Deep Learning

Introductory Lecture

Tuesday 23rd April

Dr. Nicholas Cummins

Welcome!

W

W

Deep Learning

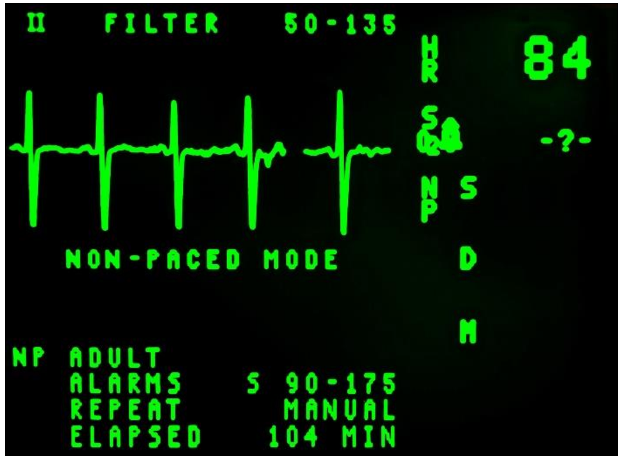
W

W

W

Welcome

Universität Augsburg University



Images Sourced:
https://pixabay.com

Welcome!

W

Who are we?

Deep Learning

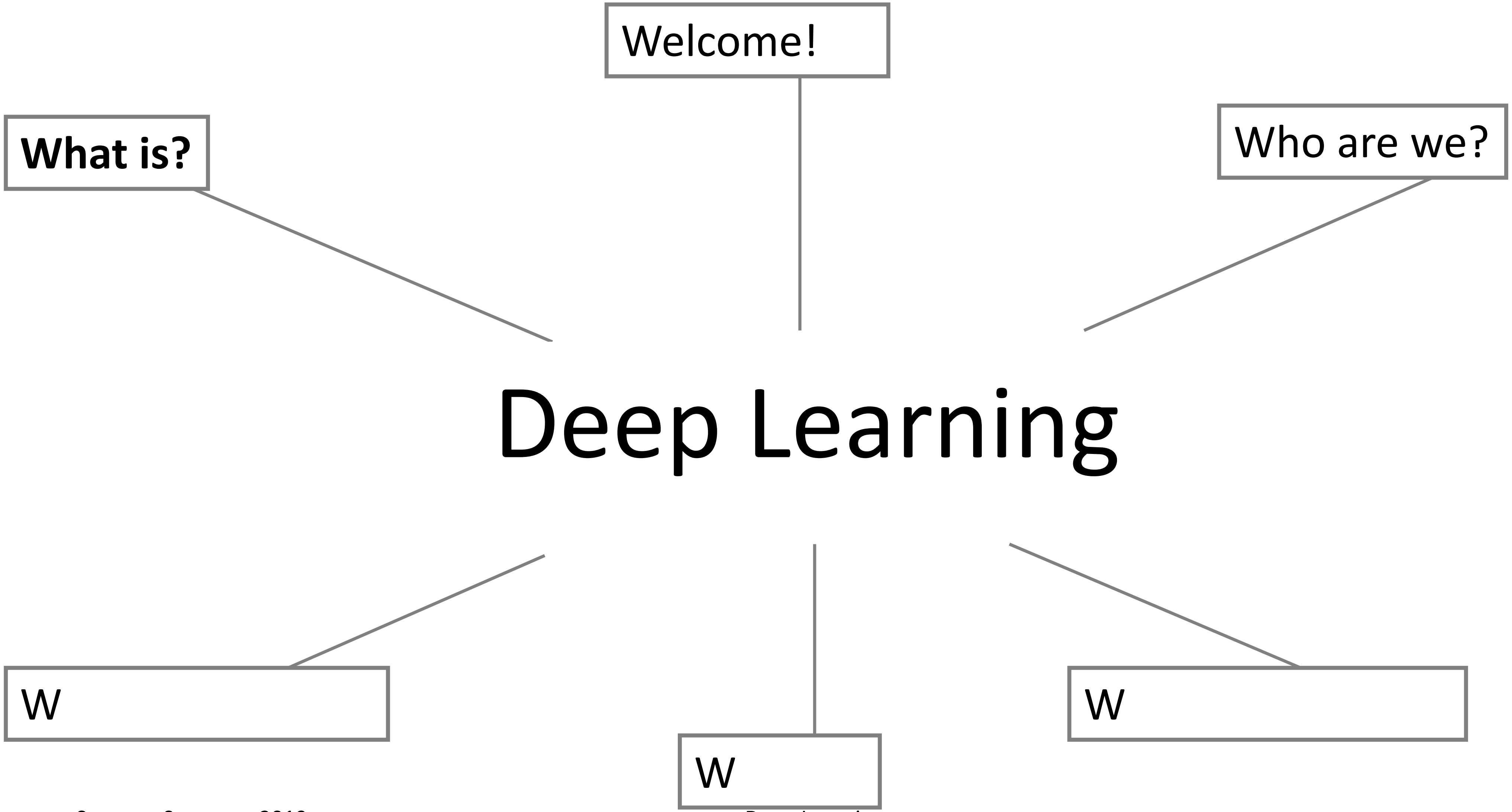
W

W

W

Contacts

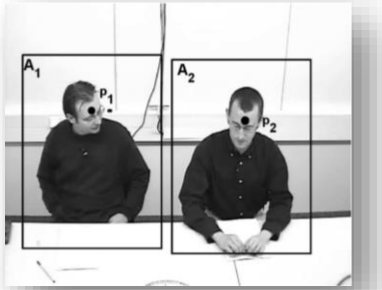
- **Prof. Björn Schuller**
  - bjoern.schuller@informatik.uni-augsburg.de
- **Dr. Nicholas Cummins**
  - nicholas.cummins@informatik.uni-augsburg.de
- **Shahin Amiriparian**
  - shahin.amiriparian@informatik.uni-augsburg.de


Welcome!
What is?
Who are we?
Deep Learning
W
W
W

History of Deep Learning

- 1943 McCulloch+Pitts: 1st formal neuron model
  → Computation of arbitrary arithmetic / logic functions
- 1949 Hebb: synaptic connection between 2 neurons is enhanced by frequent activation (Hebb rule)
- 1958 Rosenblatt: 1st Neurocomputer-Perceptron (mechanical)
- 1960 Widrow+Hoff: *Least mean square* Algorithm
  + description of *adaptive linear neuron* (*Adaline)*
- 1965 Vakhnenko+Lapa: 1st „deep Net“
- 1969 Minsky+Papert: Limits of multilayer perceptrons → stagnation
- 1979 Fukushima: convolutional nets (CNNs)
- 1986 Rumelhart+Hinton+Williams: Backpropagation in n-layer models, Dechter: „Deep Learning“ (Hinton: uses it 2006)
- 1997 Hochreiter+Schmidhuber: LSTM Net
- 2010 Hinton / Ruslan: Stacking, hierarchical feature learning
- ~2012 GPUs, Big Data, Rectified Linear Units, Dropout, …

Prof. Schuller's Deep Leaning History



- "*A Combined LSTM-RNN-HMM Approach to Meeting Event Segmentation and Recognition*", ICASSP, **2006**.
- "*Abandoning Emotion Classes - Towards Continuous Emotion Recognition with Modelling of Long-Range Dependencies*", Interspeech, **2008**.
- "*Deep neural networks for acoustic emotion recognition: Raising the benchmarks*", ICASSP, **2011**.
- "*Introducing CURRENNT: the Munich Open-Source CUDA RecurREnt Neural Network Toolkit*", JMLR, **2015**.
- "Adieu features? End-to-end speech emotion recognition using a deep convolutional recurrent network", ICASSP, **2016**.
- "End-to-end learning for dimensional emotion recognition from physiological signals", ICME, **2017**.
- "End-to-End Multimodal Emotion Recognition using Deep Neural Networks", JSTSP, **2017**.
- "End2You – The Imperial Toolkit for Multimodal Profiling by End-to-End Learning," **2018**.

What is Deep Learning?

Artificial Intelligence:

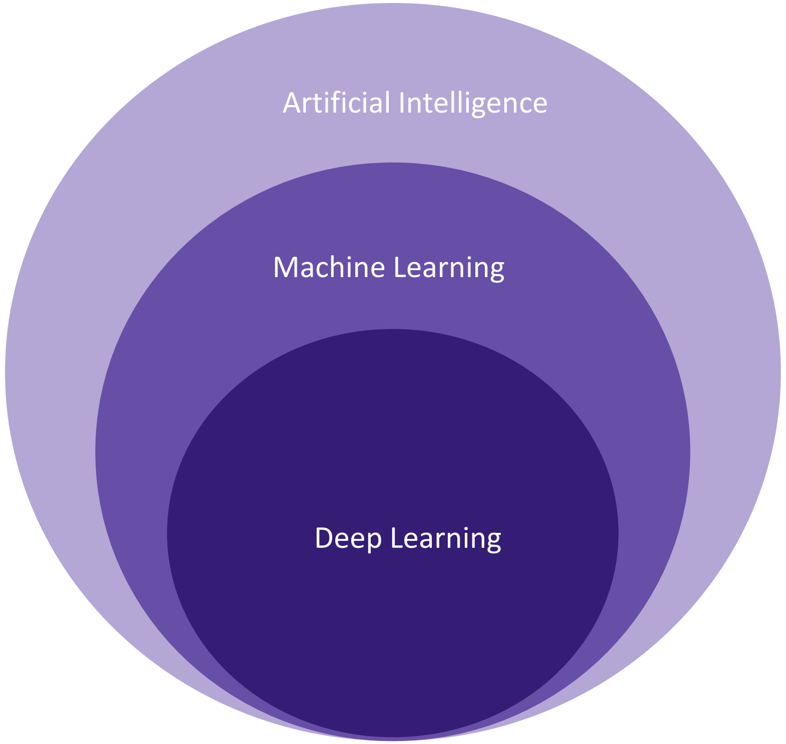
- A broad concept where machines think and act more like humans

Machine Learning:

- An application of AI where machines use date to automatically improve at performing tasks

Deep Learning:

- A machine learning technique that processes data through a multi-layered neural network much like the human brain



AI/ML Universe

Targets:
Classification
Regression
Trees
Uncertain
Missing
Drifting
Multiple
…

Features:
Time Series
Generation
Reduction
Selection
…

Optimisation:
Bayesian
Genetic
Ant Colony
Annealing
…

Algorithms:
Neural Networks
Kernel Machines
Trees
Graphical Models
Ensembles
…

Learning:
Supervised
Transfer
Active
Reinforced
Semi-Superv.
Unsupervised
Deep
Hierachical
Relational
Distributed
Parallel
Hybrid
Signal-based
Discriminative
Inference
…

