Know Center

AUSTRIA'S LEADING RESEARCH CENTER
FOR DATA-DRIVEN BUSINESS AND BIG DATA ANALYTICS

Artificial Intelligence
An Introduction

Dr. Roman Kern
Graz, 2018-12-06

KNOW-CENTER GMBH

Austria's leading research center for
Data-driven Business und **Big Data Analytics**

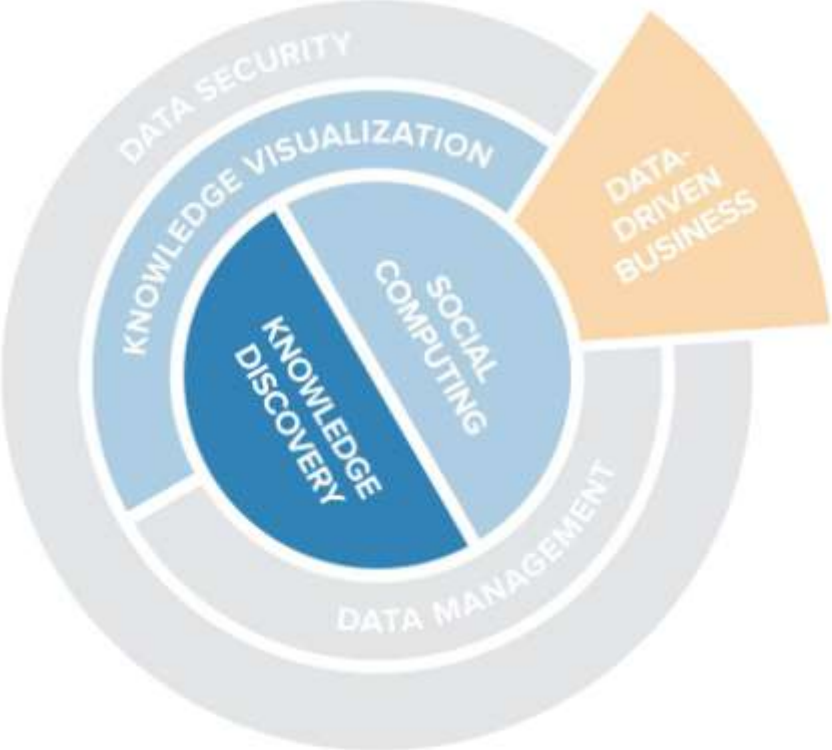
> 90 Researches
> 600 COMET* & industrial projects
> 30 international EU projects

Founded in 2001
Located at the Technical University of Graz

* COMET is an Austrian funding program that supports a sustainable knowledge transfer from science to industry.

KNOW-CENTER

RESEARCH AREAS



Knowledge Discovery

20+ data scientists, researchers and developers

ROMAN KERN
Area Head

(Big) Data → The new **oil**

Artificial Intelligence → The new **electricity**

Data Science → The new **Latin**

DEFINITIONS

https://medium.com/@wtrsld/big-data-made-me-do-it-5bfc3f46871c

BIG DATA

DEFINITIONS OF BIG DATA

Data too big for MS Excel

1. Common: **The 3 Vs**
   - Volume, Variety, Velocity
2. Data too big for a **RDBMS** to handle
3. Given enough data, **new hypothesis** will emerge out of the data
4. To **connect/align data sources**, which have been traditionally been separate
5. „Big Data“ **is just like** „Small Data“, just bigger
6. Create new **business opportunities** through data
   - The 4th V → Value (Veracity, …)

Results of Big Data Analytics

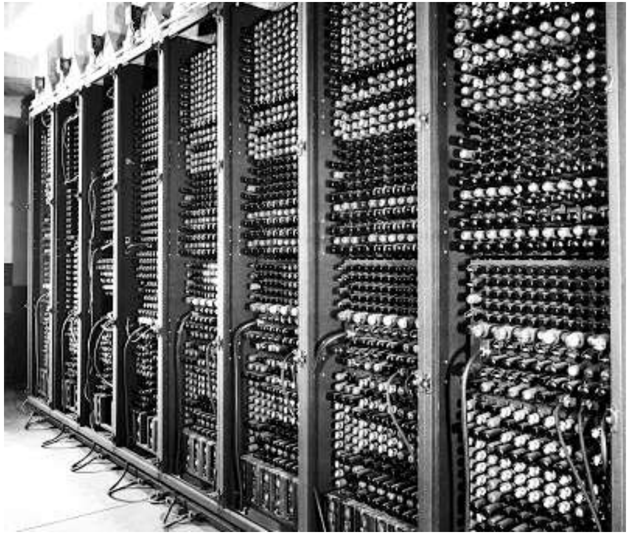
- Correlations
  - E.g. one event often precedes another event
  - Does not mean, that the one event causes the other event
- Probabilities
  - E.g. for a certain constellation of parameters there is a 80% likelihood of a certain outcome
  - Does not mean, that in all cases this will happen

