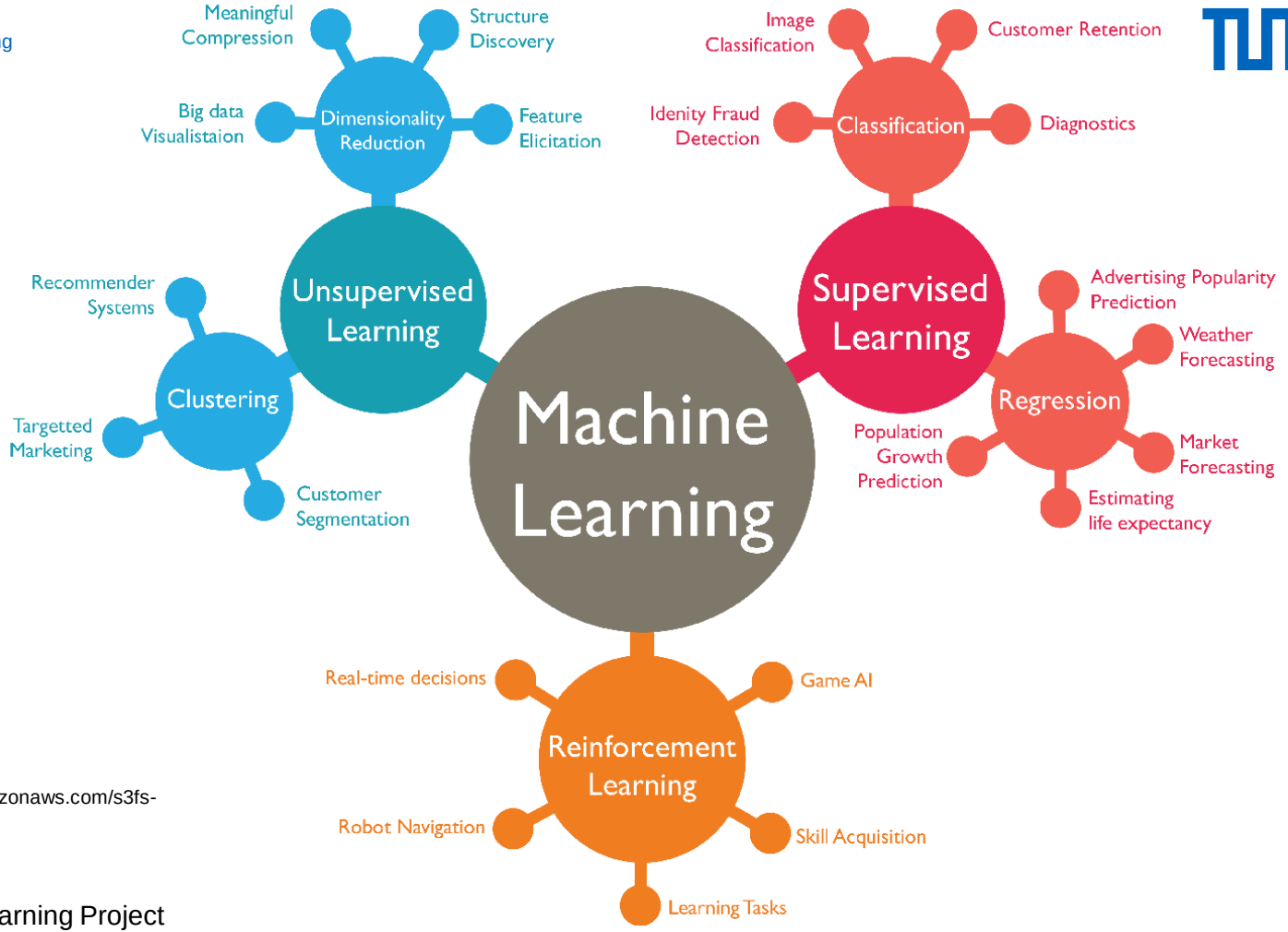
Definition / Classification of AI

Artificial Intelligence (AI) or
„Künstliche Intelligenz“ (KI)



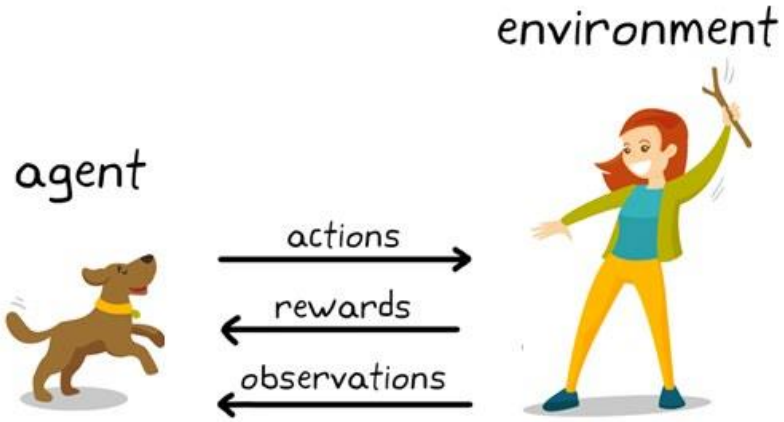
https://www.capgemini.com/de-de/2017/09/artifical-intelligence-machine-learning-und-data-science-same-same-but-different/