"Algorithms that can learn from data"

…Achieve brain-like intelligence

Applications:
Info Extraction
Computer Perception
Robotics
Communication
Transportation
Smart Infrastructure
…

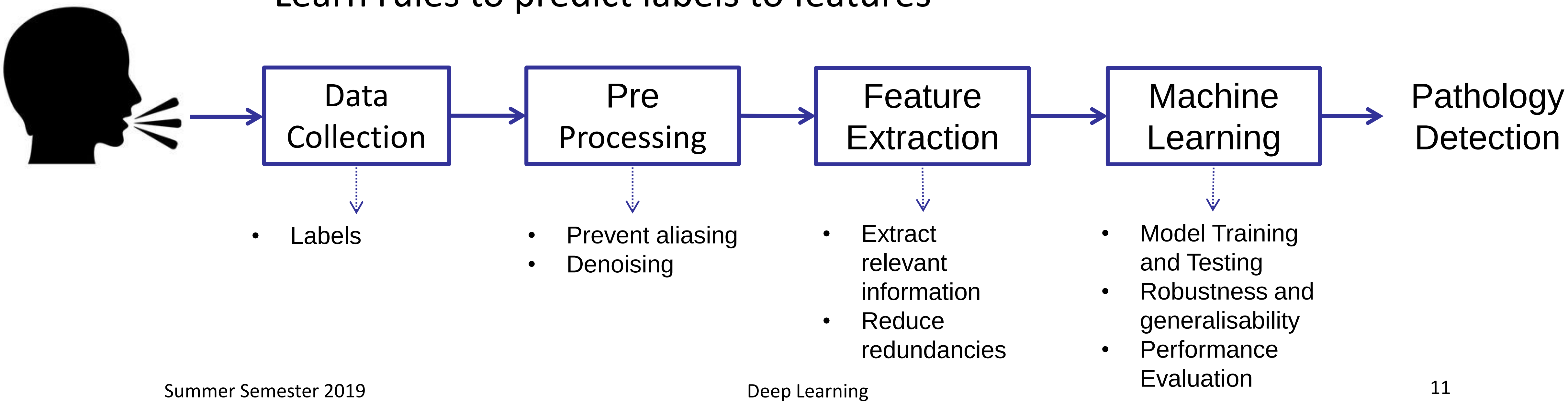
…

Theory:
General
Statistical
…

Embedding:
Efficiency
Confidences
Robustness
Low-Resource
Context
…

Involvement of Computational Intelligence

- **Feature Extraction**
  – Extracts useful information from raw data
- **Machine learning algorithm**
  – Learn rules to predict labels to features



What are features?

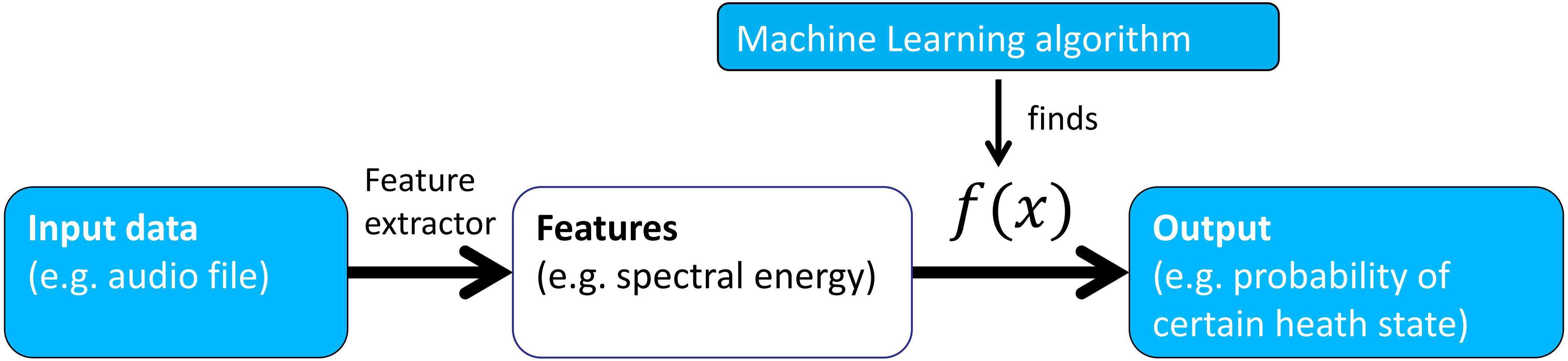
- The representation of the data presented to the machine learning algorithm
- Each feature can be thought of as a single piece of information the algorithm can use when making a decision
- Typically hundreds or thousands of such pieces of information are concatenated together to form a feature vector
- The role of the machine learning algorithm is to identify patterns from a collection of feature vectors

What is machine learning?

- Creation of (robust) models to predict/classify a particular output (y) from a selected independent variables (X – features) from a dataset
  - Primarily concerned with the identification of patterns within (large amounts of) data
  - Machine learning algorithms are used to perform the process of pattern identification via an iterative process
  - Learning phase: the algorithm optimises its parameters with the goal of improving (recognition) performance on a particular task

Typical processing chain

- Feature extractor
  - Extracts useful pieces of information from raw data
- Machine learning algorithm
  - Learn rules to predict labels to features



Goal

- Learn a *robust* predictive function $f(\cdot)$
- A mapping from the feature space $\mathcal{X}$ to the label space $\mathcal{Y}$

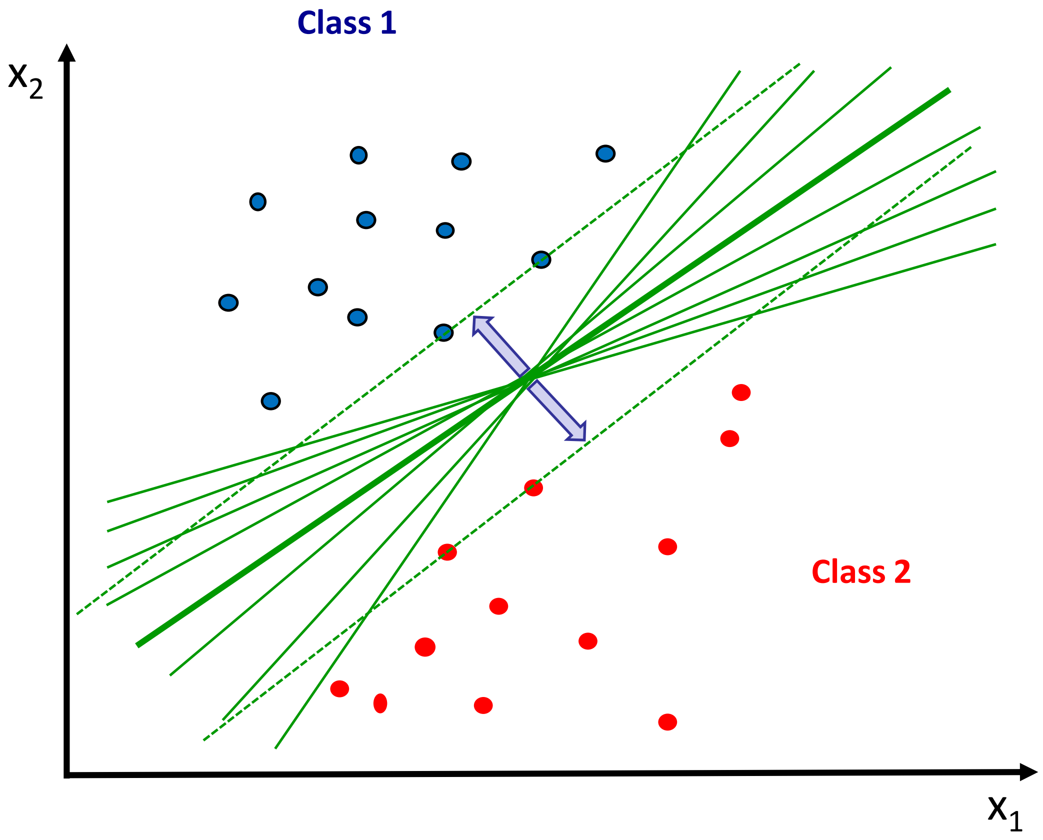
$$\mathcal{X} \xrightarrow{f(\cdot)} \mathcal{Y}$$

- Given a test sample (unknown label), the learnt function maps the test feature vector $\boldsymbol{x}_*$ into a specific label $\boldsymbol{y}_*$

$$\boldsymbol{y}_* = f(\boldsymbol{x}_*)$$

Machine learning algorithms

- Used to perform the process of pattern identification
  - Iterative techniques to find a set of optimal model parameters via the minimisation of a *cost function*
  - A **cost function** is a measure of how incorrect a model is in terms of its ability to estimate the relationship between $\mathcal{X}$ and $\mathcal{Y}$



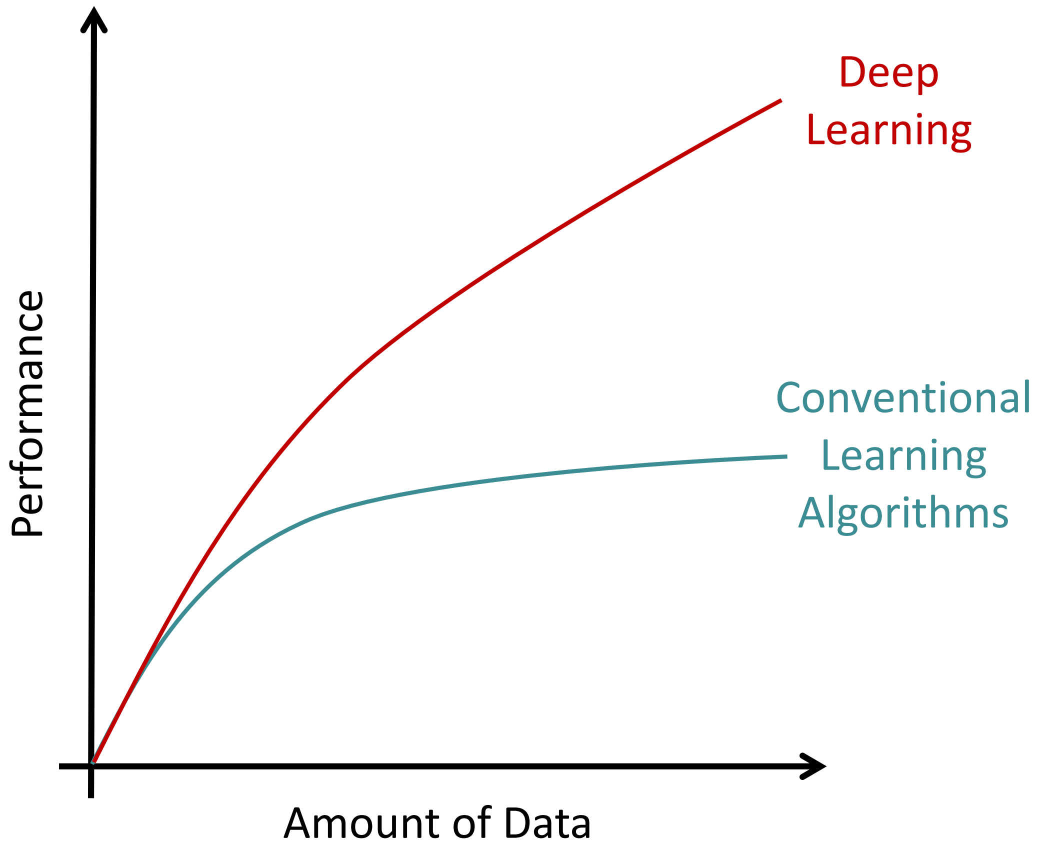
Why use Deep Learning?