Security Aspects of Big Data Analytics

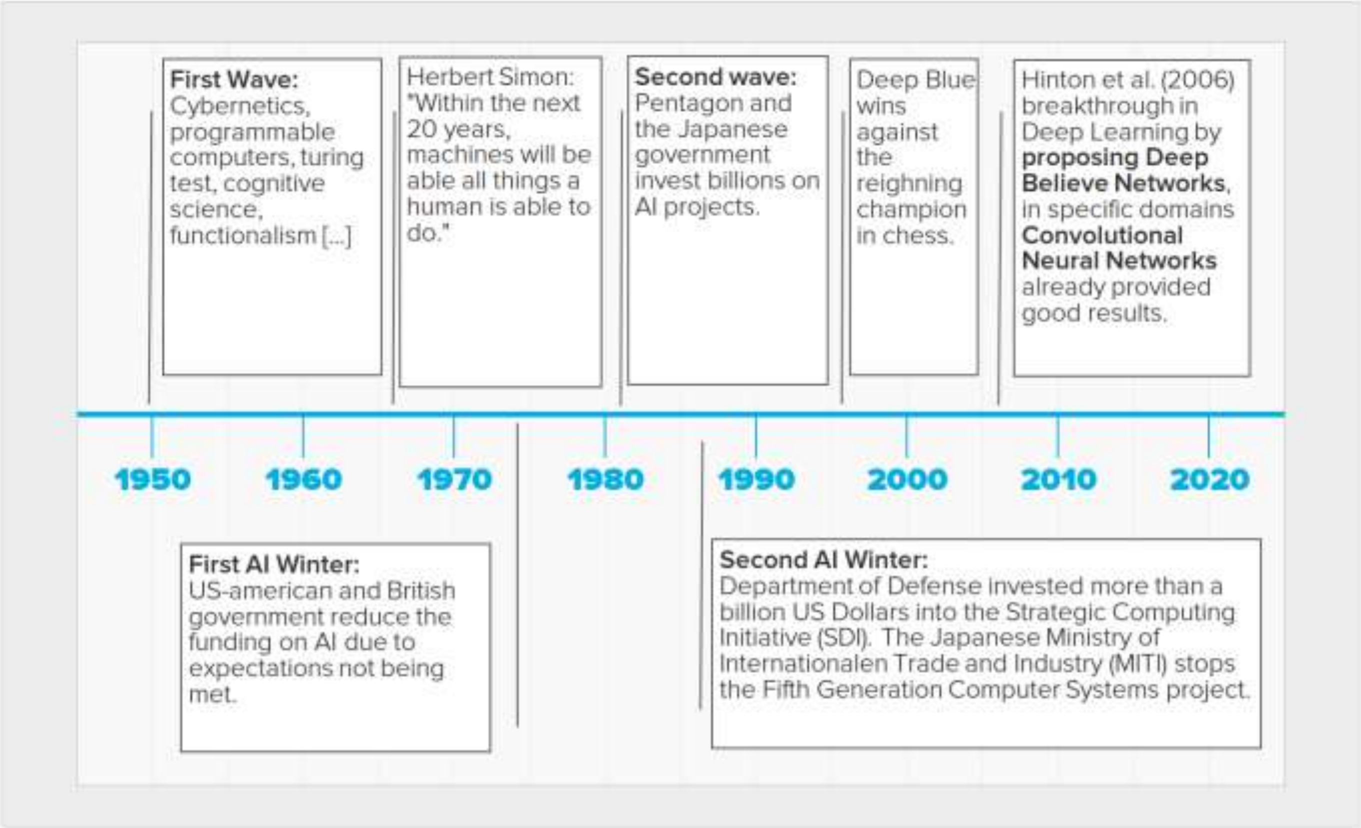
- The data itself might be sensitive
  - Processes and technologies to ensure access to data
  - Is the cloud a safe place for my data?
- The results of the analytics might be sensitive
  - Number of ethical issues

DEFINITION ARTIFICIAL INTELLIGENCE

- Term “**Artificial Intelligence**” coined in the 50ies
  - “The science and engineering of making intelligent machines.” – John McCarthy (1956)
    - Dartmouth Summer Research Project on Artificial Intelligence (DSRPAI)
- Learning has been researched even before
  - Hebbian Learning (1949)
    - Strengthen correlating connections



ARTIFICIAL INTELLIGENCE HISTORY



TYPES OF ARTIFICIAL INTELLIGENCE

- Strong AI (full AI, artificial general intelligence)
  - Machines act as if they were able to reason
- **Weak AI** (narrow AI, applied AI)
  - Pattern-based AI, data-driven AI
    - „Capability of machines to imitate intelligent human behaviour"
  - Machines support humans in relatively simple tasks
- Artificial super intelligence

Vision!

Good results, today!

ARTIFICIAL INTELLIGENCE

Top-down approach

- Study cognition using symbols
- Symbolists

Bottom-up approach

- Study natural intelligence (brain) and try to replicate
- Connectivists

ARTIFICIAL INTELLIGENCE

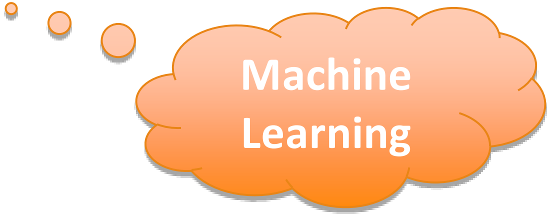
Logic-Based AI

- Rules formulated by human experts
- Knowledge represented in a formalised way
- Decisions can be explained
- Allows to automate processes
    - E.g. Workflow processes

ARTIFICIAL INTELLIGENCE

Pattern-Based AI (Data-Driven AI)

- Machine identifies patterns within data
- Based on mathematical/statistical models
  - And assumptions
- Decisions are often hard to explain
- Also profits from expert knowledge (hybrid models)



CURRENT STATE-OF-THE-ART IN AI

- If a typical person can do a mental task with **less than one second** of thought, we can probably automate it using AI either now or in the near future.
  - Andrew Ng

Decision making support

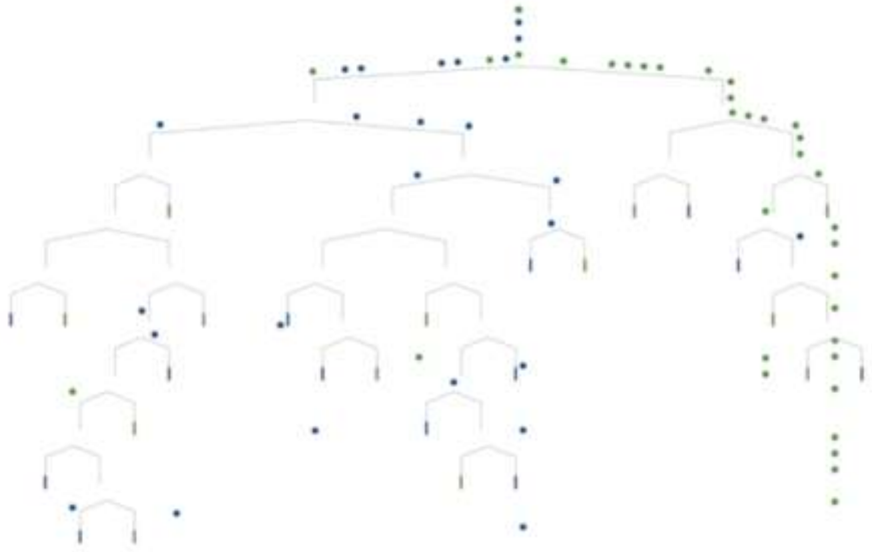
CURRENT STATE-OF-THE-ART IN AI

- For most of the tasks the **human is still superior**
  - Only for some tasks the machine currently reaches super-human performance
- Be **suspicious** if claimed
  - … works better than humans
  - … achieves 100% performance

Testing Procedure?

