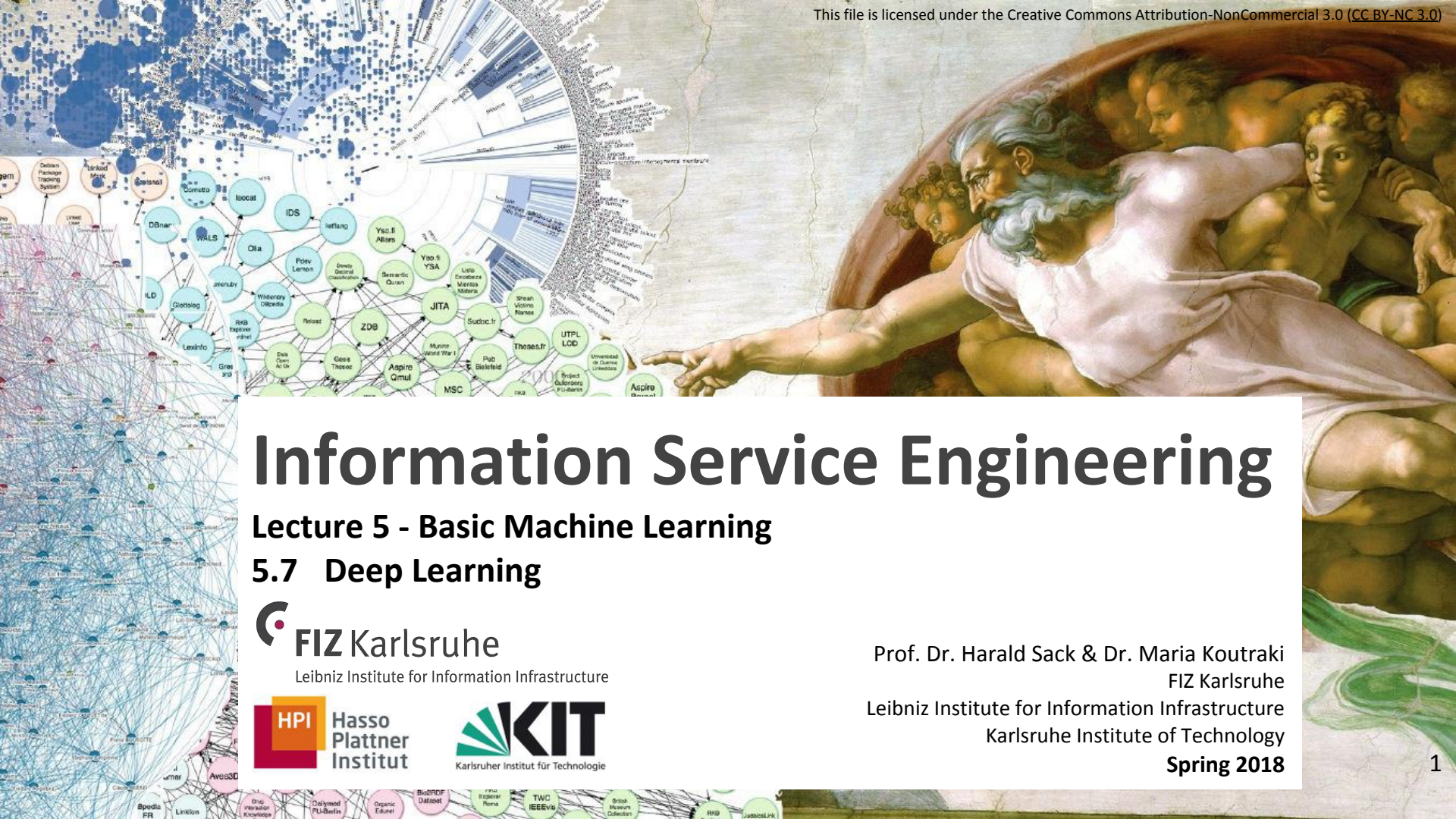




Information Service Engineering

Lecture 5 - Basic Machine Learning

5.7 Deep Learning



FIZ Karlsruhe

Leibniz Institute for Information Infrastructure



Hasso
Plattner
Institut



Karlsruher Institut für Technologie

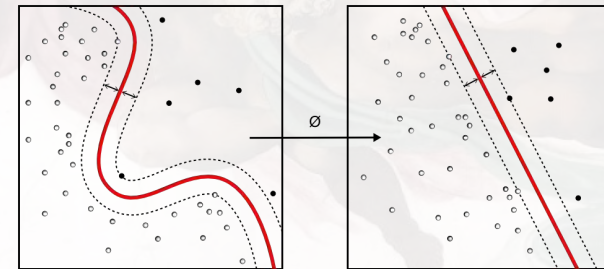
Prof. Dr. Harald Sack & Dr. Maria Koutraki
FIZ Karlsruhe

Leibniz Institute for Information Infrastructure
Karlsruhe Institute of Technology
Spring 2018