Performs well on large datasets

- Conventional algorithms do not scale well to huge amounts of data
  - Especially true in relation to complex problems in image classification, natural language processing, and speech processing

Universality

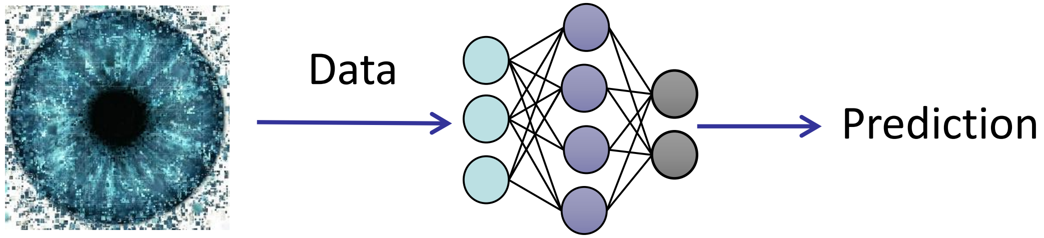
- A network with a single sufficiently large hidden layer (in theory) is adequate for the approximation of most functions



End to End Processing

- Conventional Machine learning relies of features to reduce data complexity & make patterns more visible

- Deep Learning algorithms learn high-level features from data in an incremental manner.
  - Eliminates the need of domain expertise and feature extraction.

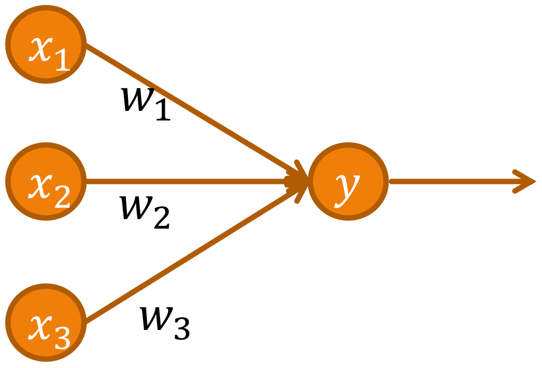


Deep Learning

Images Sourced:
https://pixabay.com

Artificial Neurons

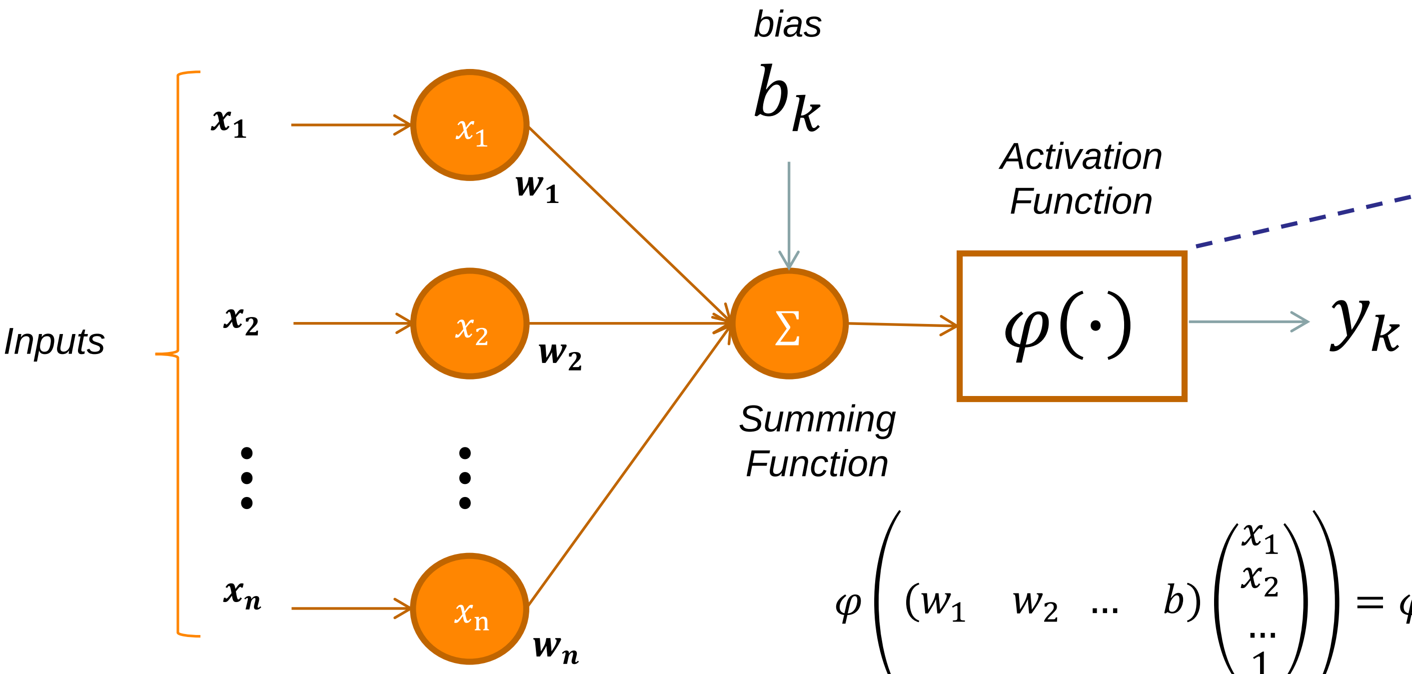
- Building block of neural networks
- Combines different inputs to make a single output

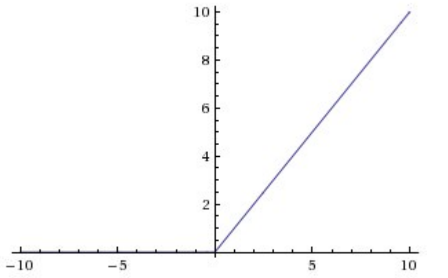


$$y = \begin{pmatrix} w_1 & w_2 & w_3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = w_1 x_1 + w_2 x_2 + w_3 x_3$$

Artificial Neurons

- Nonlinearities
  - Activations Function



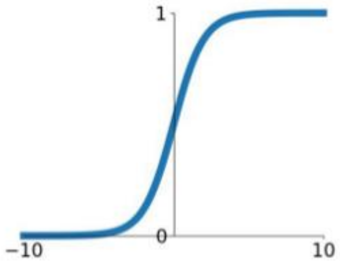


Rectified Linear unit
$f(x) = \max(0, x)$

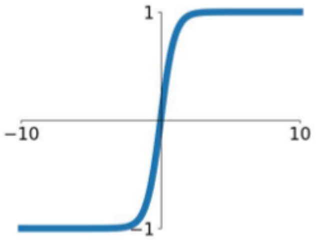
$$\varphi\left((w_1 \quad w_2 \quad \dots \quad b) \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ 1 \end{pmatrix} \right) = \varphi(w_1 x_1 + w_2 x_2 + \dots + b\,) = y_k$$

Activation Functions

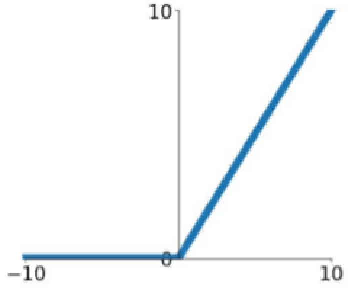
- Uses the weighted input value to determine the level of output activation
  - Introduces nonlinearities into network
- Typical activation functions include
  - Identity → $f(x) = x$
  - Logistic → $f(x) = 1/(1 + \exp(-x))$
  - Tanh → $f(x) = \tanh x$
  - Rectified Linear unit → $f(x) = \max(0, x)$
  - Sigmoid → $f(x) = 1/(1 + \exp^{x})$