Institute of Automotive Technology
Department of Mechanical Engineering
Technical University of Munich



https://wordstream-files-prod.s3.amazonaws.com/s3fs-public/machine-learning.png

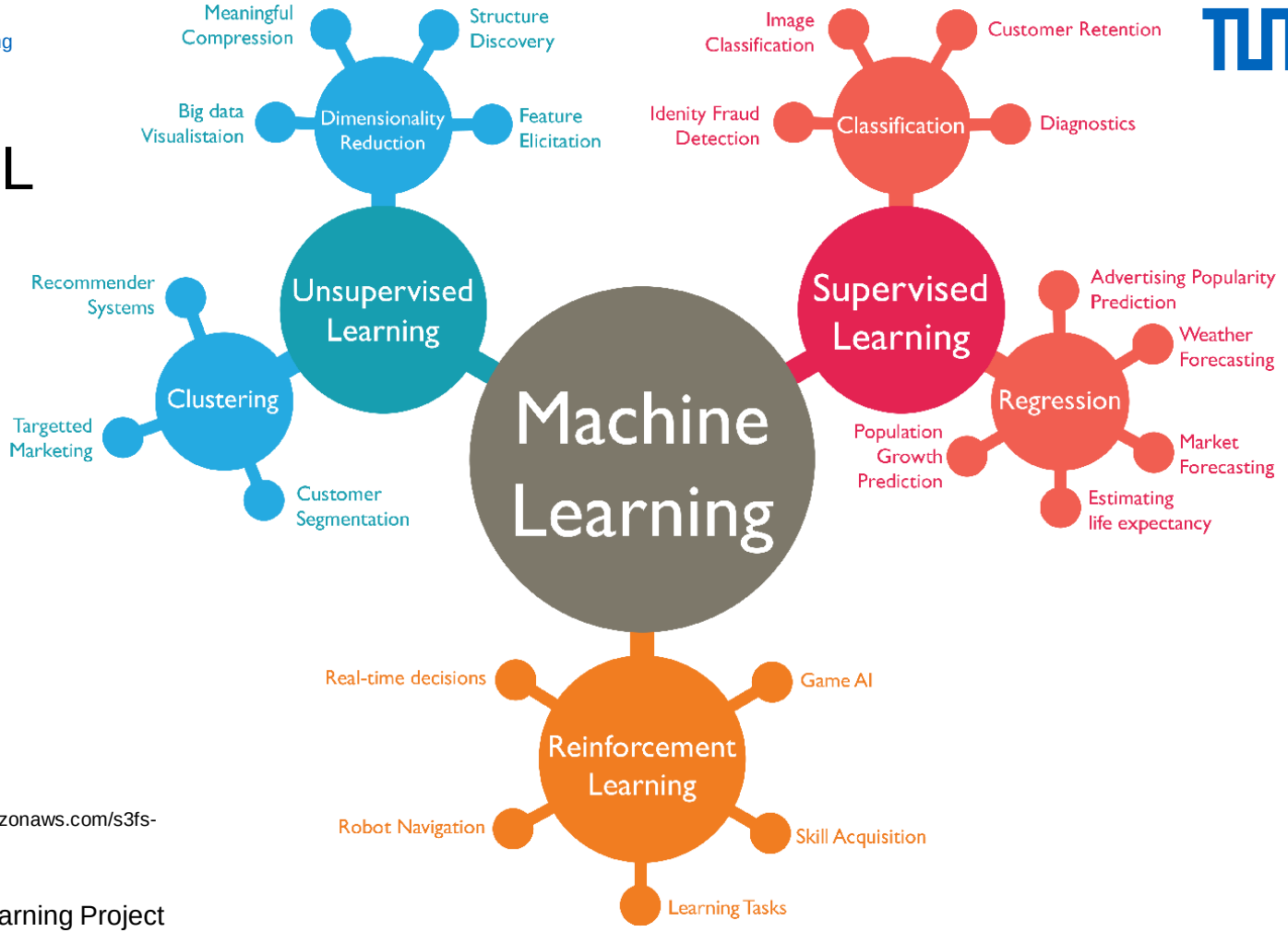
Reinforcement Learning



- No data required in advance
- Trial and error procedure
- Maximize rewards

Institute of Automotive Technology
Department of Mechanical Engineering
Technical University of Munich