Information Service Engineering

Lecture 5: Basic Machine Learning

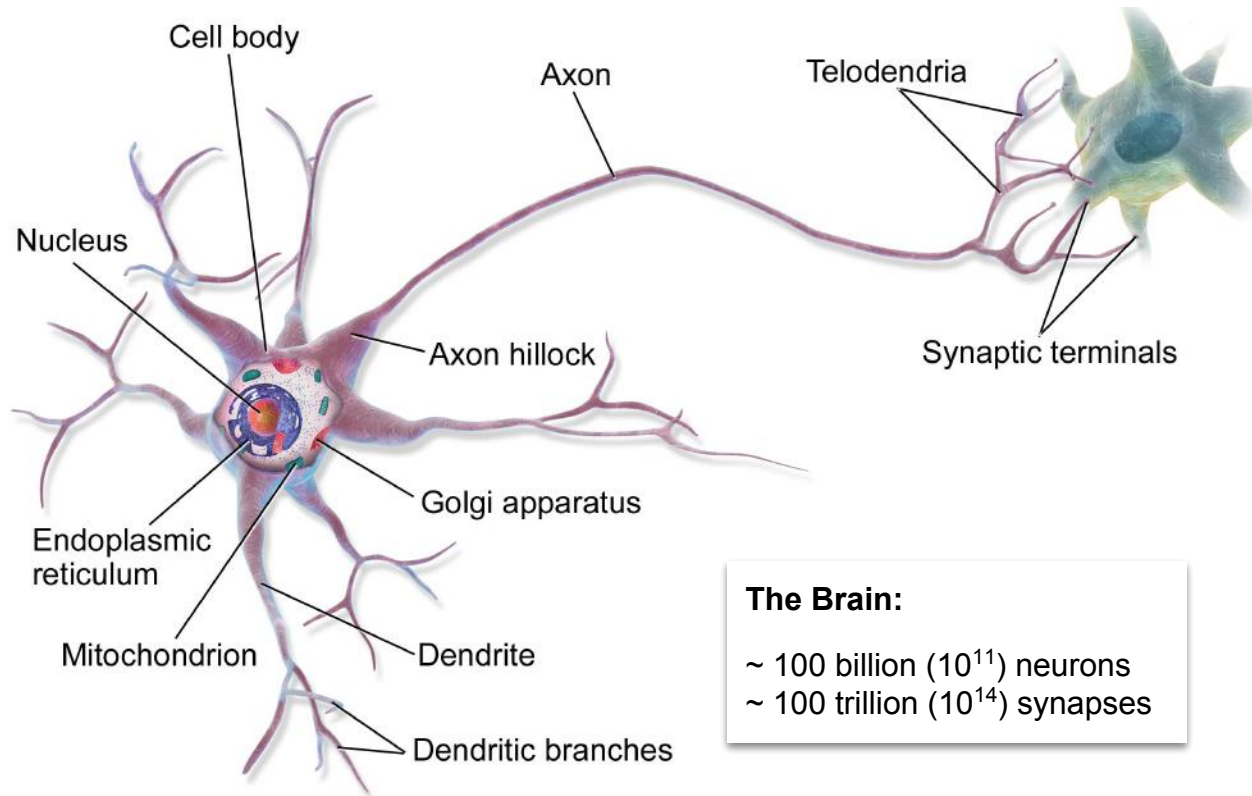
- 5.1 A Brief History of AI
- 5.2 Introduction to Machine Learning
- 5.3 Evaluation and Generalization Problems
- 5.4 Basic ML Algorithms
- 5.5 Basic ML Algorithms II
- 5.6 Unsupervised Learning
- 5.7 Deep Learning**



Biological Neural Networks

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.6 Unsupervised Learning

McCulloch & Pitts (1943)



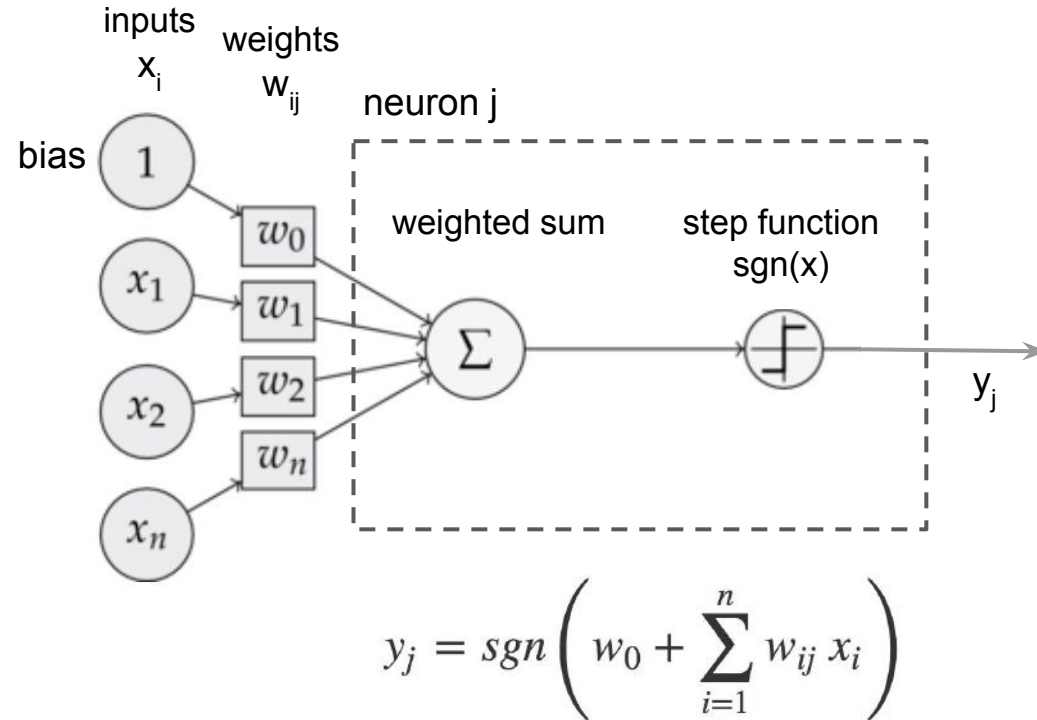
The Brain:

~ 100 billion (10^{11}) neurons
~ 100 trillion (10^{14}) synapses

- The majority of **neurons** encode their outputs or activations as a series of brief electrical pulses.
- **Dendrites** are the receptive zones that receive activation from other neurons.
- The **cell body (soma)** of the neuron's processes the incoming activations and converts them into output activations
- **Axons** are transmission lines that send activation to other neurons.
- **Synapses** allow weighted transmission of signals (via **neurotransmitters**) between axons and dendrites to build up large neural networks.

McCulloch-Pitts Neurons

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning



- **Artificial neurons** have the same basic components as biological neurons.
- The simplest ANNs consist of a set of **McCulloch-Pitts neurons** labelled by indices i, j
- There is a **threshold bias** $x_0=1$
 - allows bias to be captured from input neurons
 - allows normalization of output thresholds without loss of generality

Building an Artificial Neural Network

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning

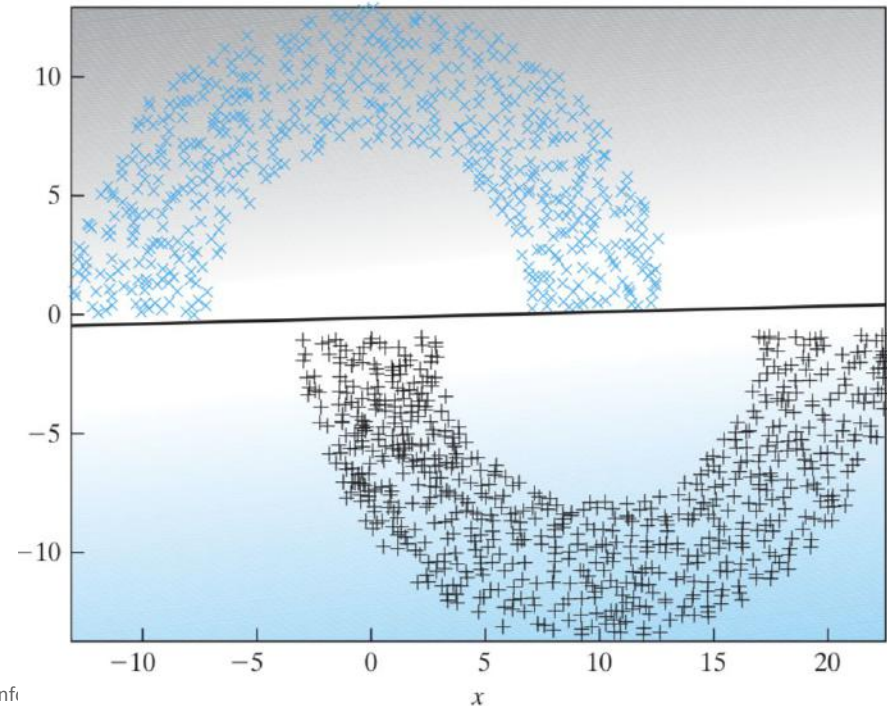
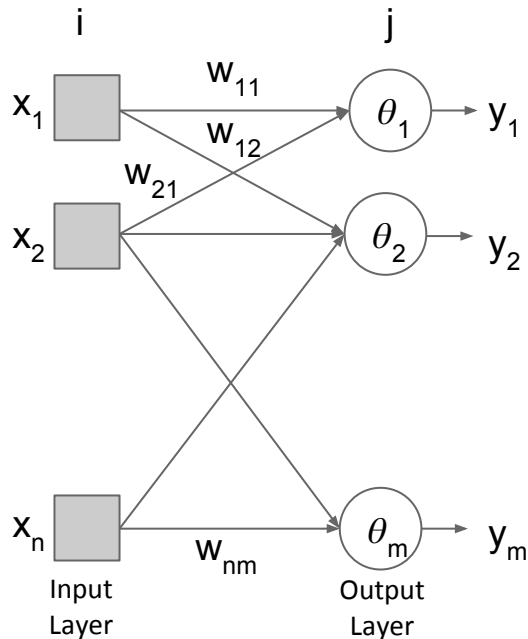
Using artificial neural networks to solve real problems is a multi-stage process:

1. **Understand** and **specify the problem** in terms of inputs and required outputs.
2. Take the **simplest form of network** that might be able to solve the problem.
3. Try to find appropriate connection **weights** and neuron **thresholds** so that the network produces appropriate outputs for each input in its training data.
4. **Test** the network on its **training data**, and also on **new** (validation/testing) **data**.
5. If the network doesn't perform well enough, go back to 3 and work harder.
6. If the network still doesn't perform well enough, go back to 2 and work harder.
7. If the network still doesn't perform well enough, go back to 1 and work harder.
8. Problem solved – move on to next problem.