Sigmoid



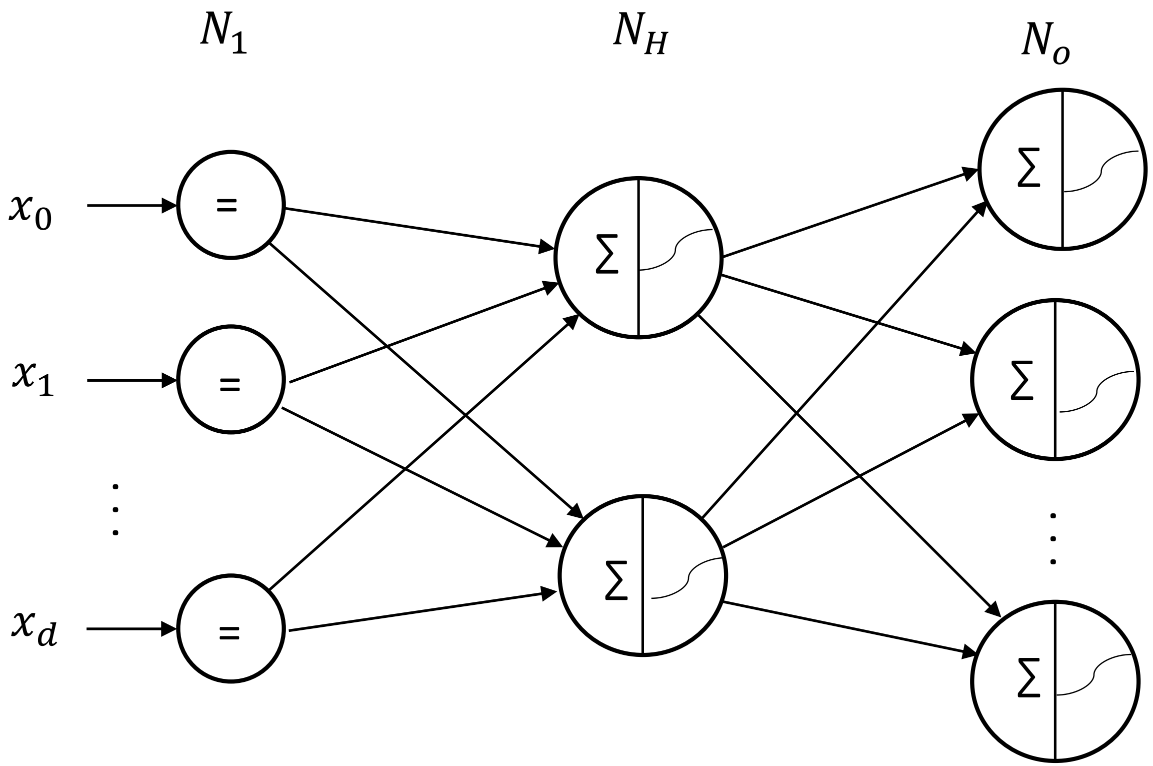
Tanh



ReLU

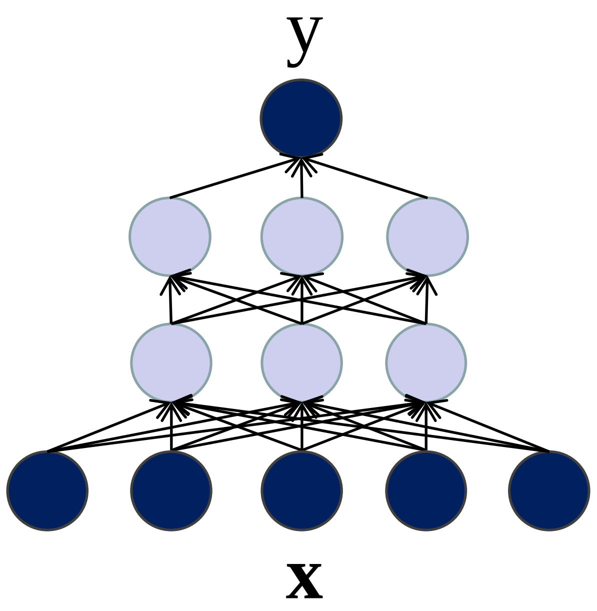
Artificial Neural Networks

- Multiple layers of (non-linear) processing



Definition

- Minimum ca. 4 layers
- AlexNet (2012): 8 (learning) layers
  - 60 million parameters
- Inception-v3 (2015): 47 (learning) layers
  - 25 million parameters



Expectations

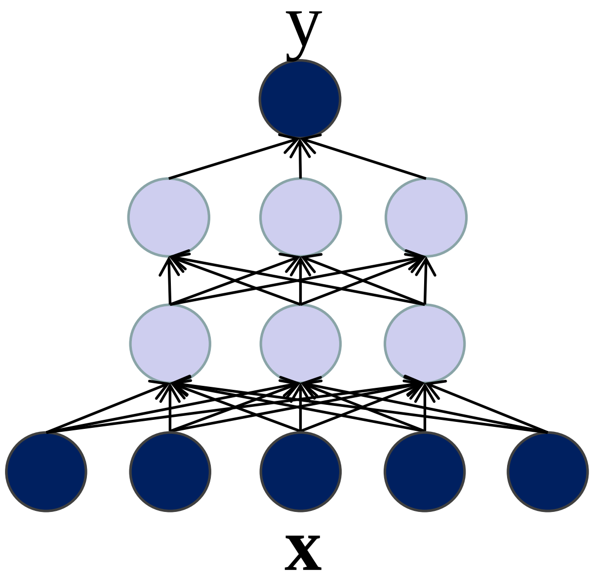
- Learning of highly complex functions enabled
- Applicable on 'hard' tasks
- Reuse of parts of the net for other tasks
  - Transfer-Learning

Universität Augsburg University

Deep Neural Network

- DNN = MLP…
- Training makes the difference:
  - Build them layer-wise
  - Less uninitialized parameters
- Use “raw” features as input
  - Net learns own higher-level features
- Use more training data
- On-line learning (gradient descent)
- Use ReLUs, drop-out learning, …

$$\mathbf{y} = f_{DNN}(\mathbf{x}) = o\left(\mathbf{W}^K \sigma\left(\mathbf{W}^{K-1}\sigma(\dots \mathbf{W}^1\mathbf{x})\right)\right)$$

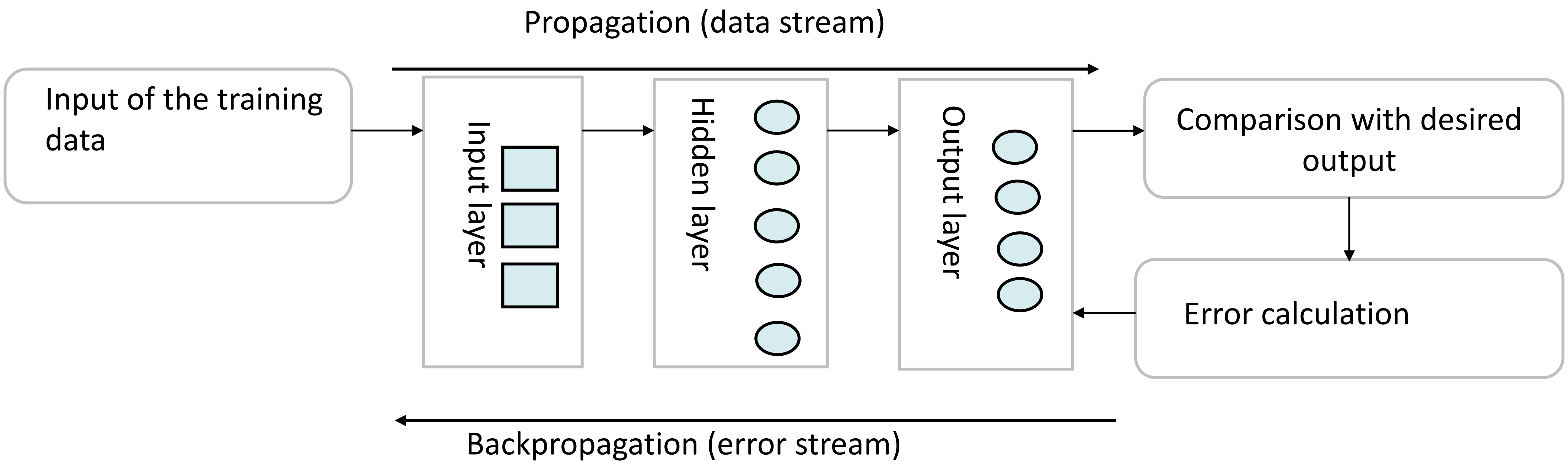


$$E_{Tr}(\mathbf{y}, \mathbf{y}^*) = \sum\nolimits_{\mathbf{x}\in Tr} D(f_{DNN}(\mathbf{x}), \mathbf{y}^*)$$

$$w^{(i+1)} = w^{(i)} - \eta \frac{\partial E_B}{\partial w}(w^{(i)}),\ B \subset Tr$$

Training via Backpropagation

- *Information flow*

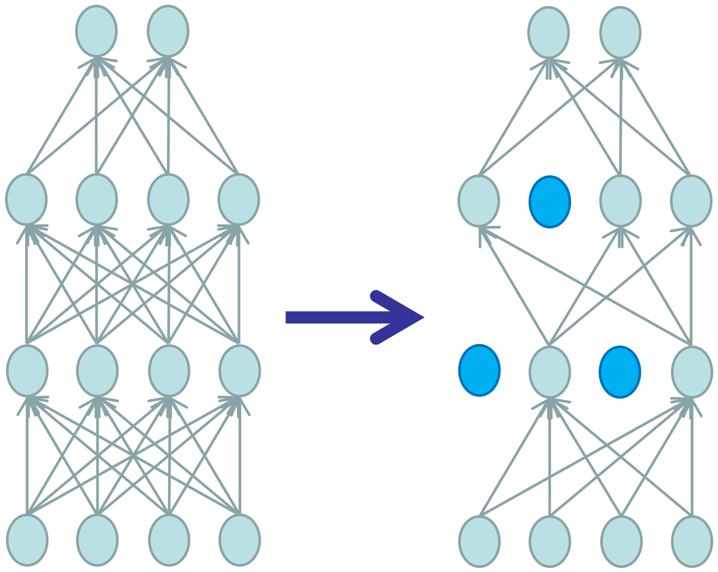


Regularistion

- Nets may become very deep
- Increased risk of overfitting
- Regularisation avoids this

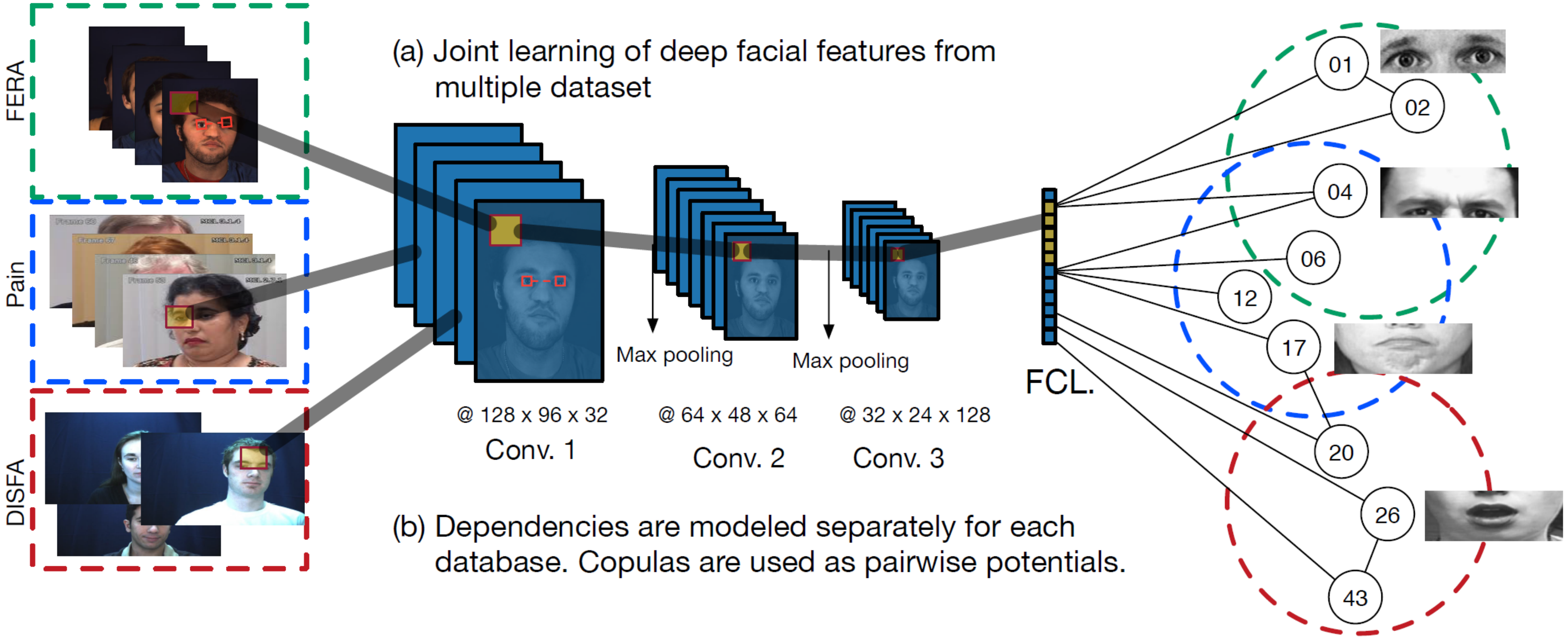
Example: Dropout

- During Training. units are removed randomly
- Aim: higher robustness of learnt parameters
- Disadvantage: slightly slows down training
- Units cannot “rely” on each other’s presence
- Units learn features which are useful by themselves