DEFINITION - MACHINE LEARNING

- **Machine learning**
  - "Machine programs itself" – Arthur Samuel, 1959
    - No explicit instructions
  - Typically, making use of (large) amounts of data



https://www.techemergence.com/what-is-machine-learning/

"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, 1997

CHALLENGES OF MACHINE LEARNING

- Requires large amounts of data
  - Typically clean, unambiguous data
- Interpretable models
  - Understand why the machine made a decision
- Skills
  - Select appropriate approaches
  - Put results in context
- Limited generalisation

Greedy!

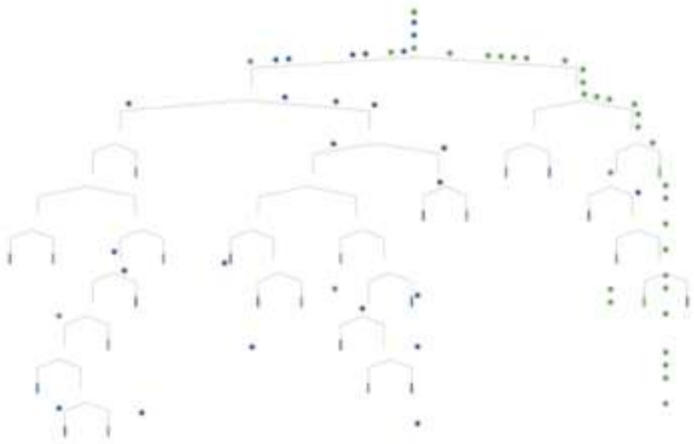
Black Boxes!

Scarce!

"Machine learning can't get something from nothing…what it does is get more from less." – Pedro Domingo.

TYPES OF MACHINE LEARNING

- Supervised learning
  - Response for a given input
    - A → B
- Unsupervised learning
  - Find pattern in data
- Reinforcement learning
  - Learn via interaction with environment

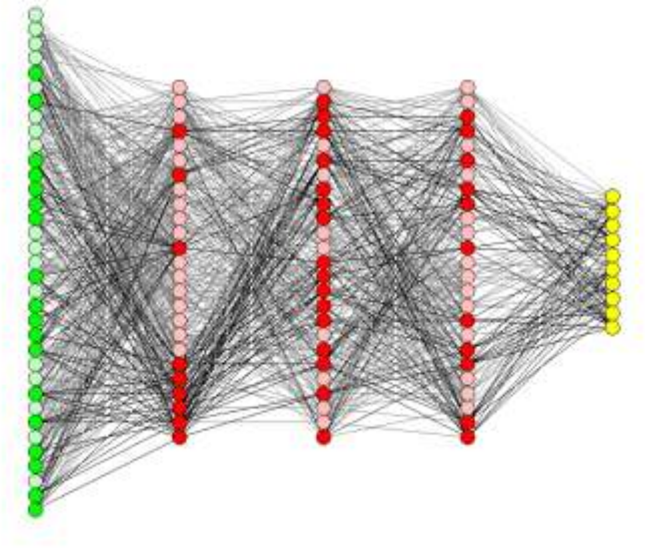


Source: https://www.techemergence.com/what-is-machine-learning/

“We are drowning in information and starving for knowledge” -- John Naisbitt.

DEFINITION - MACHINE LEARNING

- **Deep learning**
  - Synonymous with AI
    - Typically neural networks
  - Expensive to compute
    - Amount of data
    - Amount of computational resources
  - Impressive results



Source: https://www.techemergence.com/what-is-machine-learning/

Example: BigGAN

Copyright: Andrew Brock

CHALLENGES OF MACHINE LEARNING / AI

- Requires **large amounts of data**
  - Typically clean, unambiguous data

Greedy

- **Interpretable models**
  - Understand why the machine made a decision

Black-Boxes

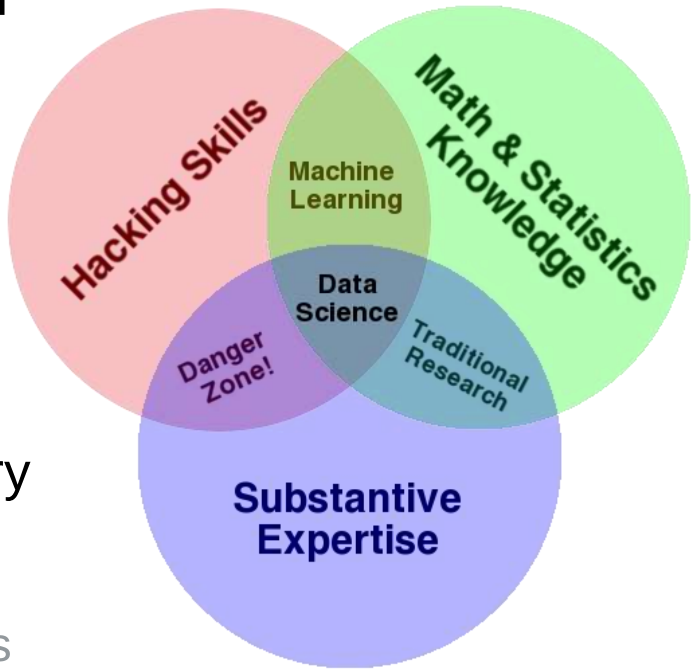
- **Skills**
  - Select appropriate approaches
  - Put results in context