Perceptron

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning

An arrangement of **one input layer of activations feeding forward to one output layer** of McCulloch-Pitts neurons is known as a **simple Perceptron**

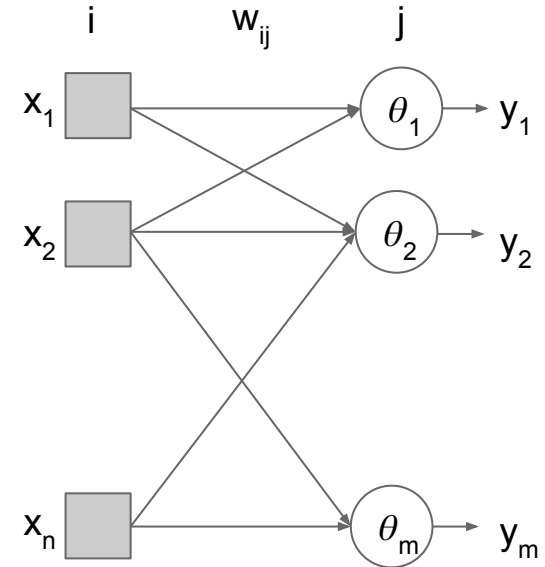


What the Perceptron can Do

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning

Learning with a Perceptron

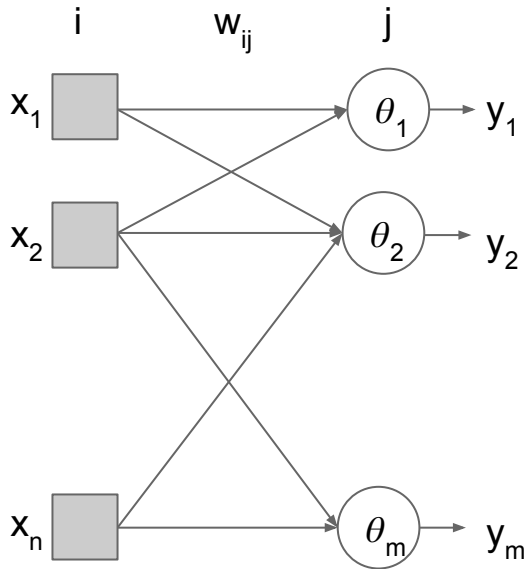
- Initialize weights randomly
- Take one sample x_i and predict y_i
- For erroneous predictions update weights
 - If the output was $y_i = 0$ and desired output $d_i = 1$, increase weights
 - If the output was $y_i = 1$ and desired output $d_i = 0$, decrease weights
- Repeat until no errors are made



Perceptron (2)

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning

An arrangement of **one input layer of activations feeding forward to one output layer** of McCulloch-Pitts neurons is known as a simple **Perceptron**



Network activation: for each neuron in output layer $j=1, \dots, m$

$$y_j = \text{sgn} \left(\sum_{i=1}^n w_{ij} x_i - \Theta_j \right)$$

Perceptron learning function:

for each feature weight $i=1, \dots, n$ of neuron in output layer $j=1, \dots, m$

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}(t)$$

$$\Delta w_{ij} = \eta (d_j - y_j(t)) x_i$$

learning rate
coefficient

desired
output

actual
output

Perceptron Learning Algorithm

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning

Input vector	$\mathbf{x}(t) = [+1, x_1(t), x_2(t), \dots, x_n(t)]^T$
Weight vector	$\mathbf{w}(t) = [b, w_1(t), w_2(t), \dots, w_n(t)]^T$, b = bias
Actual response	$y(t)$ (quantized)
Desired response	$d(t)$
Learning-rate	η , $0 < \eta \leq 1$

1. **Initialization.** Set $\mathbf{w}(0) = 0$. Then perform the following computations for time-step $t = 1, 2, \dots$
2. **Activation.** At time-step t , activate the perceptron by applying continuous-valued input vector $\mathbf{x}(t)$ and desired response $d(t)$.
3. **Computation of Actual Response.** Compute the actual response of the perceptron as $y(t) = \text{sgn}[\mathbf{w}^T(t) \mathbf{x}(t)]$
4. **Adaptation of Weight Vector.** Update the weight vector of the perceptron to obtain, where
$$\mathbf{w}(t+1) = \mathbf{w}(t) + \eta [d(t) - y(t)] \mathbf{x}(t) \quad d(t) = \begin{cases} +1 & \text{if } x(t) \text{ belongs to class } c_1 \\ -1 & \text{if } x(t) \text{ belongs to class } c_2 \end{cases}$$
5. **Continuation.** Increment time step $t = t+1$ by one and go back to step 2.