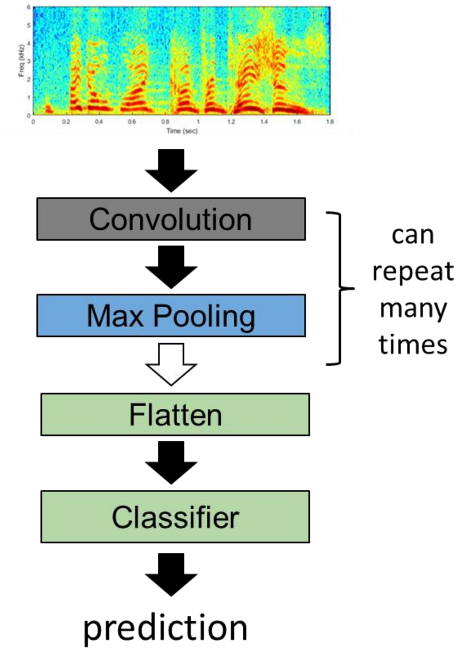


Convolutional Neural Networks

- Inspired by structure of visual processing in the brain
- Use same neurons for repetitive tasks → less parameters
- ‘Kernels’ recognise features in image



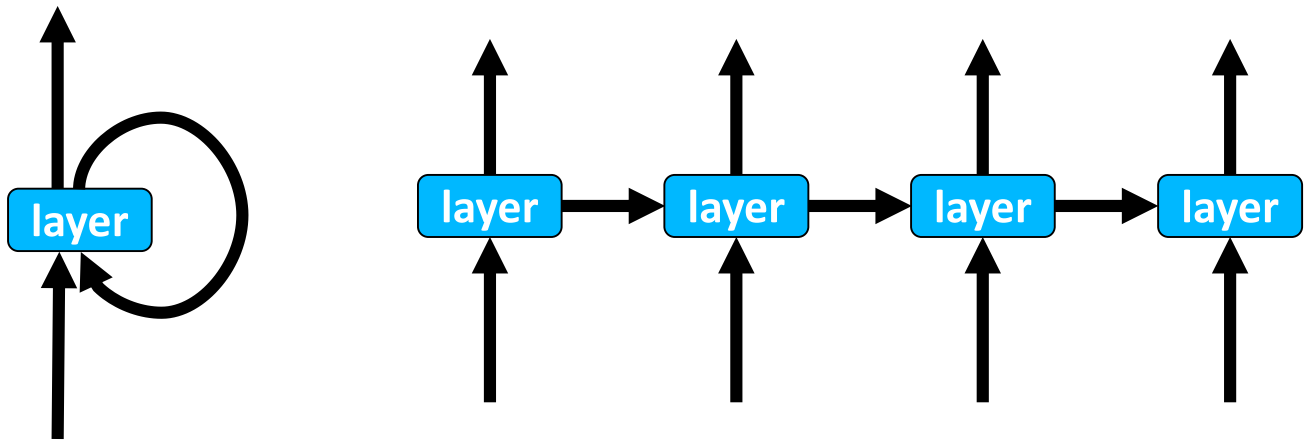
Convolutional Neural Networks



- Learning features directly from data
  - Extraction of task specific features
- **Convolutional layers**
  - Sets of filters used to identify patterns within the signal
  - Network learns weights associated with each filter
- **Max Pooling layers**
  - Keep maximum output within a small neighbourhood of filter output
  - Down sampling operation to reduce information
  - Learnt features invariant to small translations (shifts) in the input

Recurrent Neural Networks

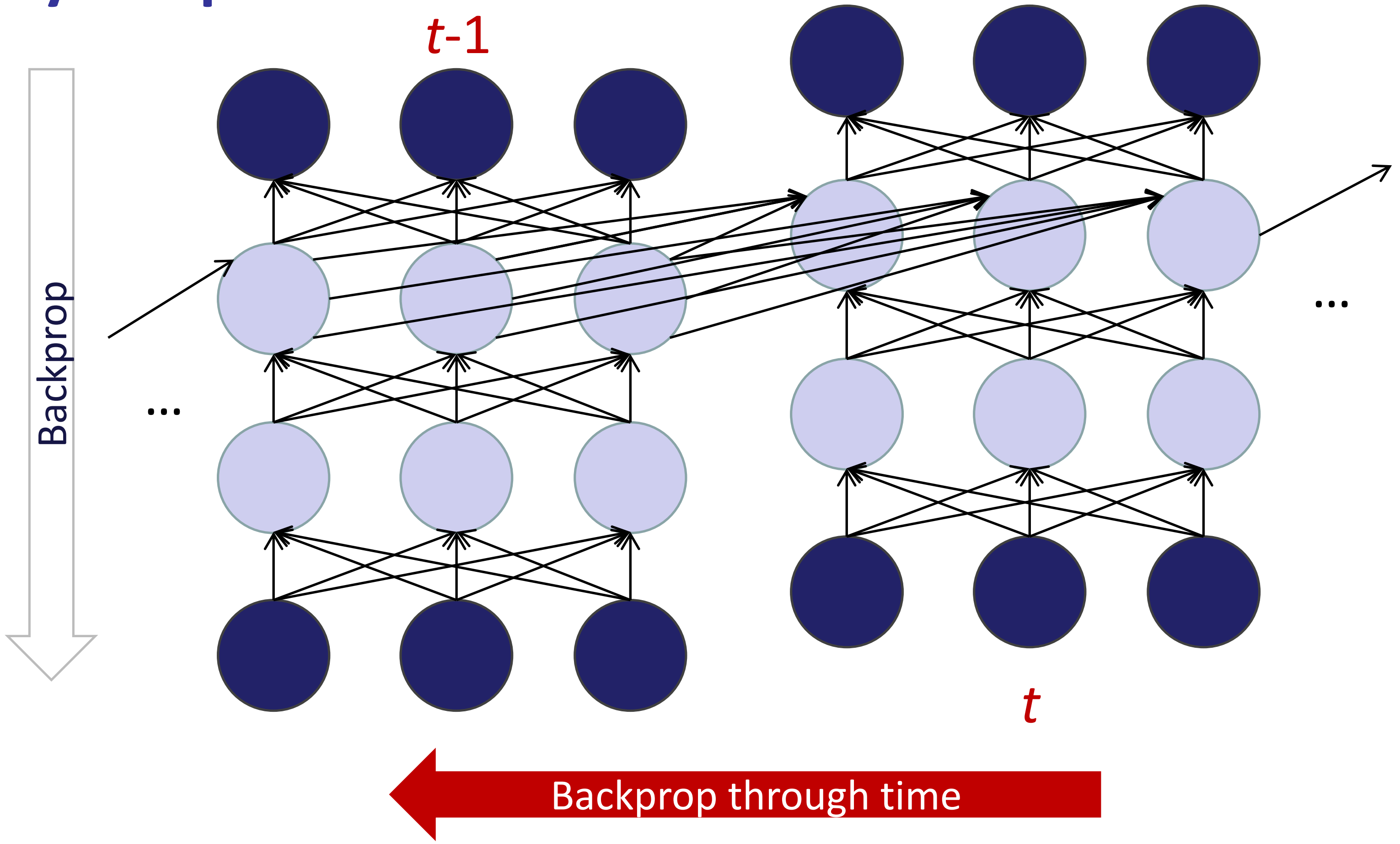
- Neural Network with feedback
- Advantages of doing this
  - Network capable of analysing streams of data
  - Capture temporal / contextual information
  - Good when training samples have interdependencies



Deep Recurrent Neural Networks (Deep RNNs)

$$\mathbf{y}_t = f_{RNN}(\mathbf{x}_0, \dots, \mathbf{x}_t)$$
$$= o\left(\mathbf{W}^o \sigma\left(\mathbf{W}\mathbf{x}_t + \mathbf{R}\,\sigma(\dots\mathbf{W}\mathbf{x}_0 + \mathbf{R}\mathbf{h}_0)\right)\right)$$

'Very Deep' NN



Long Short-Term Memory.

Recurrent Neural Networks (RNN):

- 80s: Gradients by unfolding (BPTT, etc.)

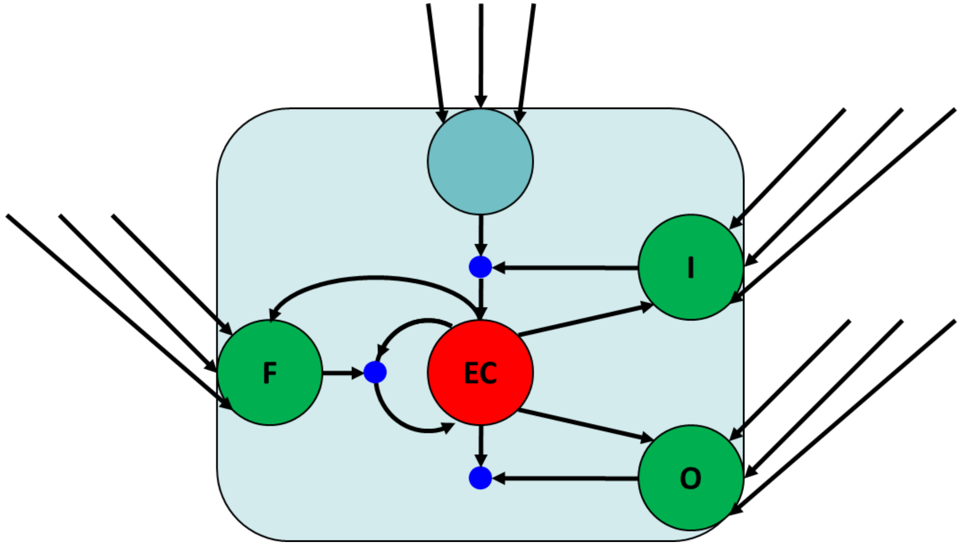
Vanishing Gradient:

- RNN “forget” time lags > 10
- 90s: History compression, etc.