Scarce

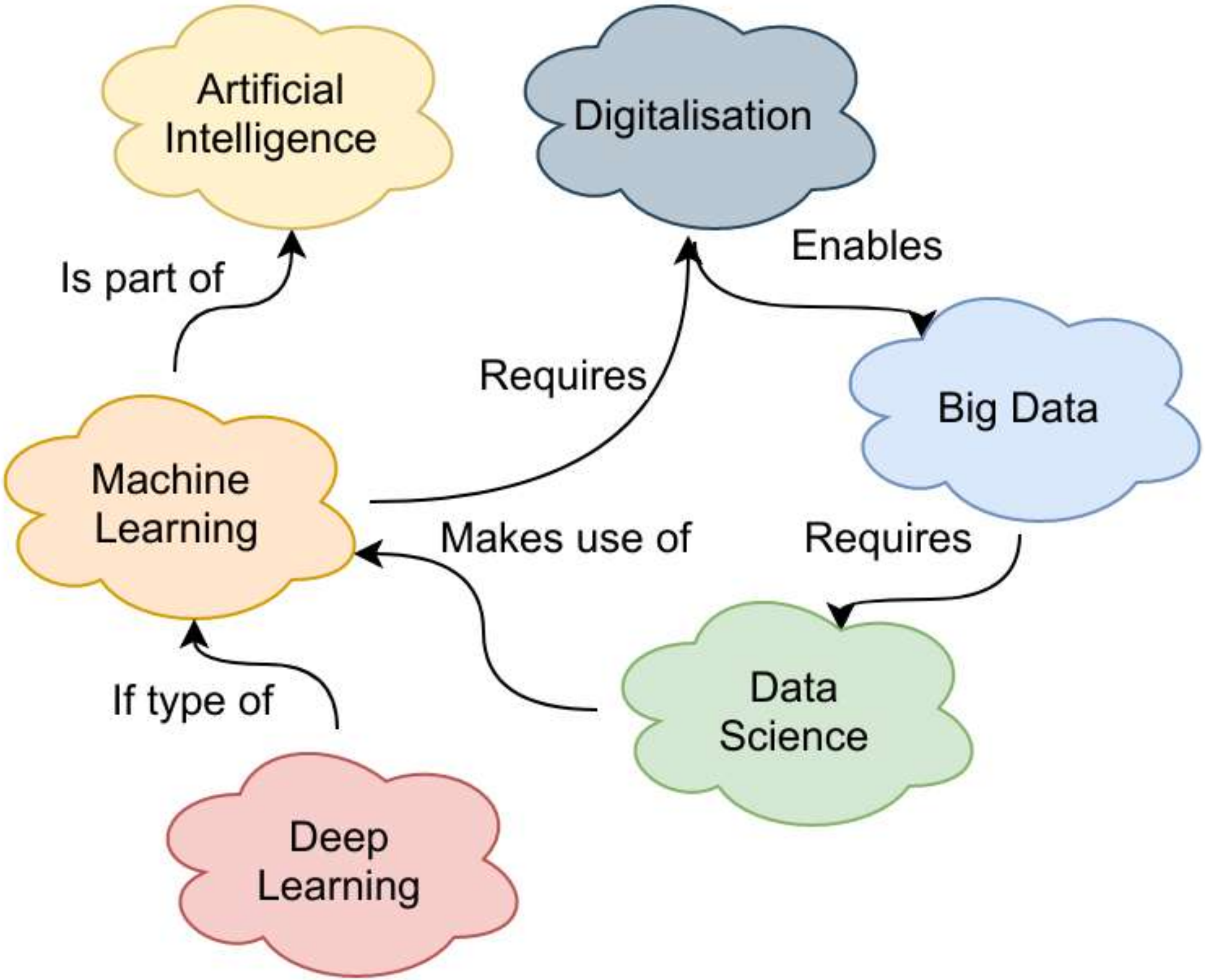
- **Limited generalisation**
  - Works in the lab, not in practice

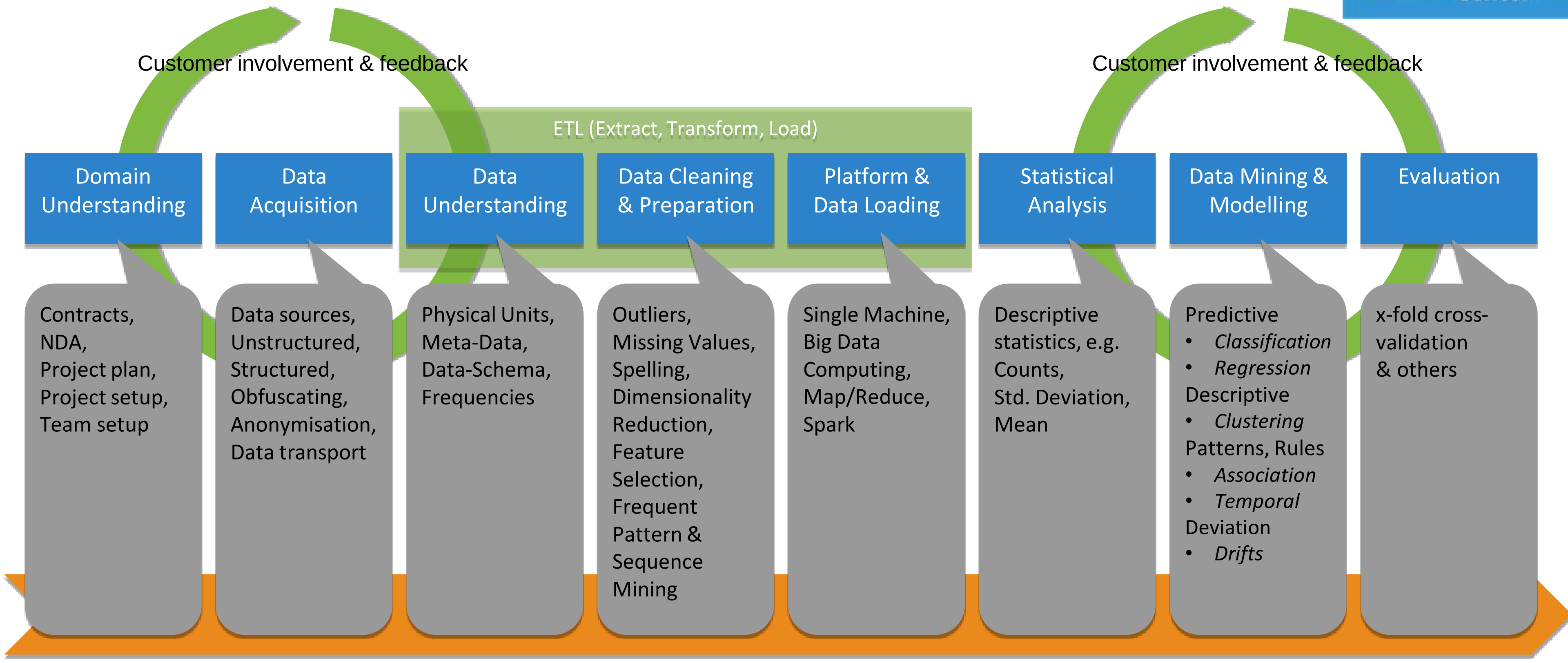
DEFINITION - DATA SCIENCE

- Term “**Data Science**” first appeared
  - Book “Concise Survey of Computer Methods”
    – Peter Naur (1974)
- The **new generation of statistics**
  - Consolidation of several interdisciplinary fields
    - Characterised by a combination of skills



http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram


Artificial Intelligence
Digitalisation
Is part of
Enables
Requires
Big Data
Machine Learning
Makes use of
Requires
If type of
Data Science
Deep Learning