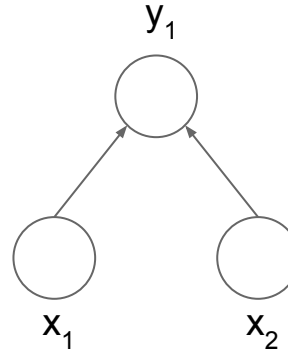
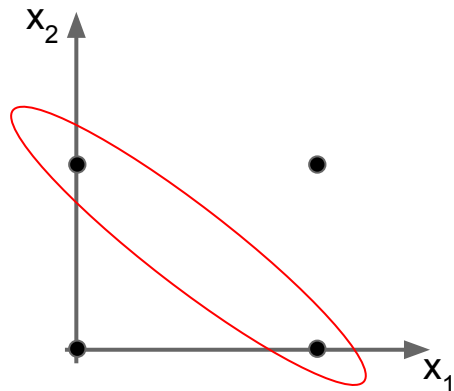
The Problem with the XOR

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning

- However, the simple **exclusive or (XOR)** cannot be solved by perceptrons

[Minsky and Papert, "Perceptrons", 1969]

- $0 w_1 + 0 w_2 < \theta \rightarrow 0 < \theta$
- $0 w_1 + 1 w_2 > \theta \rightarrow w_2 > \theta$
- $1 w_1 + 0 w_2 > \theta \rightarrow w_1 > \theta$
- $1 w_1 + 1 w_2 < \theta \rightarrow w_1 + w_2 < \theta$



x_1	x_2	y_1
0	0	0
0	1	1
1	0	1
1	1	0

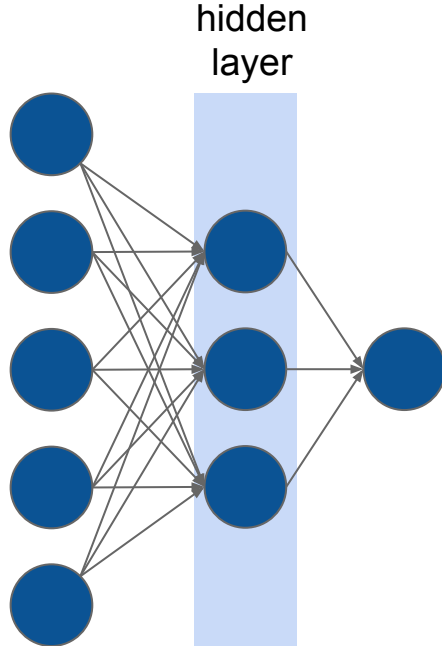
The First “AI Winter”

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning



Multilayer Neural Network

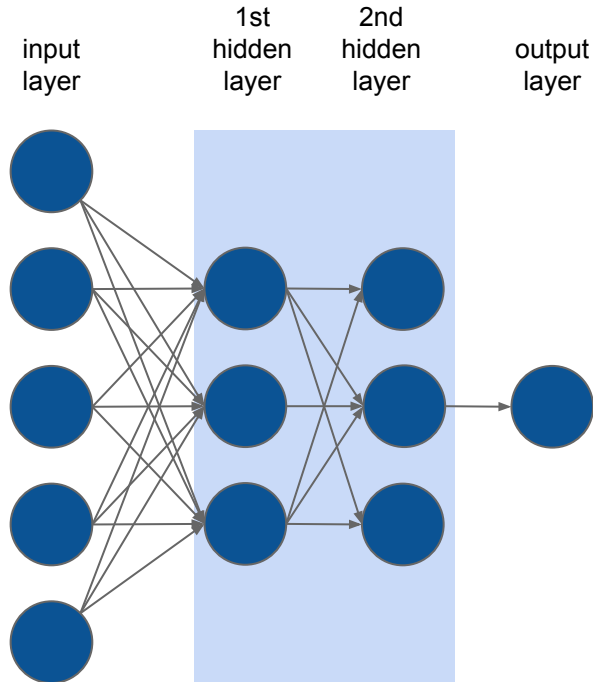
Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning



- The perceptron is limited to the classification of linearly separable patterns
- **Solution:**
 - use **non-monotonic activation functions**
 - use multilayer neural networks, which include a least **one hidden layer** of neurons with non-linear activations functions
- **Problem:** Desired output for hidden nodes is not known
- **New Learning Algorithm:** Backpropagation Algorithm

Backpropagation Algorithm

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning



- **forward phase**

- synaptic weights of the network are fixed
- input signal is propagated through the network layer by layer, until it reaches the output.
- changes only in activation potentials and outputs

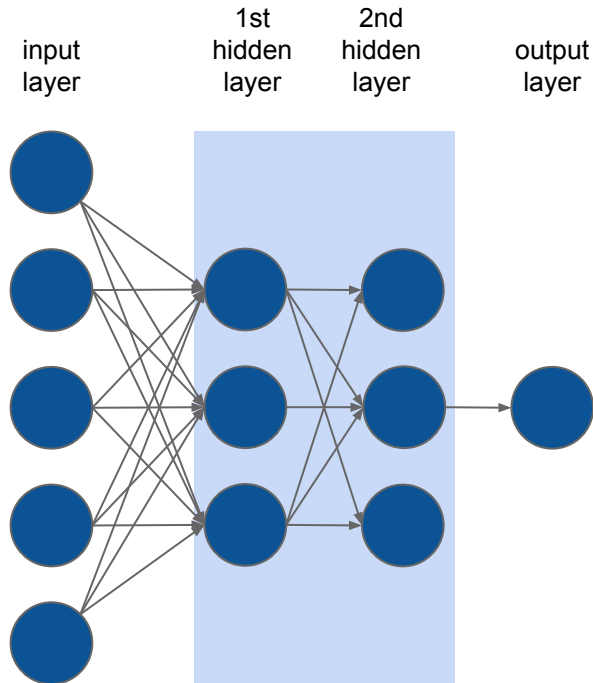
- **backward phase**

- An error signal is produced by comparing network output with desired response
- resulting error signal is propagated through the network layer by layer in the backward direction
- successive adjustments are made to the synaptic weights of the network.

(1) → forward propagation
(2) ← backward propagation

Feature Engineering

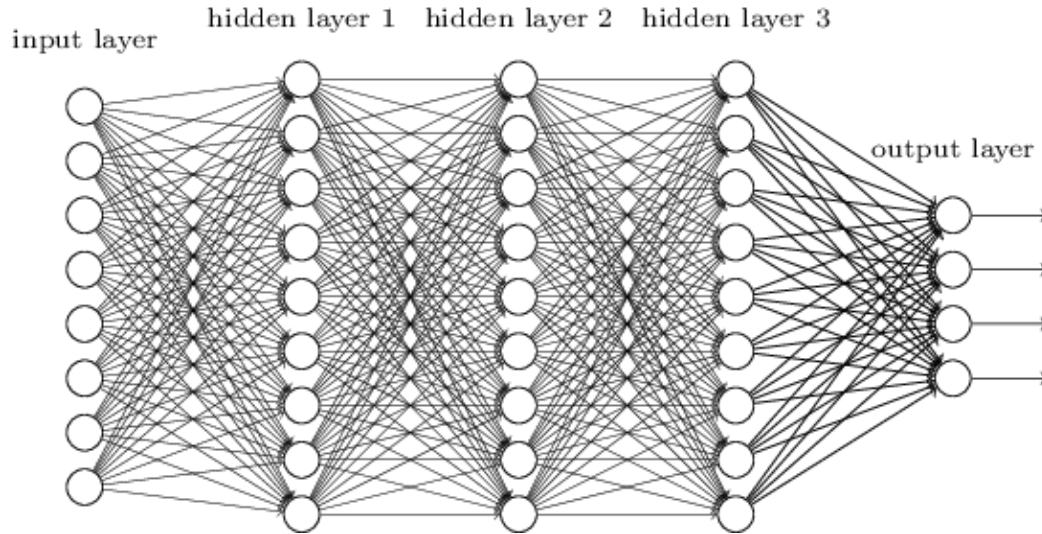
Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning



- The quality of your Machine Learning approach depends on the appropriate assembly of the **features** used for learning
- **Feature engineering** is the process of using domain knowledge of the data to create features that make machine learning algorithms work
- When working on a machine learning problem, feature engineering is manually designing what the input x_i should be.
- Coming up with features is
 - difficult,
 - time consuming,
 - requires expert knowledge.

Deep Learning

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning

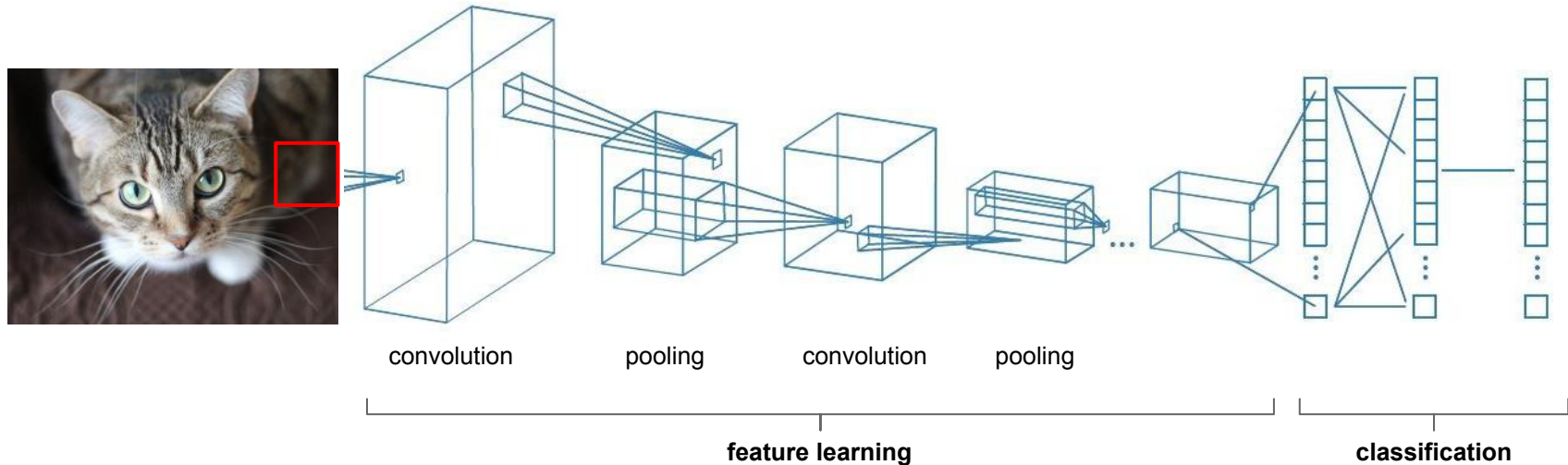


- Deep Learning uses deep (multi layer) neural networks
- Assume hidden layers with **1000 nodes each**
- In the given example, we need **2 Mio parameters** to optimize between the hidden layers only
- To classify images or for object detection, there would be no spatial invariance, i.e. **lack of generalization**