Long Short Term Memory (LSTM):

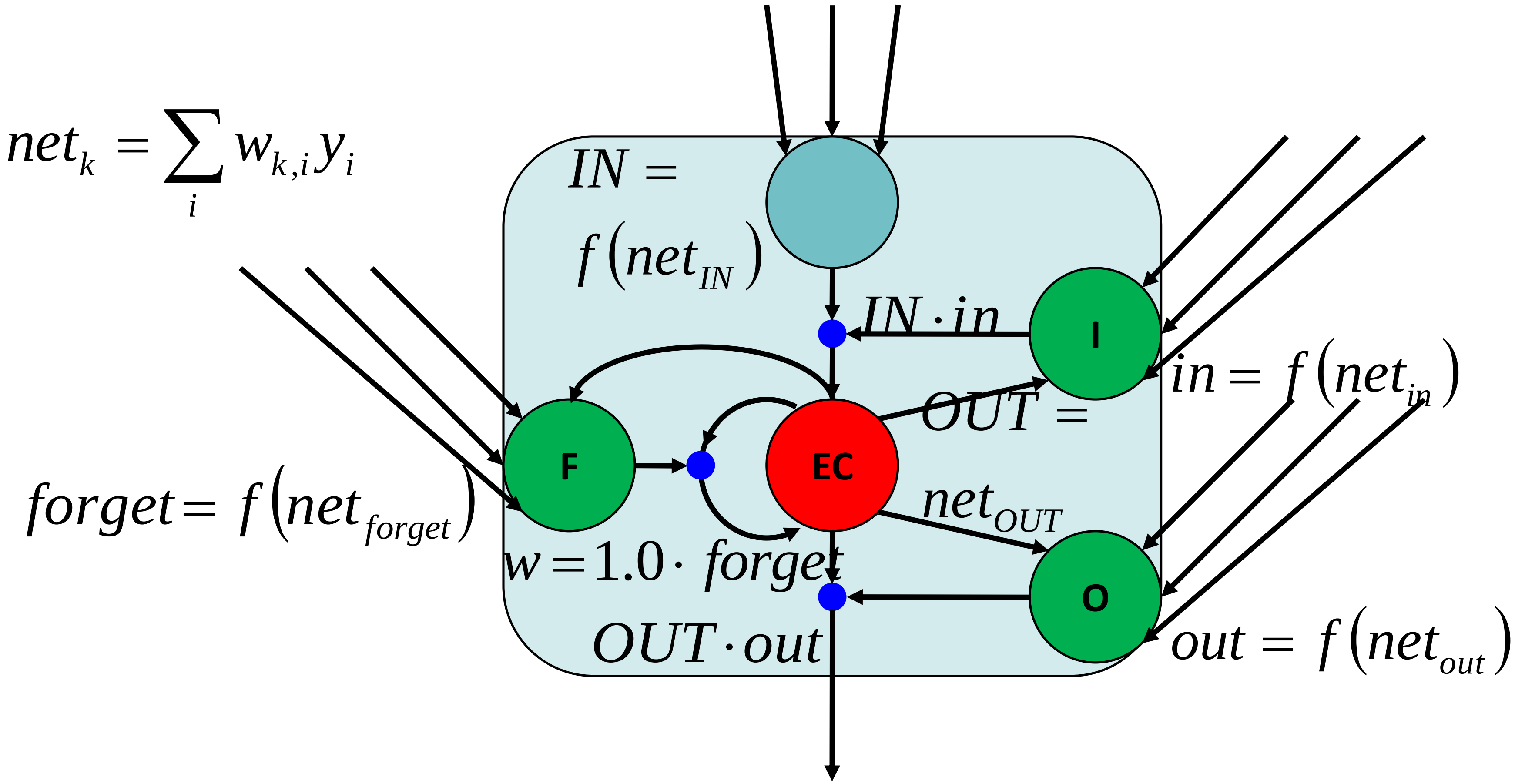
- Late 90s: Hochreiter and Schmidhuber
- Later: Graves and Schmidhuber
- ... And others...

Input, Output, Forget

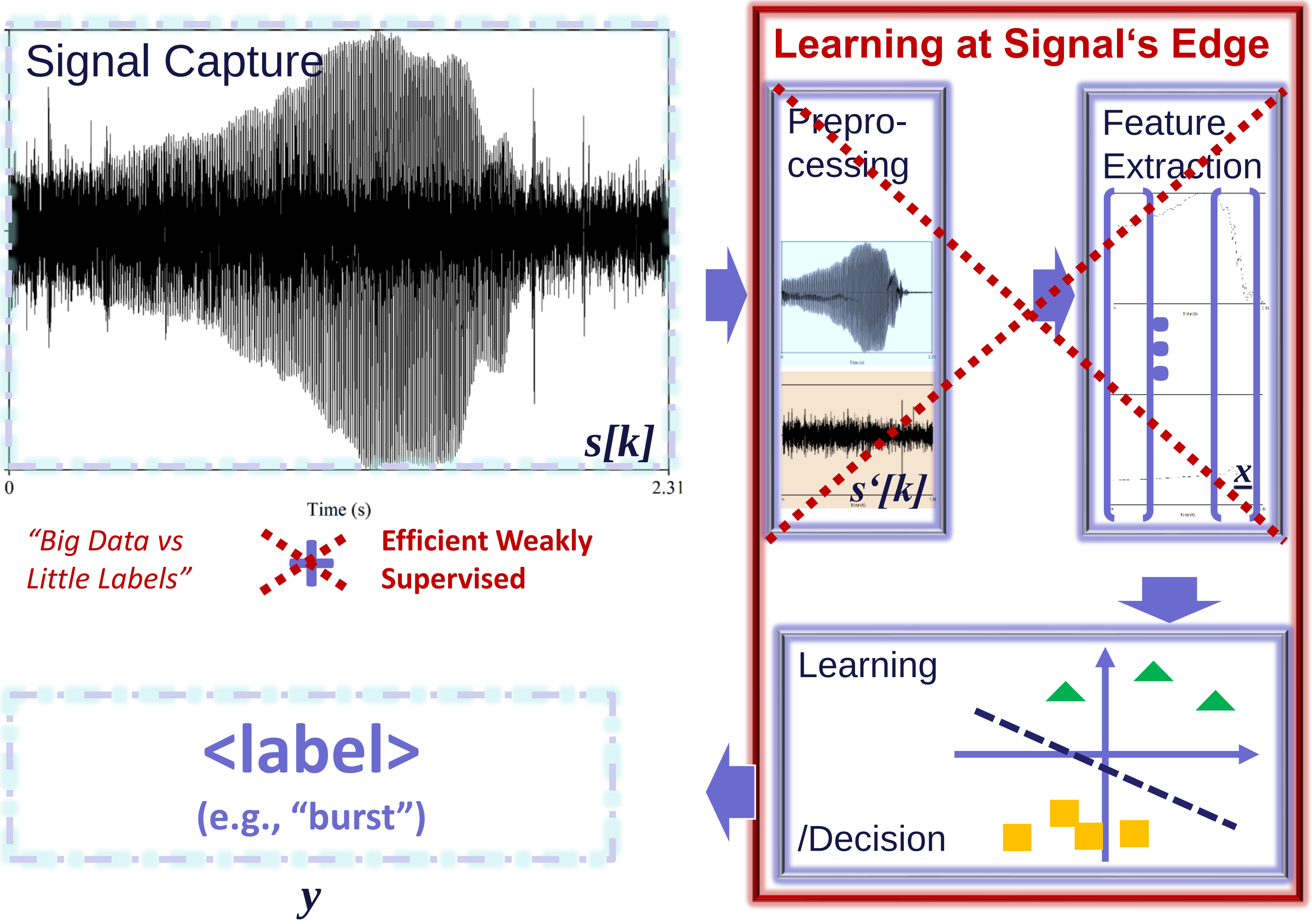


Universität Augsburg University

Current Cell (Input, Output, Forget)