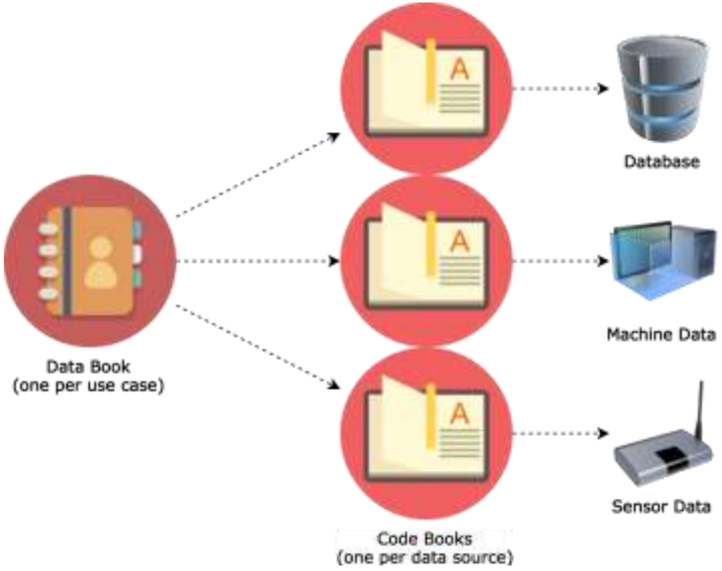

Customer involvement & feedback
Customer involvement & feedback
ETL (Extract, Transform, Load)
Domain Understanding
Data Acquisition
Data Understanding
Data Cleaning & Preparation
Platform & Data Loading
Statistical Analysis
Data Mining & Modelling
Evaluation
Contracts, NDA, Project plan, Project setup, Team setup
Data sources, Unstructured, Structured, Obfuscating, Anonymisation, Data transport
Physical Units, Meta-Data, Data-Schema, Frequencies
Outliers, Missing Values, Spelling, Dimensionality Reduction, Feature Selection, Frequent Pattern & Sequence Mining
Single Machine, Big Data Computing, Map/Reduce, Spark
Descriptive statistics, e.g. Counts, Std. Deviation, Mean
Predictive
• Classification
• Regression
Descriptive
• Clustering
Patterns, Rules
• Association
• Temporal
Deviation
• Drifts
x-fold cross-validation & others

DATA SCIENCE PROJECTS

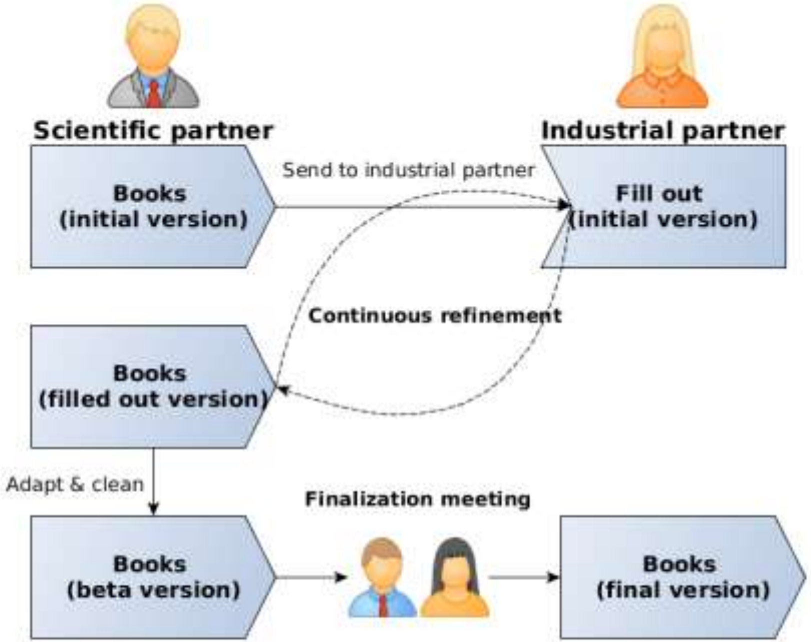
- Types of data science projects
  - **Hypothesis-driven**
    - E.g. Is there a quality impairment, if parameter X is changed?
    - E.g. Can the quality be approximated by process measurements
  - **Data-driven**
    - What insights can be generated from the data?
    - Do the data contain critical changes?
  - **Simulation-driven**
    - Can Machine Learning being utilised to simulate (and then optimise) a process?

TIDY DATA FOR DATA SCIENCE

- **Data quality** is essential for successful project
  - Often optimistic starting point
- Apply techniques to **document the available data** sources
  - Combined with an iterative process
  - Final result is tidy data set



CODE & DATA BOOK PROCESS



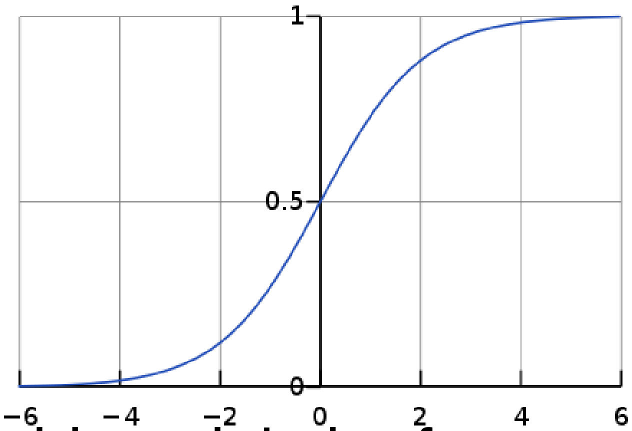
LOGISTIC REGRESSION

Regression for
binary dependent
variables

BINARY CLASSIFICATION EXAMPLE

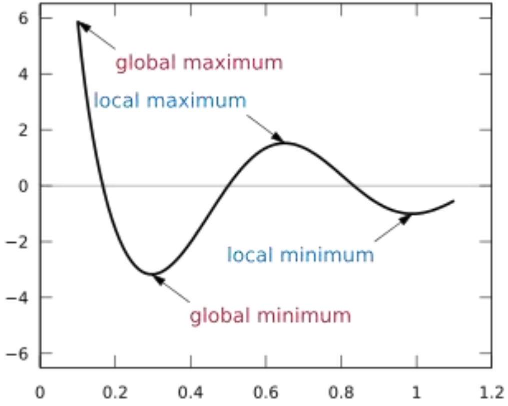
- Example dataset consisting of
  - 8 instances (N), 2 features (M) and 2 classes