Convolutional Neural Networks

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning

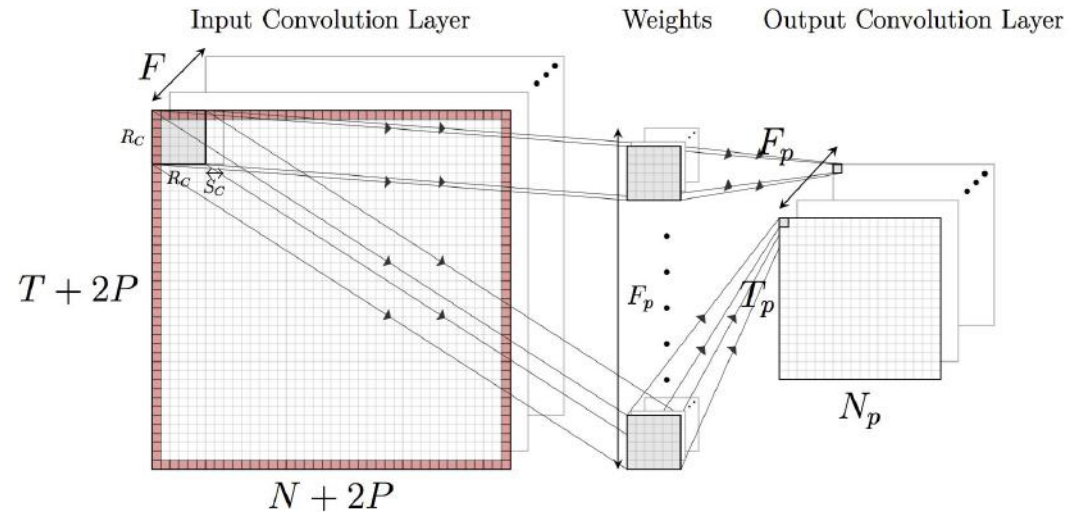
- CNNs are a special type of neural networks for processing **spatially arranged data**
- CNNs are particularly adapted for **visual analysis tasks**
- fundamental building blocks of CNNs: **convolution** and **pooling**



Convolution Layer

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning

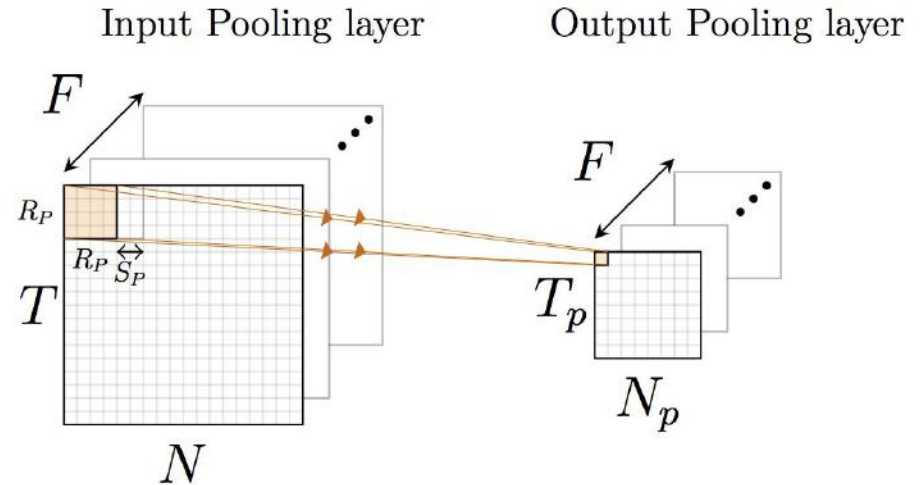
- **Purpose:** examine (filter) the input from different perspectives
- Each neuron checks a specific area of the input field using a **filter kernel**
- A filter examines the image for a certain **feature**, e.g. color, edges or brightness.
- Result of a filter is the weighted input of a range and is stored in the **convolutional layer**
- The depth of the convolutional layer is defined by the number of filters



Pooling Layer

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning

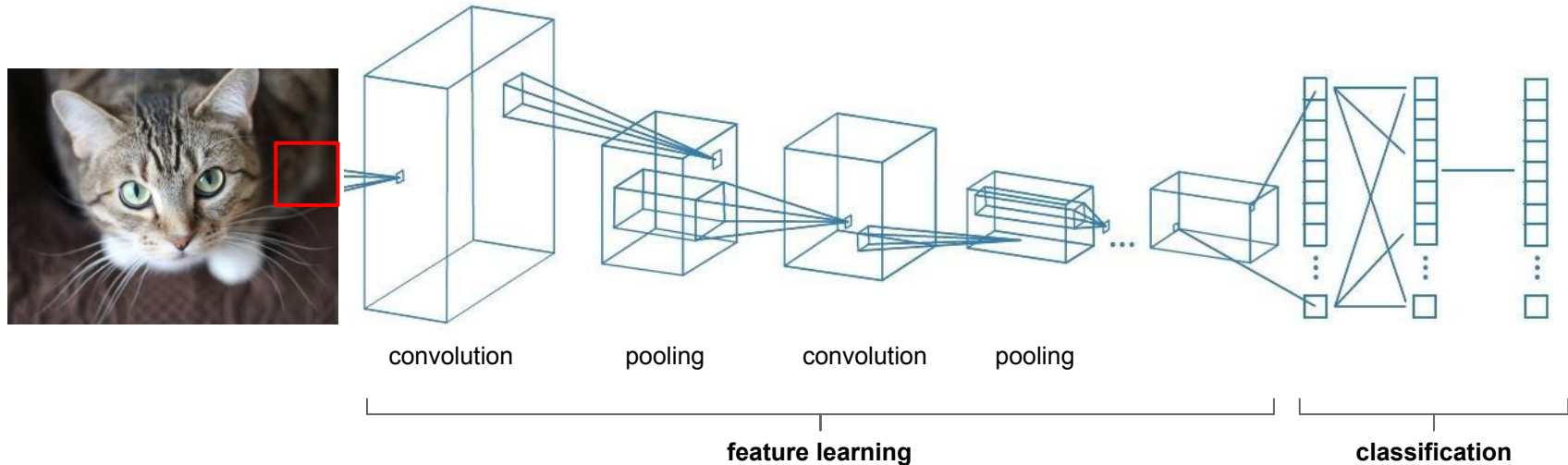
- **Purpose:** subsampling
- Information is compressed between the individual convolutional layers via **pooling layers** (dimensionality reduction)
- Pooling layers run through the **feature maps** created by the filters and compress them

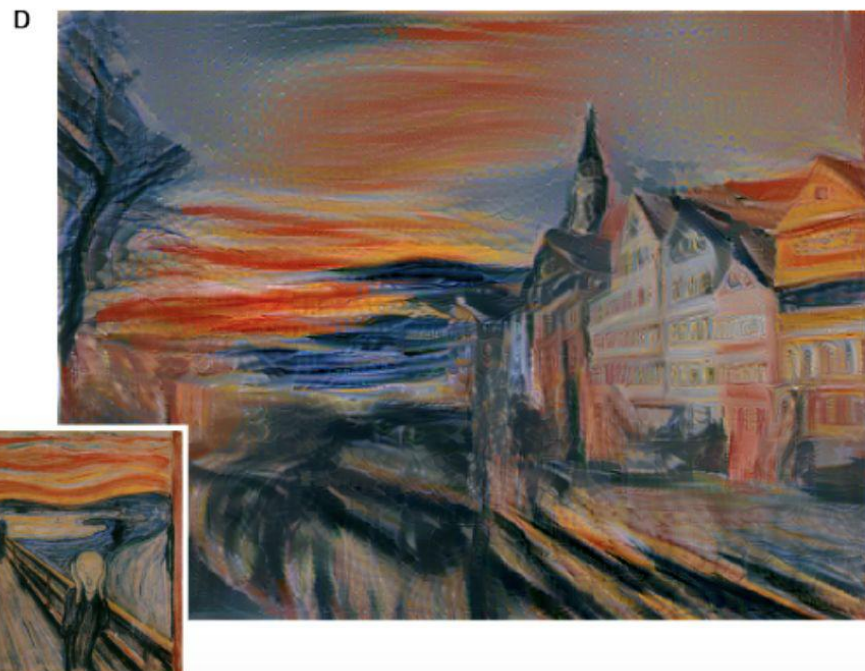


Convolutional Neural Networks

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning

- CNNs are a special type of neural networks for processing **spatially arranged data**
- CNNs are particularly adapted for **visual analysis tasks**
- fundamental building blocks of CNNs: **convolution** and **pooling**





CNNs - Example Applications

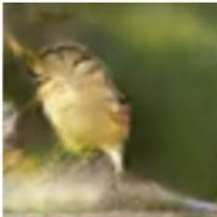
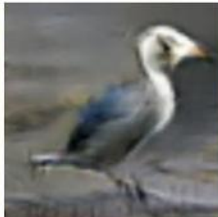
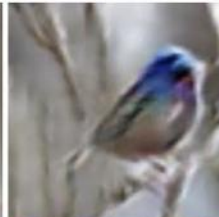
Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning



<https://deepart.io/>

Generative Adversarial Networks - GANs

Information Service Engineering - Lecture 5 Basic Machine Learning - 5.7 Deep Learning

Text description	This bird is blue with white and has a very short beak	This bird has wings that are brown and has a yellow belly	A white bird with a black crown and yellow beak	This bird is white, black, and brown in color, with a brown beak	The bird has small beak, with reddish brown crown and gray belly	This is a small, black bird with a white breast and white on the wingbars.	This bird is white black and yellow in color, with a short black beak
Stage-I images							
Stage-II images							

Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, Dimitris N. Metaxas:

[StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks.](#)

CoRR abs/1710.10916 (2017)



Putting it all together

Next Lecture...

Picture References:

- P.11: photo Pixabay, <https://pixabay.com/en/ice-snow-iceland-plateau-mountains-2817112/>
- P.16: photo Pexels, <https://www.pexels.com/photo/cat-whiskers-kitty-tabby-20787/>
- P.23: David Teniers the Younger, Archduke Leopold Wilhelm in his Gallery in Brussels (1651).
[https://commons.wikimedia.org/wiki/File:David Teniers the Younger - Archduke Leopold William in his Gallery at Brussels - Google Art Project.jpg](https://commons.wikimedia.org/wiki/File:David_Teniers_the_Younger_-_Archduke_Leopold_William_in_his_Gallery_at_Brussels_-_Google_Art_Project.jpg)