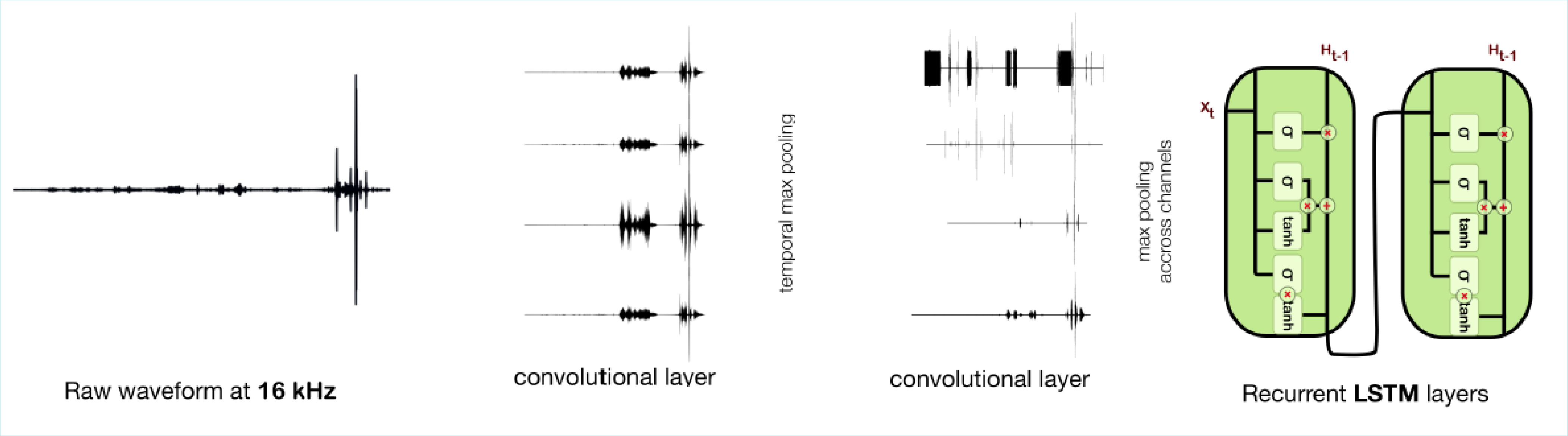
End-to-End Learning



Universität Augsburg University

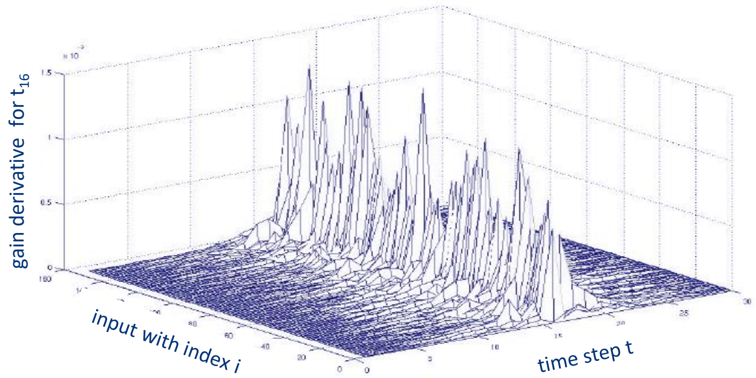
End-to-End Systems

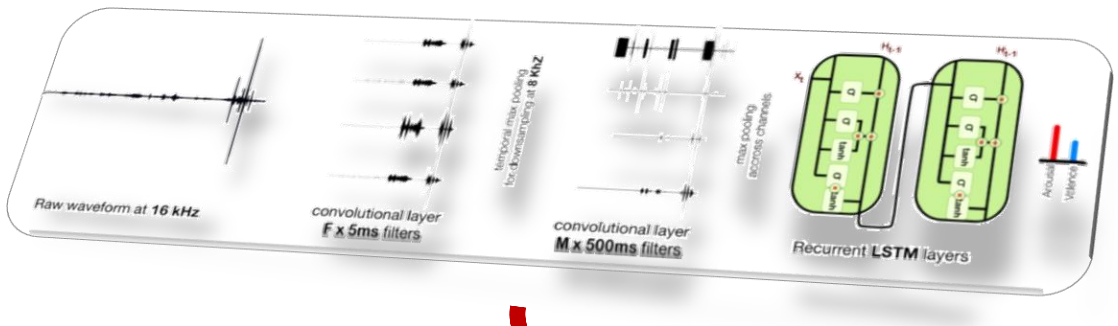
- The whole processing chain is learnt simultaneously.
- Basic system architecture:
  1. CNN used extract features from the raw time representation
  2. A subsequent RNN to learn temporal dependancies
  3. Fully Connected Layer



End-to-End – a black box?

- CNN activations correlate with standard speech features

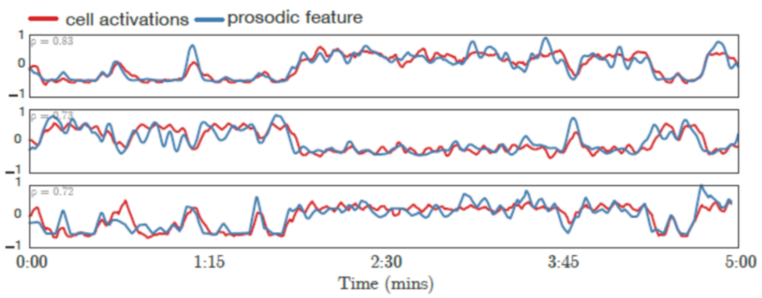


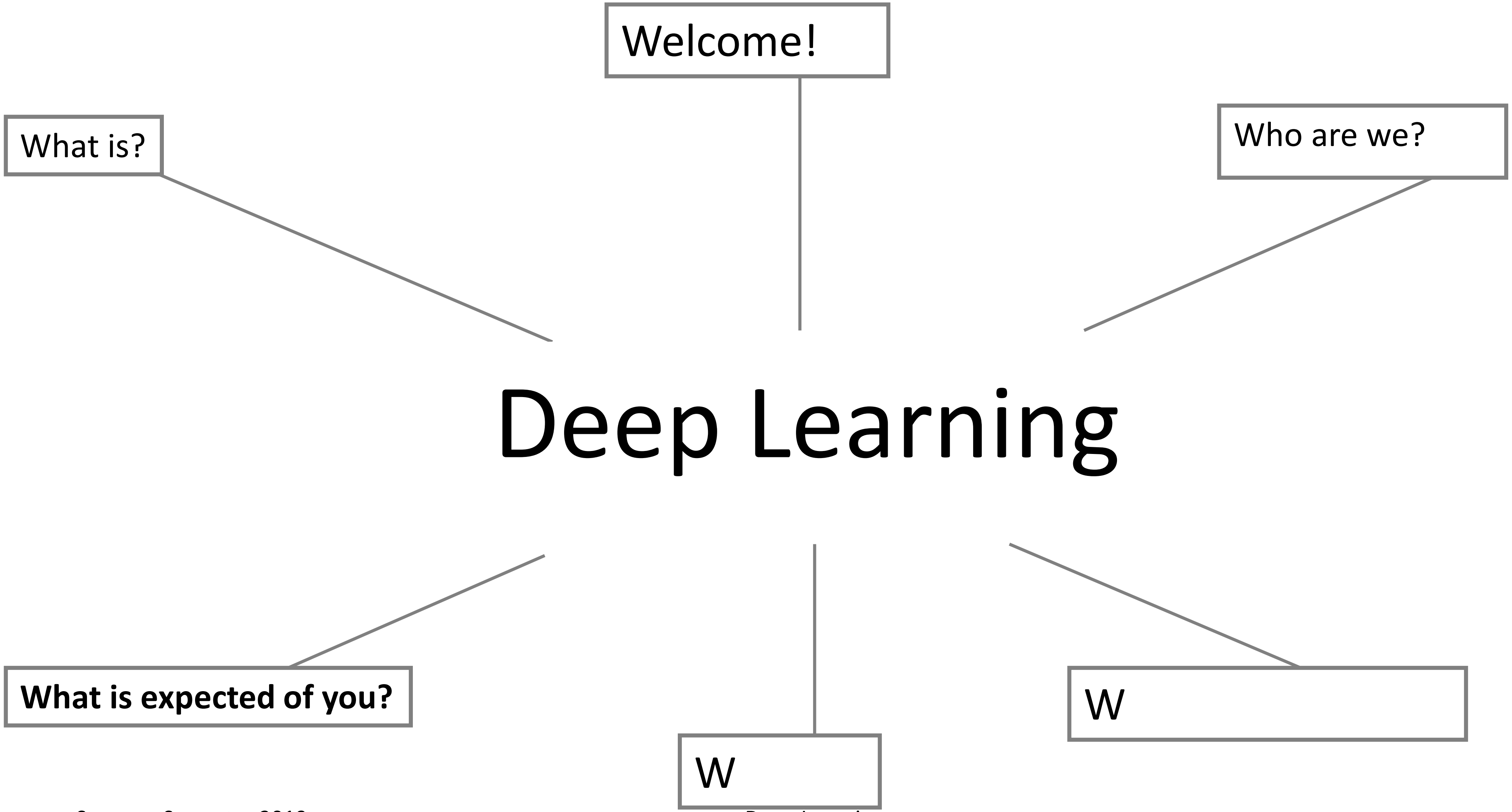


energy range (.77)

loudness (.73)

F0 mean (.71)




Welcome!
What is?
Who are we?
Deep Learning
What is expected of you?
W
W

Knowledge

- The lecture content includes an overview of Neural Network developments, covering all aspects from basic concepts to complex models. The course will cover Neural Network architectures that are suitable for a variety of data types and signals from different domains, such as audio, speech, vision, and text

Skills

- In the Tutorials, students will be familiarised with the latest Deep Learning toolkits such as Tensorflow and Keras and learn how to train and evaluate Deep Neural Networks for different applications.

Rest of the course

Lectures: Tuesdays 12:15 – 13:45 in 2045 N

- Topics Covered:
  - Feed Forward, Convolutional & Recurrent Networks
  - Attention Modelling and Connectionist Temporal Classification
  - Machine Learning
  - Natural language processing

Tutorials: Wednesdays 12:15 - 13:45 in 1005 N

- Skills gained in using
  - Python, Tensorflow, Keras

Welcome!

What is ISAM?

Who are we?

Deep Learning

What is expected of you?

W

What do we expect?

What do we expect?

Basic knowledge

- In signal processing and additional knowledge in the fields of Machine Learning, Data Mining, or more generally Pattern Recognition is of advantage.

Target groups

- Include post-graduate students of Electronics and Telecommunications Engineering, Information Technology, Computer Sciences and related studies.

The teaching language is English.

Interest & Interaction!

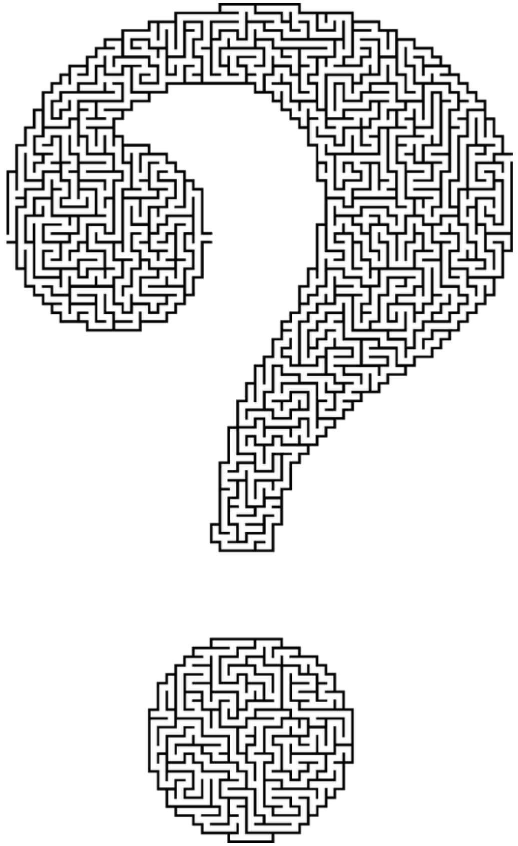


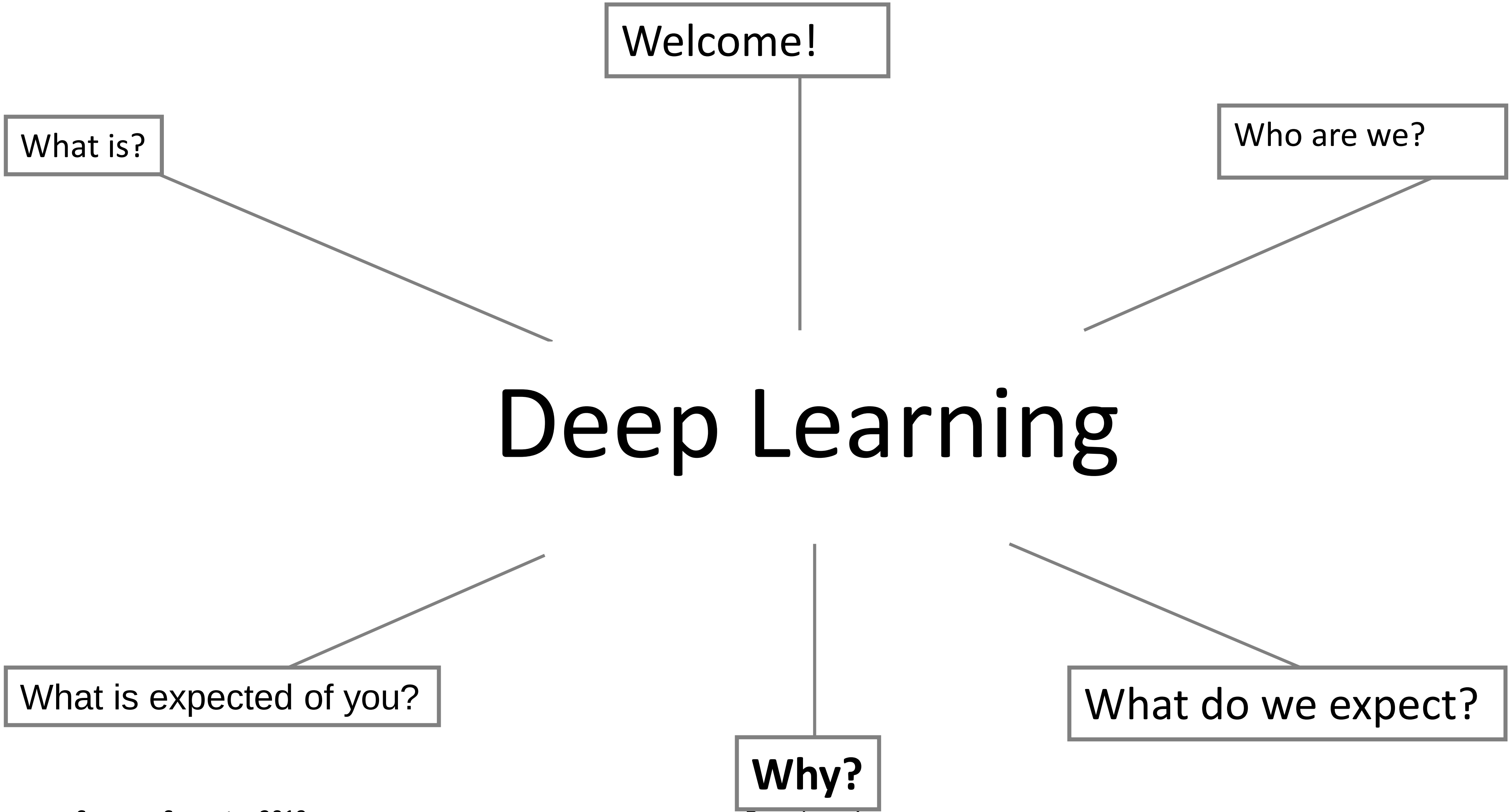
Image Source: https://pixabay.com

Lecture Timetable

Date	Topic	Lecturer
23.04.2019	Introduction	Prof. Schuller
30.04.2019	Applied Maths / Machine learning concepts	Dr. Cummins
07.05.2019	Feed Forward Networks	Dr. Cummins
14.05.2019	Regularisation in Neural Networks	Dr. Cummins
21.05.2019	Convolutional Neural Networks	Dr. Cummins
28.05.2019	Recurrent Neural Networks	Dr. Cummins
04.06.2019	Introduction to NLP	Dr. Cummins
11.06.2019	No Lecture (Dienstag nach Pfingsten)	-
18.06.2019	Attention and Connectionist Temporal Classification	Ass. Prof Zhao
25.06.2019	Data Representation Learning	Dr. Cummins
02.07.2019	Current Challenges and where to next	Dr. Cummins
09.07.2019	Deep Learning Master Class	Prof. Schuller
16.07.2019	Deep Learning Master Class	Prof. Schuller
23.07.2019	Wrap-Up Lecture	Dr. Cummins

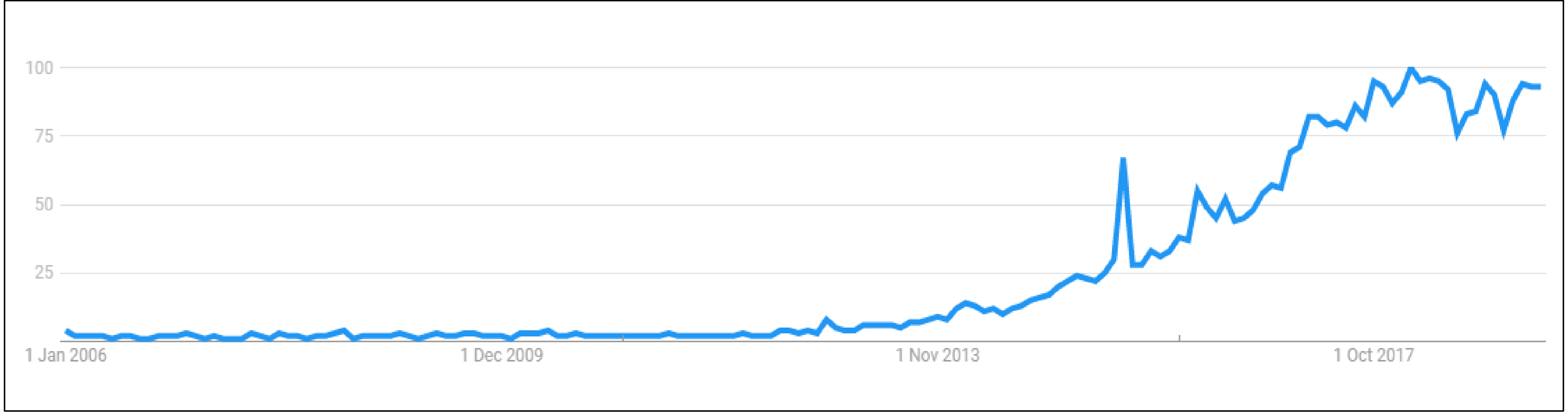
Tutorial Timetable

Date	Topic	Main Tutor
24.04.2019	Python Introduction	S. Amiriparian
01.05.2019	No Tutorial: Tag der Arbeit	-
08.05.2019 15.05.2019	Feed Forward Networks	S. Amiriparian
22.05.2019 29.05.2019	Convolutional Neural Networks	S. Amiriparian
05.06.2019 12.06.2019	Recurrent Neural Networks	S. Amiriparian
19.06.2019 26.06.2019	Encoder Decoder Networks	S. Amiriparian
03.07.2019 10.07.2019 17.07.2019 24.07.2019	Group Challenge	S. Amiriparian


Welcome!
What is?
Who are we?
Deep Learning
What is expected of you?
Why?
What do we expect?

Growing Interest

- Google trends graph on the term ‘Deep Learning’
  - Clear trend in increasing world-wide interest



Why study Deep Learning?

Growing Job Market

- Germany seeking to become a global AI powerhouse
  - 3 Billion Euro to be invested by 2025
    - Mobility, Health, Energy, Industry 4.0
  - 100 new AI chairs to be created
  - 12 new centres for research
  - Attractive working conditions & salaries

Video source: https://www.ki-strategie-deutschland.de

Money!

- AI experts can command huge salaries
  - AI specialist can make between $300,000 and $500,000 a year in salary and stock
    - E.g. Average wage at Google Deep Mind ~ $345,000
    - E.g. OpenAI employed 52 people in 2016 and spent more than $7 million on salaries
      - Source: New York Times


The New York Times
Tech Giants Are Paying Huge Salaries for Scarce A.I. Talent


The New York Times
A.I. Researchers Are Making More Than $1 Million, Even at a Nonprofit

Plenty of Research Still needed

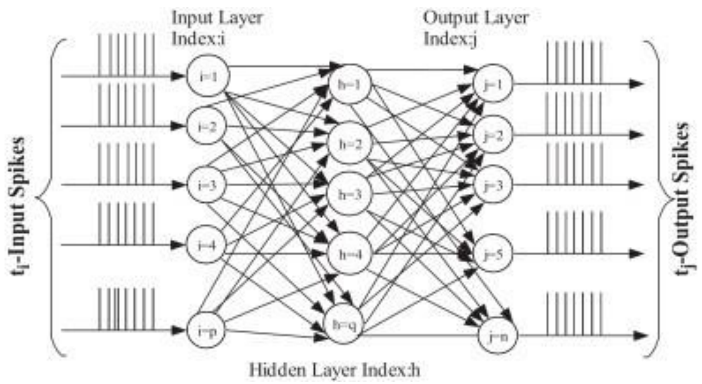
- Prominent AI failures
  - **UBER, 2018**: An experimental autonomous Uber vehicle struck and killed a pedestrian in Tempe, Arizona
  - **Goldman Sachs, 2018**: AI predictions the 2018 World Cup were almost all wrong
  - **Amazon 2017:** AI-enabled recruiting software is found to be gender biased
  - **IBM, 2016:** Watson for Oncology cancel after spending $62 million on the project after system was found to give unsafe advice
  - **Microsoft, 2016:** Twitter chatbot began stating 9/11 conspiracy theories and Nazi sentiments
  - **Google 2015**: Image recognition algorithms in Google Photos was found to be labelling black people as gorillas

Plenty of Research Still needed

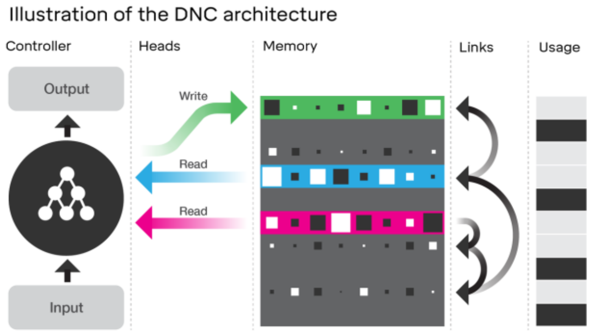
- **Data Requirements**
  - Robust deep learning solutions require massive amounts of training data
- **Privacy Concerns**
  - Large deep neural networks are difficult to implement in edge computing scenarios
- **Explainable AI**
  - Deep learning systems are considered ‘black-boxes’
  - Further, they contain millions of data connections
  - Very difficult to determine exactly how a decision was reached
- **Butterfly effects / Adversarial attacks**
  - Small variations in the input data can lead to drastically different results.
  - By adding small amounts of noise it is possible to fool deep learning networks

Plenty of Research Still needed

- Next Generation of AI examples
  - **Spiking Neural Networks**
    - Data driven event based computation
    - Encode information as events in time
  - **Differentiable Neural Computer**
    - Neural network with a memory matrix from which it can read and write data
    - Allows network to have long-term memory
  - **Meta-Learning**
    - Networks trained to be able to learn the learn processes



Source: https://towardsdatascience.com/



Source: https://becominghuman.ai