LOGISTIC REGRESSION BASICS



- General intuition
  - Separate two classes (C_1, C_2)
  - $p(\mathcal{C}_1|\phi) = y(\phi) = \sigma\left(\mathbf{w}^{\mathrm{T}}\phi\right)$
  - With σ(·) being the logistic sigmoid, and ϕ the feature vector
- Task
  - Given the data, find the most probable **w**

LOGISTIC REGRESSION SOLUTION #1



- Solution idea
  - Define a loss
    - i.e., how good/bad is the solution
  - Find the weights with minimal loss
- General optimisation idea
  - Find the minimum via the derivative of the function
  - Find the point where it is zero

LOGISTIC REGRESSION SOLUTION #1

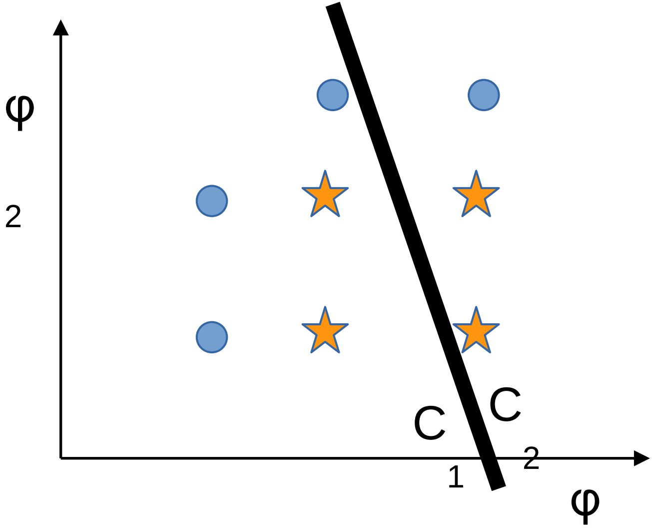
- For logistic regression with *N* instances
  - The likelihood function: $p(\mathbf{t}|\mathbf{w}) = \prod_{n=1}^{N} y_n^{t_n} \{1 - y_n\}^{1-t_n}$
    - With t_n being the true class {0, 1}, and y_n the output [0, 1]
  - Take the negative logarithm
  - $$E(\mathbf{w}) = -\ln p(\mathbf{t}|\mathbf{w}) = -\sum_{n=1}^{N} \{t_n \ln y_n + (1 - t_n) \ln(1 - y_n)\}$$
  - Also known as the cross-entropy error function

LOGISTIC REGRESSION SOLUTION #1

- Find derivative
  - With y_n being the output of the logistic sigmoid σ
    - The derivative is: σ(1-σ)
  - Thus the gradient of the error function is
  - $$\nabla E(\mathbf{w}) = \sum_{n=1}^{N} (y_n - t_n)\phi_n$$

LOGISTIC REGRESSION SOLUTION #2

- Alternative idea
  - Start with a random initialisation and iteratively improve

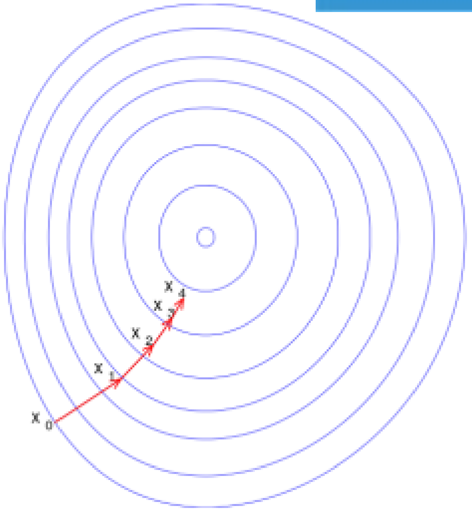


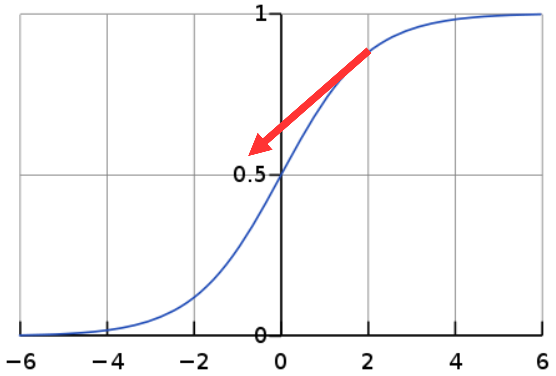
LOGISTIC REGRESSION SOLUTION #2

- Given the initial weights
  - Iteratively refine the weights
  - To reduce a target objective
    - Loss function
  - Open question
    - In which direction and by how much to change the weights?

LOGISTIC REGRESSION SOLUTION #2

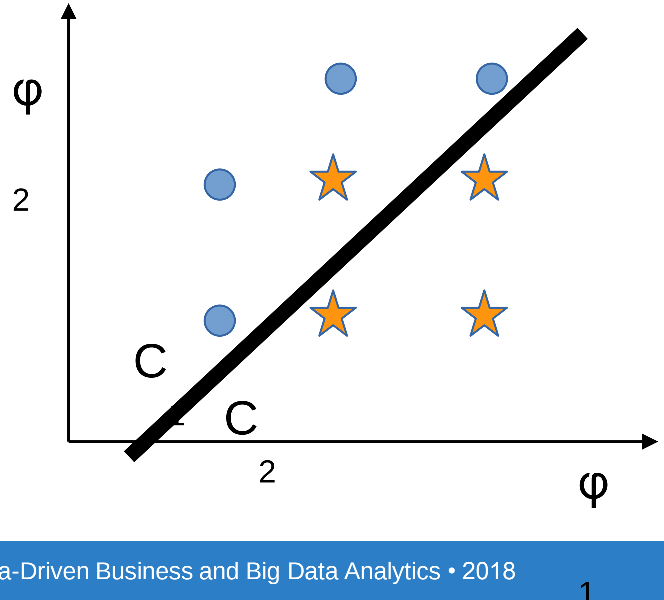
- General idea for iterative optimisation:
  - For each data point
  - Take the steepest gradient to minimise loss
    - Move according to the gradient (derivative at the current point)
  - Only take small steps
    - Learning rate η





LOGISTIC REGRESSION SOLUTION #2

- Stochastic gradient descent
  - Locally optimal solution to separate C_1 and C_2