Übersicht ML



https://wordstream-files-prod.s3.amazonaws.com/s3fs-public/machine-learning.png

ML Methods

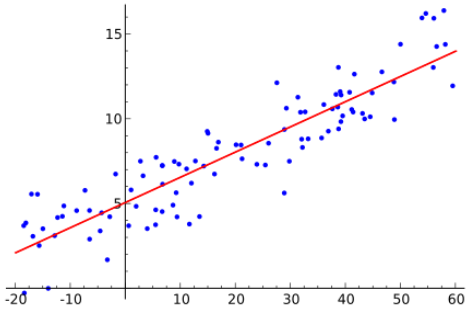
Supervised

Find a predictive model based on input and **labeled** output data

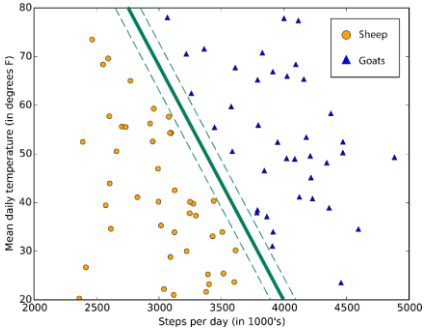
Unsupervised

Find similarities in input data and interprete them

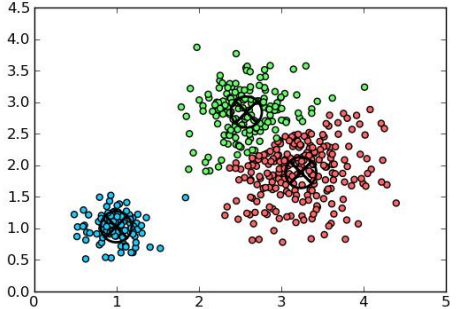
Regression



Classification



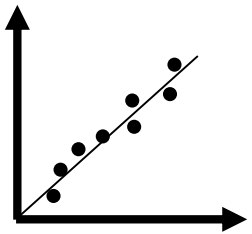
Clustering



Source: https://upload.wikimedia.org/wikipedia/commons/thumb/3/3a/Linear_regression.svg/438px-Linear_regression.svg.png
https://docs.microsoft.com/en-us/azure/machine-learning/studio/media/algorithm-choice/image7.png
http://blog.mpacula.com/2011/04/27/k-means-clustering-example-python/

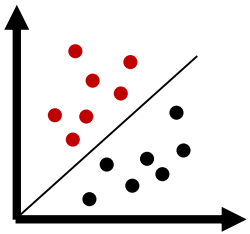
Examples

Regression



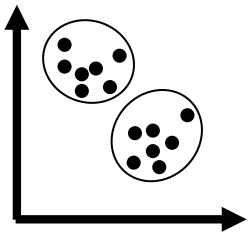
- House pricing
- Sales
- Persons weight

Classification



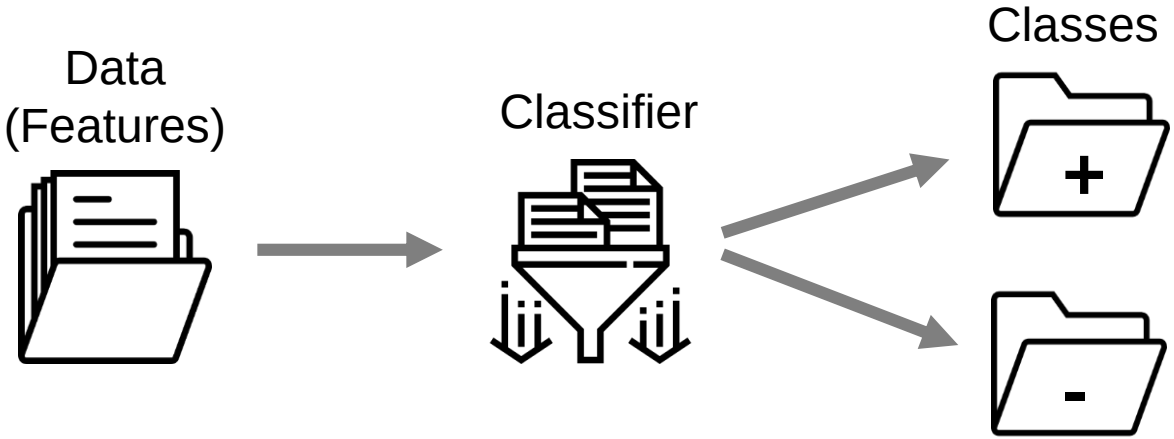
- Object detection
- Spam detection
- Cancer detection

Clustering



- Genome patterns
- Google news
- Pointcloud (Lidar) processing

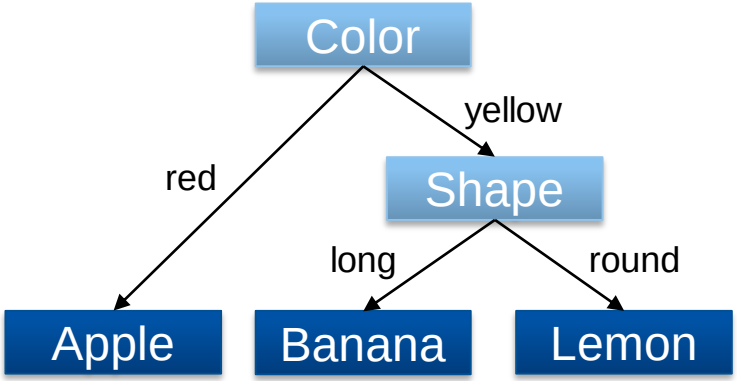
Classification Algorithms - General Approach



Email (Keywords, …) Spam?

Object (Color, …) What type?
Cat/Car/Fruit/…

Classic Method - Decision tree



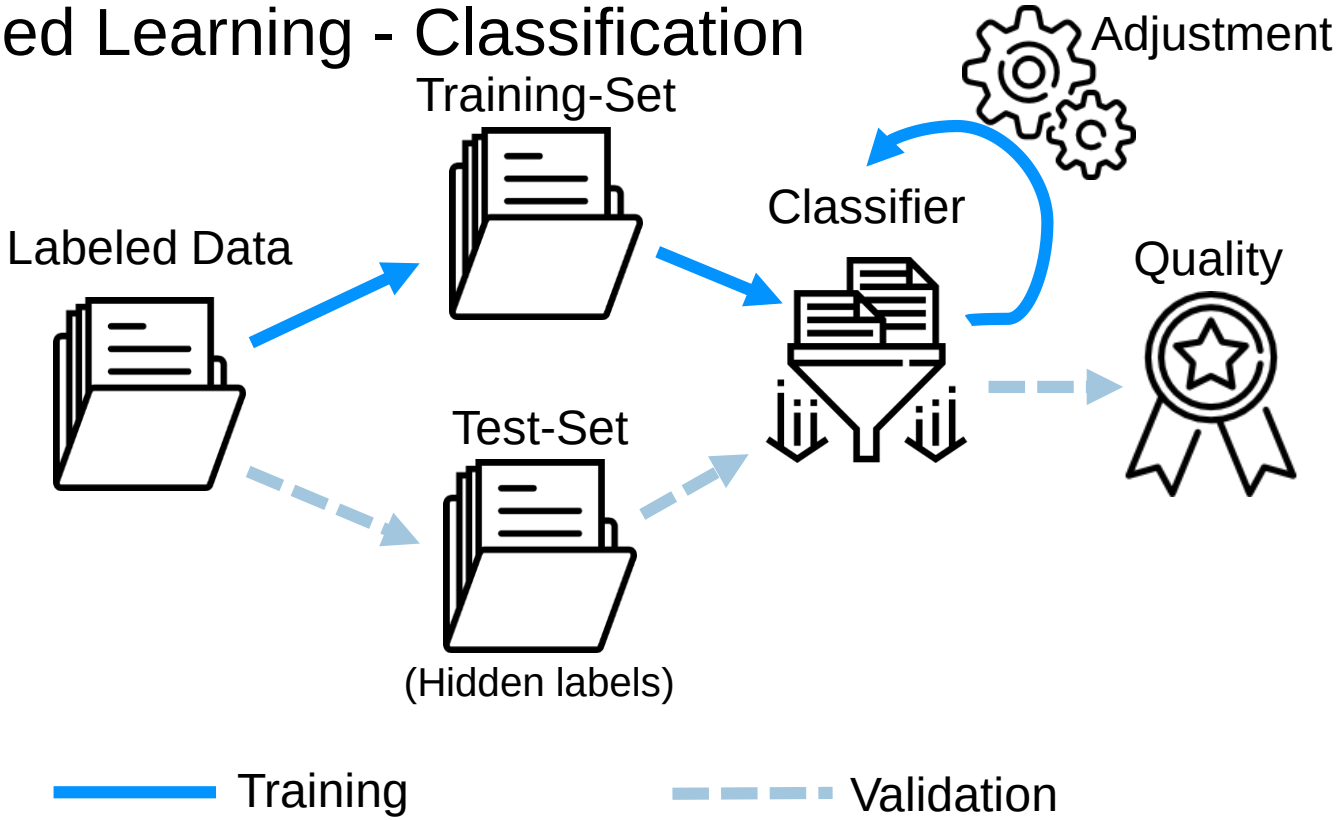
E.g. Decision tree

- Use **a-priori knowledge** to formulate classification rules

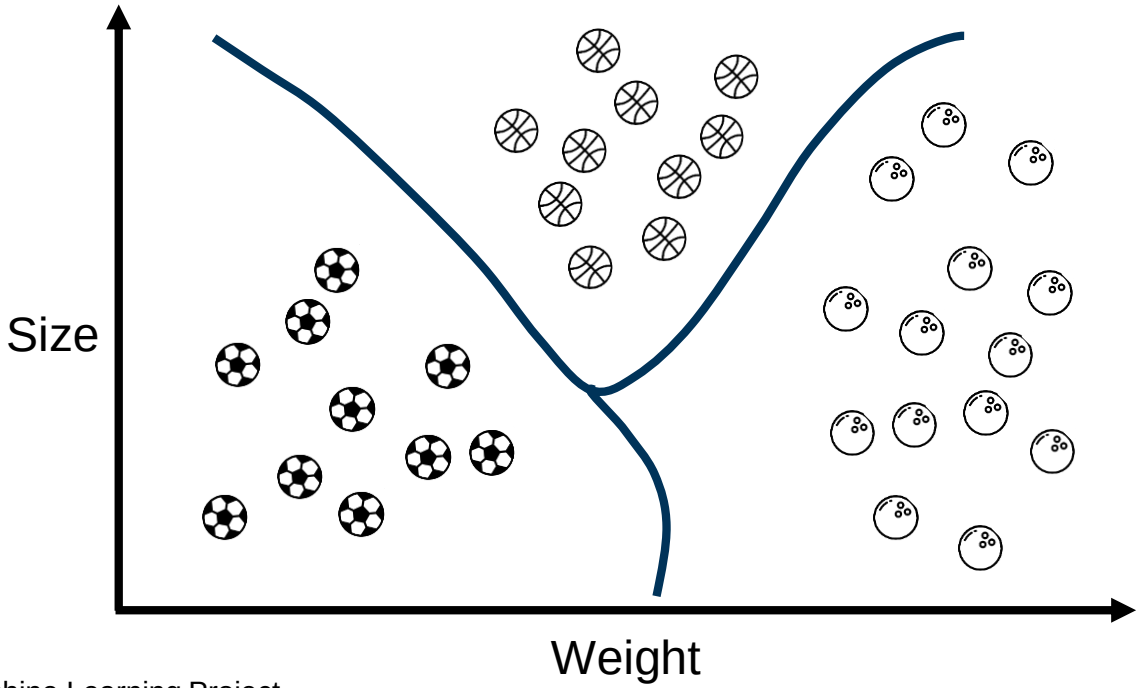
- Advantages of machine learning
  - Automatic generation of a-priori knowledge
  - Automatic generation of complex classification rules

  → Suitable for extreme large datasets

Supervised Learning - Classification



Classifier Training



Classification Methods

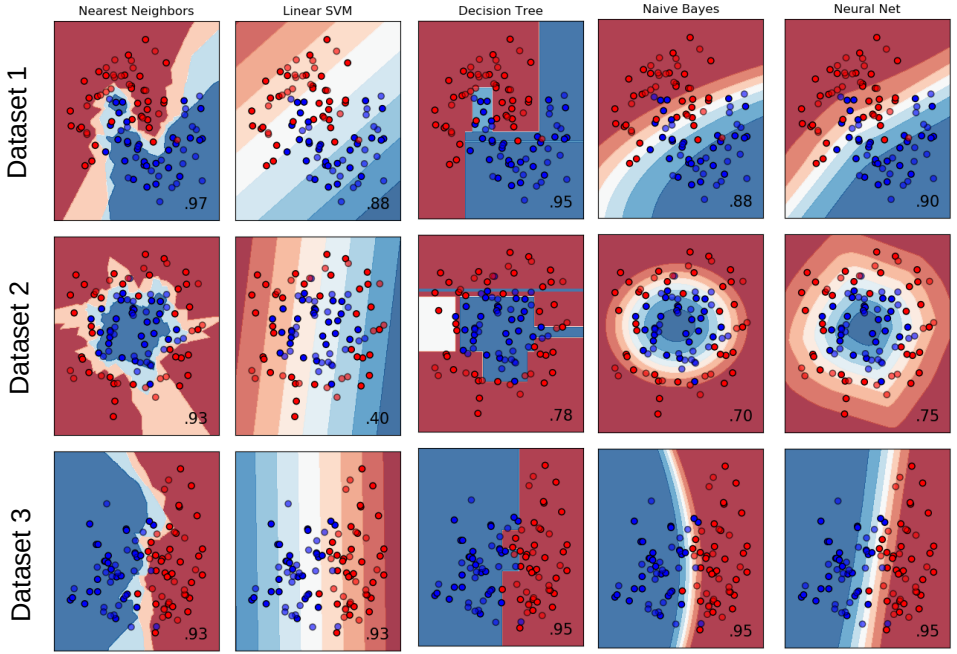
Nearest Neighbors

Support Vector Machine

Logistic Regression

Neural Networks

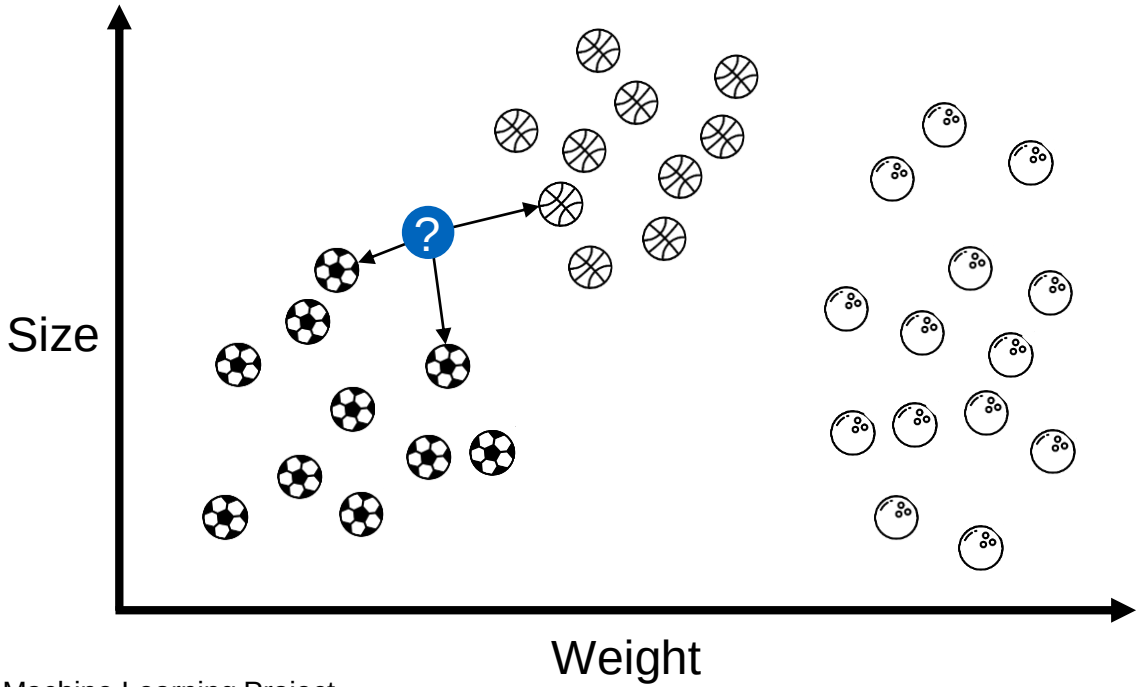
…



Source: http://scikit-learn.org/stable/auto_examples/classification/plot_classifier_comparison.html

Institute of Automotive Technology
Department of Mechanical Engineering
Technical University of Munich

K-Nearest-Neighbors (k-NN) for Classification

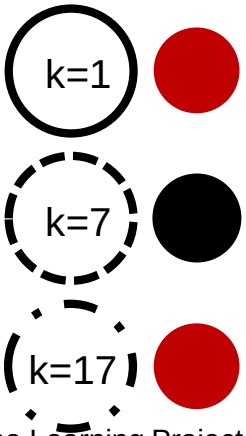


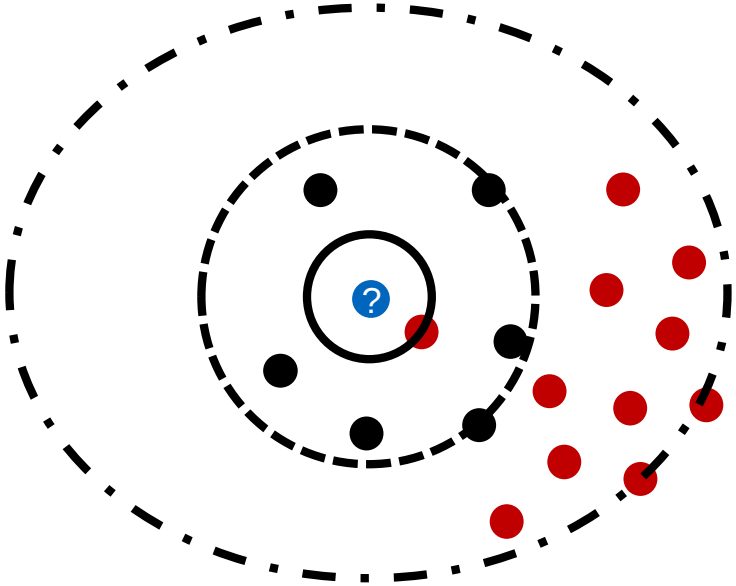
- Consider k nearest neighbors (k>1)

k-NN for Classification

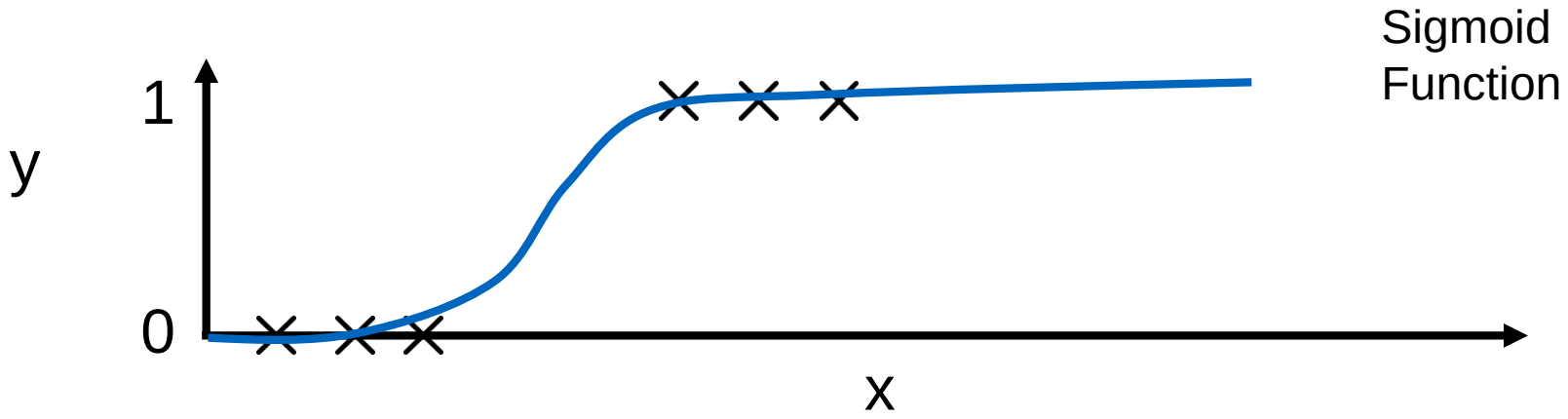
How to choose k?

- Generalization vs Overfitting
- Large k: Many objects from different classes
- Small k: Sensitivity against outliers
- Practice: 1 << k < 10





Logistic Regression for Classification



$$y = h_\theta(x), \qquad y \epsilon \mathbb{R}$$

AI-Frameworks – Scikit-learn Library

Free software framework
Machine Learning Library
Language: Python, C, C++
Operating System: Linux, macOS, Windows
Includes: Clustering, Regression, Clustering with algorithms like SVM, Nearest Neighbors, Gaussian Process, Decision Trees

Pros: Everything you need, Good documentation, powerful, GPU boost
Cons: Not for hardcore statistics, limited in parameters

Source: https://scikit-learn.org/stable/

AI-Frameworks – Matlab

Commercial software (free for **students)**
Machine and Deep Learning toolbox
Language: Matlab, Simulink
Operating System: Linux, macOS, Windows
Includes:

- Machine Learning
- Deep Learing

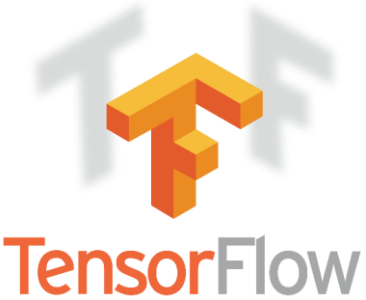
Pros: Easy to use, good documentation, GPU boost
Cons: Closed environment, performance

Source: https://www.cbcity.de/portfolio/matlab-simulink

AI-Frameworks – Tensorflow

Free software framework
Deep Learning software framework
Language: Python, C++
Operating System: Linux, macOS, Windows
Includes: CNN, RNN, → Voice and image recognition

Pros: High performance, multiple GPU, connects research and production, true portability, tensorboard for visualization, good documentation
Cons: Hard to learn in comparison to other frameworks



Source: https://www.tensorflow.org/

Practice

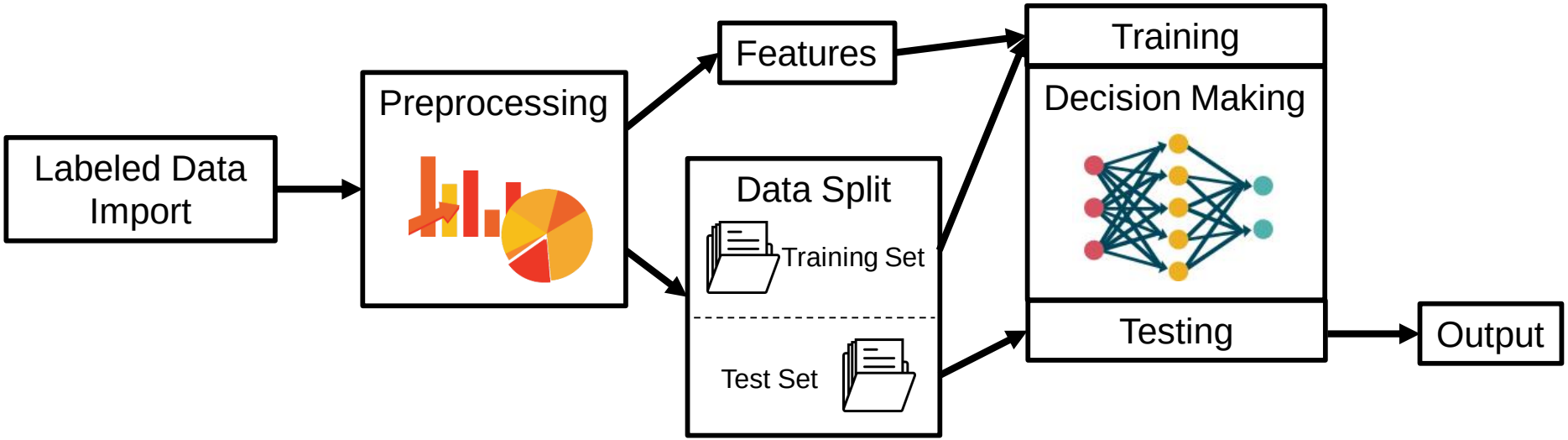
Practice Task

Open Jupyter Notebook „*Classification_Practice.ipynb*“

Follow the task step by step:

1. Import „*Iris.csv*“
2. Preprocessing & Feature generation
3. Train / Test split
4. ML-Classification Models
5. Evaluation classification accuracy

Procedure Steps



Machine Learning Project

Dashboard / Machine Learning Project

General conditions

- Built teams of 3-4 Students (division of tasks!) -> Group name
- Initial data set „David“ + „London Data“
- Mode classification with Machine Learning
- Dashboard visualization

- Presentation of project results: 19.01.2022
- Each team should give a 15 min presentation (concept, implementation, results)

Dashboard / Machine Learning Project

Challenge: Implement the best mode classification!

1. **Generate and visualize metadata in your dashboard**
   a. Number of tracks per transportation mode
   b. Distance / duration per mode
   c. Average and maximum velocity / acceleration per mode
   d. Average distance / duration / velocity per track per mode

2. **Visualize the data with a heatmap in your dashboard**

3. **Create a classification using Machine Learning**
   a. Import and preprocess labeled data
   b. Extract relevant features
   c. Split training and test data
   d. Generate and train one model based on the features
   e. Evaluate and visualize the output in your dashboard

Part 02: OSM/GIS
Routing / Map Matching

Part 03: Basic Methods
Statistical Measures / Plots

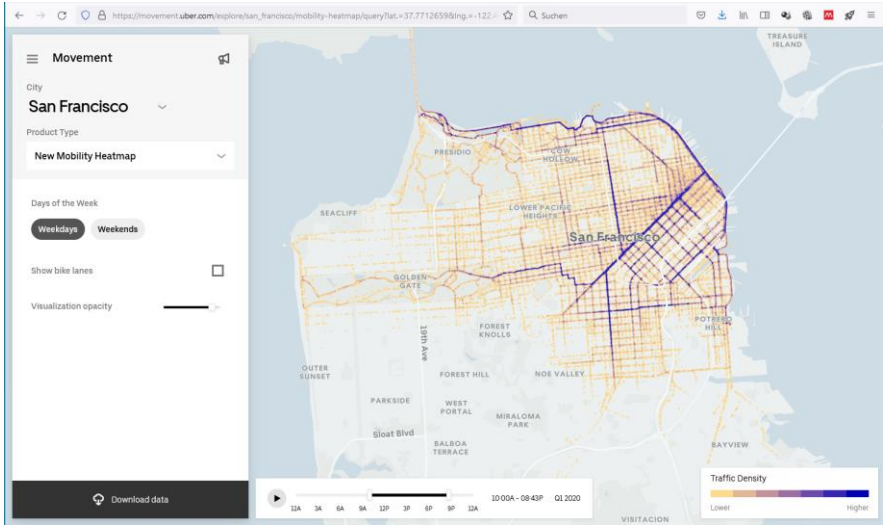
Part 04: Spatio-temporal Data
Heatmap

Part 07: Machine Learning
Classification

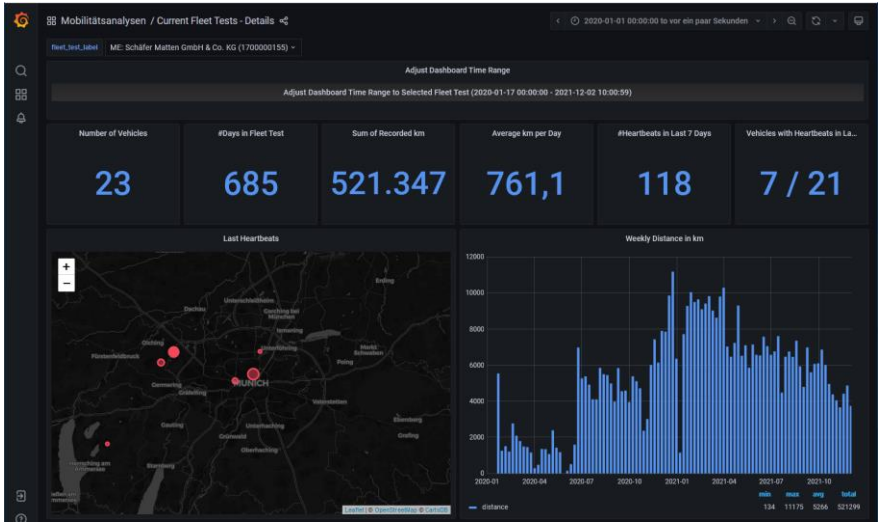
Part 07: Dashboards

Live-Demo

Dashboard | Examples



https://movement.uber.com/explore/san_francisco/mobility-heatmap/



TU München - FTM

Dashboard | Grafana

What are the best open source dashboard frameworks?
https://www.slant.co/topics/9940/~open-source-dashboard-frameworks

Machine Learning Project

Data set „London Data“

- GPS trajectory data set
- 2 users labeled their trajectories with transportation mode (car, bus, bike, walk ...)
- Every folder of this dataset stores a user’s GPS log files

3.1 00inf.txt

Metadata file comprising information about the recording.

Line	What
1	User ID (e.g. User1, User2)
2	timemsmin: first sample time in milliseconds.
3	timemsmax: last sample time in milliseconds.
4	Recording start date in human readable format.
5	Recording length in milliseconds.
6	Recording ID, i.e. name of the folder in which the data is stored (e.g. ‘120617’)

Data set „London Data“

3.10 <position>_Location.txt

This file contains one line per sample. The columns are as follows

Column	What
1	Time ms
2	Ignore
3	Ignore
4	Accuracy of this location (accuracy as the radius of 68% confidence) [m]
5	Latitude [degrees]
6	Longitude [degrees]
7	Altitude [m]

Institute of Automotive Technology
Department of Mechanical Engineering
Technical University of Munich

Data set „London Data“

3.2 <position>_Motion.txt

This file contains one line per sample, all sampled at 100 Hz. Some columns may contain NaN if the information is not available (e.g. not all sensors start sampling at the exact same time). The columns are as follows:

Column	What
1	Time ms
2	Acceleration X [m/s^2]
3	Acceleration Y [m/s^2]
4	Acceleration Z [m/s^2]
5	Gyroscope X [rad/s]
6	Gyroscope Y [rad/s]
7	Gyroscope Z [rad/s]
8	Magnetometer X [µT]
9	Magnetometer Y [µT]
10	Magnetometer Z [µT]
11	Orientation w
12	Orientation x
13	Orientation y
14	Orientation z
15	Gravity X [m/s^2]
16	Gravity Y [m/s^2]
17	Gravity Z [m/s^2]
18	Linear acceleration X [m/s^2]
19	Linear acceleration Y [m/s^2]
20	Linear acceleration Z [m/s^2]
21	Pressure [hPa]
22	Altitude derived from the pressure sensor; for all recordings of User1 after 110517 this value is 0.
23	Temperature derived from the pressure sensor; for all recordings of User1 after 110517 this value is 0.

Data set „London Data“

3.12 labels_track_main.txt

This file contains one line per label. The columns are as follows

Column	What	
1	Label start time in millisecond	
2	Label end time in millisecond	
3	Activity label	
	Still;Stand;Outside:	0
	Still;Stand;Inside:	1
	Still;Sit;Outside:	2
	Still;Sit;Inside:	3
	Walking;Outside:	4
	Walking;Inside:	5
	Run:	6
	Bike:	7
	Car;Driver:	8
	Car;Passenger:	9
	Bus;Stand:	10
	Bus;Sit:	11
	Bus;Up;Stand:	12
	Bus;Up;Sit:	13
	Train;Stand:	14
	Train;Sit:	15
	Subway;Stand:	16
	Subway;Sit:	17