LOGISTIC REGRESSION GOALS

- Applications for logistic regression
  - Model and estimate the probabilities
    - e.g. of an event occurring
  - Prediction of an binary outcome
    - based on continuous and binary variables
  - Binary classification
    - by estimating the probability of class membership

LOGISTIC REGRESSION EXAMPLE

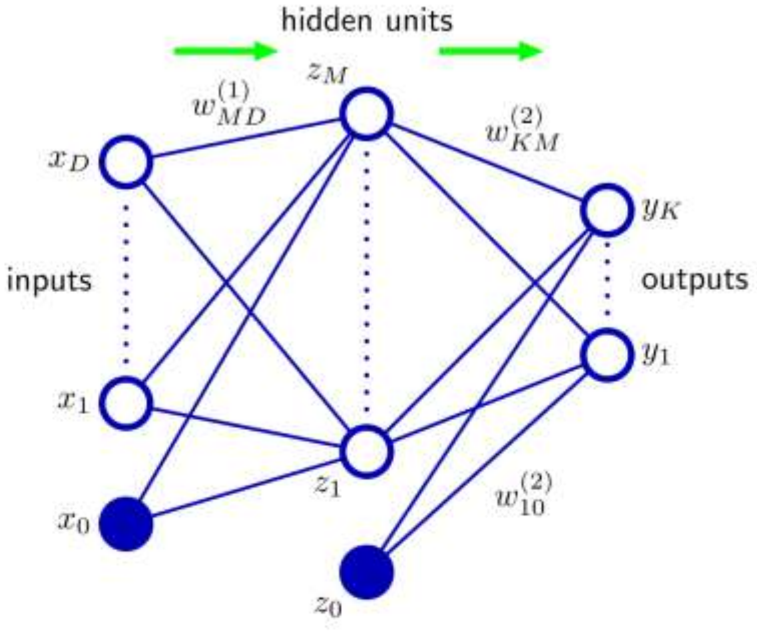
- Demand forecast
  - Demand from certain customers will raise or fall?
    - Given historical training data
      - Data sources: OECD, Eurostat, News, Web, Social Media
    - Compute a model to predict
      - … for each customer raise/fall
    - Additional feature selection is necessary
      - … based on the input of domain experts

ARTIFICIAL NEURAL NETWORKS

ARTIFICIAL (DEEP) NEURAL NETWORKS

- Artificial neural networks (ANN)
  - Layers of artificial neurons
    - e.g perceptrons

- Deep learning
  - Learning of multiple layers
    - Not limited to ANNs



ARTIFICIAL NEURAL NETWORKS

- Key elements of learning of (deep) neural networks
  - Activation function
    - e.g. logistic function, softmax
  - Cost function
    - e.g. cross entropy
  - Optimisation/learning algorithm
    - e.g. stochastic gradient descent
  - Update rule for hidden layers
    - e.g. backpropagation

Forward pass

Backward pass

BACKPROPAGATION

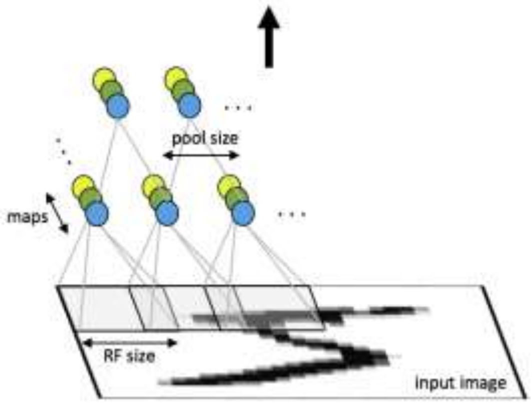
- Start with the output layer
  - Loss can be computed at the output layer
    - i.e. difference between expected and computed output
    - Compute the gradient from the loss
- Step backwards in direction of input layer
  - Need to update the inner layers
    - Chain of gradients
      - Multiply the local gradient with the outer gradient
      - Local gradient is the gradient at the point of the input

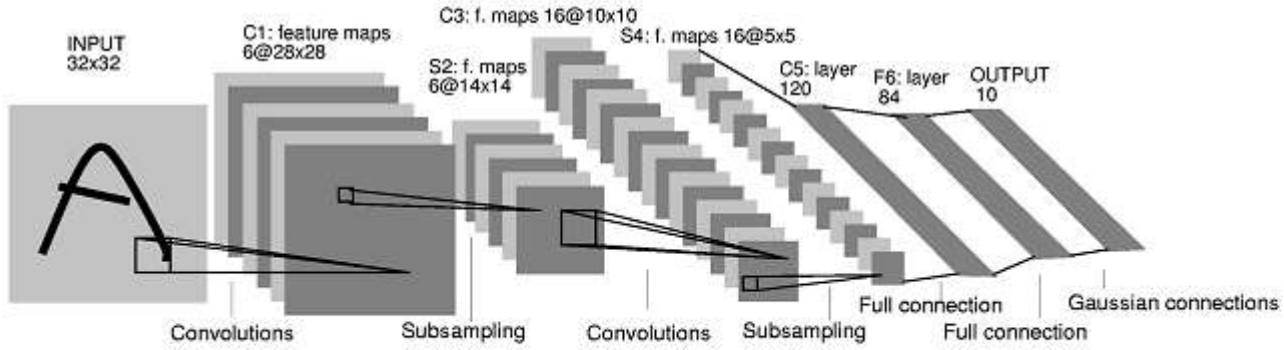
BACKPROPAGATION

- Intuition
  - The loss is how much we have to change our output
  - The (output) gradient tells us in which direction to move
    - … to achieve the reduction in loss
  - The chain rule computes the contribution of each input
    - … how much has each input to change
  - … so the gradients flows through the network

CONVOLUTIONAL NEURAL NETWORKS

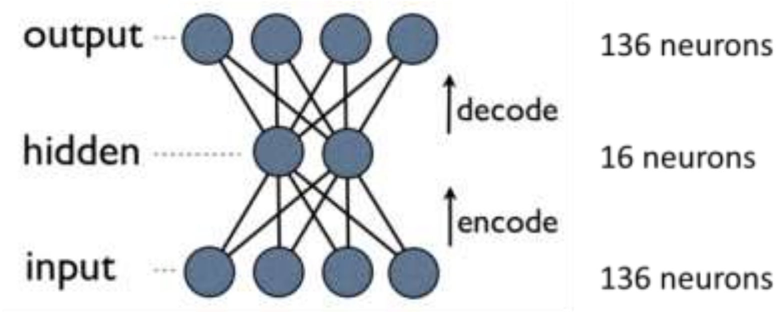
- Special type of feed forward network
  - To limit the amount of necessary weights
  - Convolutions
    - Instead of taking the full image, only look at a small field (receptive field)





AUTOENCODER

- Special type of feed forward network
  - As many outputs as inputs
    - 1+ hidden layers
    - decoder(encoder(x))
    - compressed representation
  - Networks is trained to reproduce the input
- Can be used for: dimensionality reduction, clustering, outlier detection, denoising of images

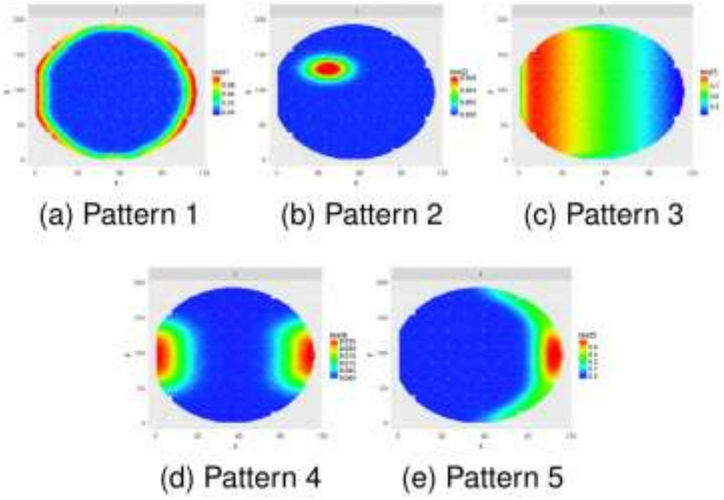


VARIATIONAL AUTOENCODER

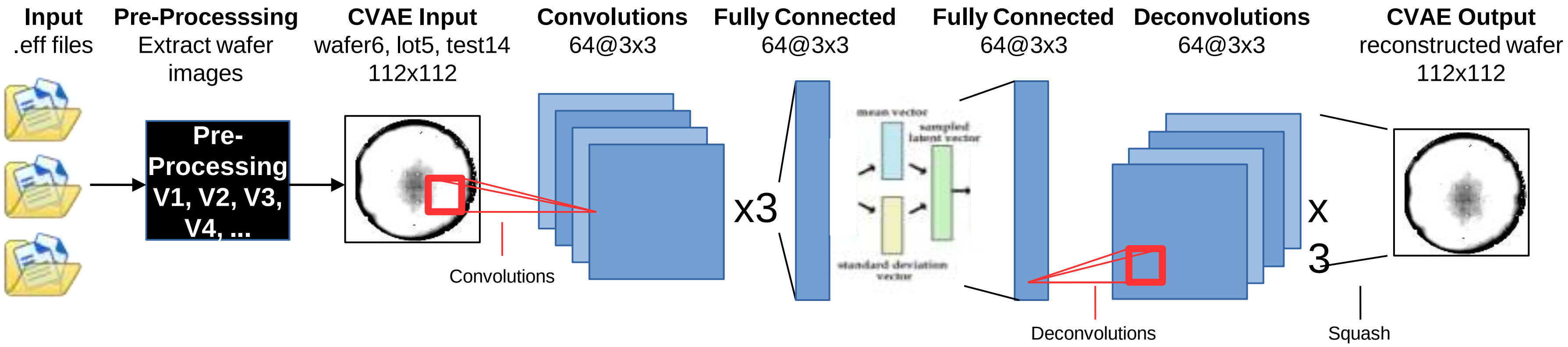
- Special type of auto encoder
  - The networks learns the parameters of a distribution, e.g. μ, σ
    - … instead of single values
    - Generative model
      - e.g. can produce realistically looking images
    - Backpropagation through a distribution not possible
- Training requires the so-called reparameterisation trick
  - Treat μ, σ as deterministic variables with an auxiliary noise (regularisation), i.e. should be as close to $\mathcal{N}(0,1)$ as possible

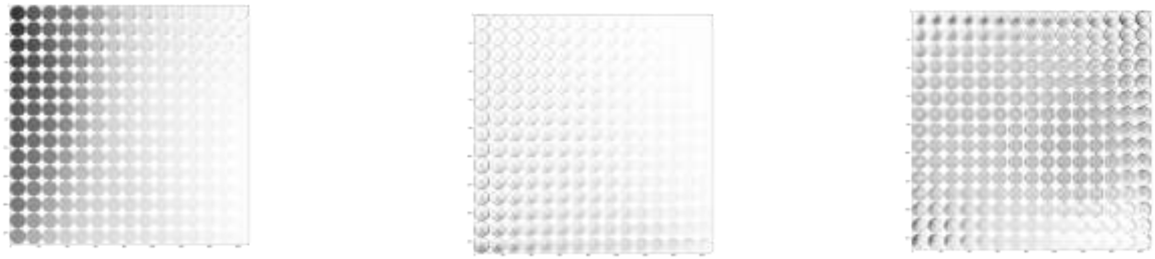
DEEP LEARNING APPLICATION EXAMPLE

- Goal: find pattern in analog wafer end test data
  - Algorithmic solution
    - Convolutional Variational Auto Encoder (CVAE)
  - Trained with real-world examples
    - Pre-processing needed
    - Finds groups and can be used to generate new „images“



DEEP LEARNING APPLICATION EXAMPLE





CONCLUSIONS

ARTIFICIAL INTELLIGENCE IN REAL-LIFE

- **Data acquisition**
  - High quality data (complete, up-to-date, documented …)
    - Labelled data preferred
  - Often requires dedicated storage techniques
- **Data analysis & modelling**
  - Cooperation b/w ML expert and domain experts
- **Successful projects**
  - Start with a small, focused topic (e.g. hypothesis)
    - Expectation management, e.g. rare events hardly be modelled
  - Iteratively refine and improve

Data quality check!

Skills & trust!

Success stories!

AUSTRIA'S LEADING RESEARCH CENTER
FOR DATA-DRIVEN BUSINESS AND BIG DATA ANALYTICS

Dr. Roman Kern
Area Head
rkern@know-center.at

Know-Center GmbH
Research Center for Data-Driven Business and Big Data Analytics
Inffeldgasse 13/6
8010 Graz, Austria

Firmenbuchgericht Graz
FN 199 685 f
UID: ATU 50367703

gefördert durch das Programm COMET (Competence Centers for Excellent Technologies), wir danken unseren Fördergebern: