ETH zürich

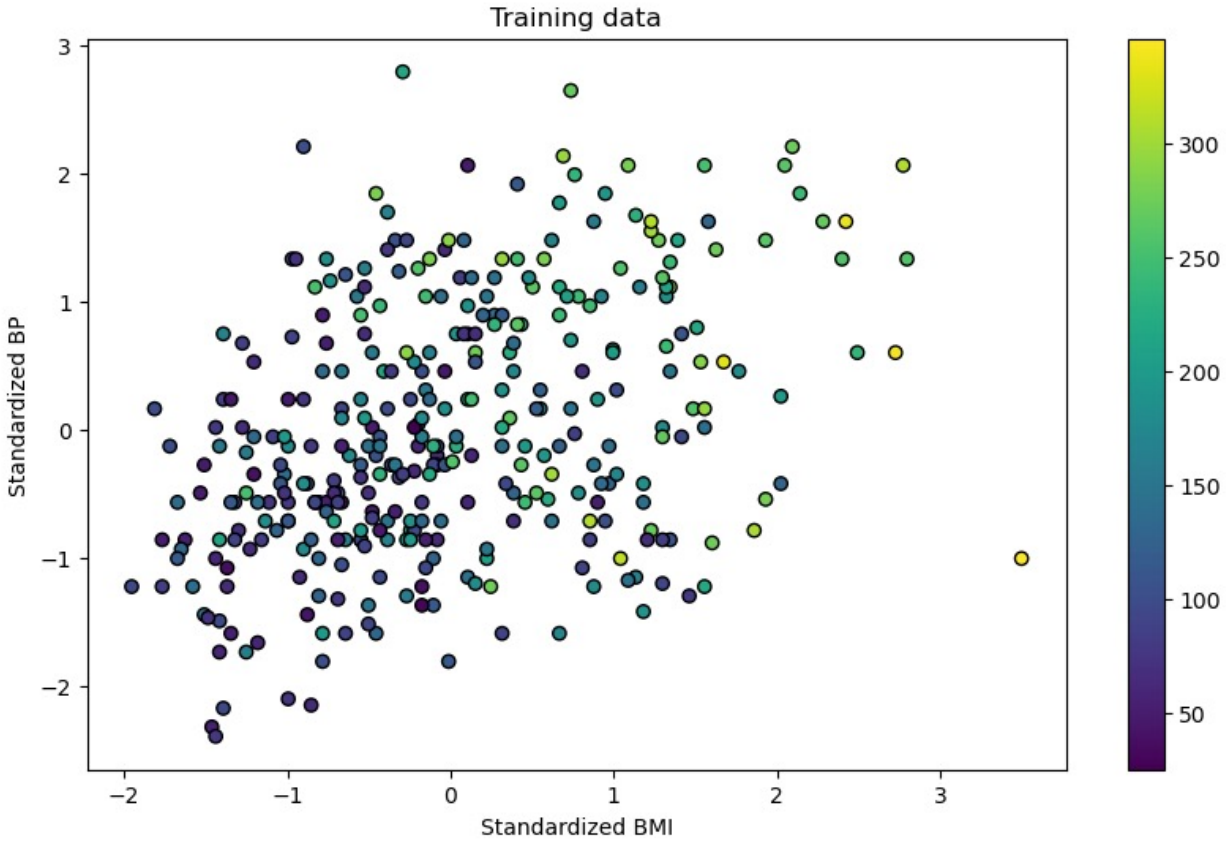
D-INFK Lehre

Cross-Validation und Deep Learning

Carlos Cotrini, Marcel Lüthi

Datenanalyse in der Physik 2024

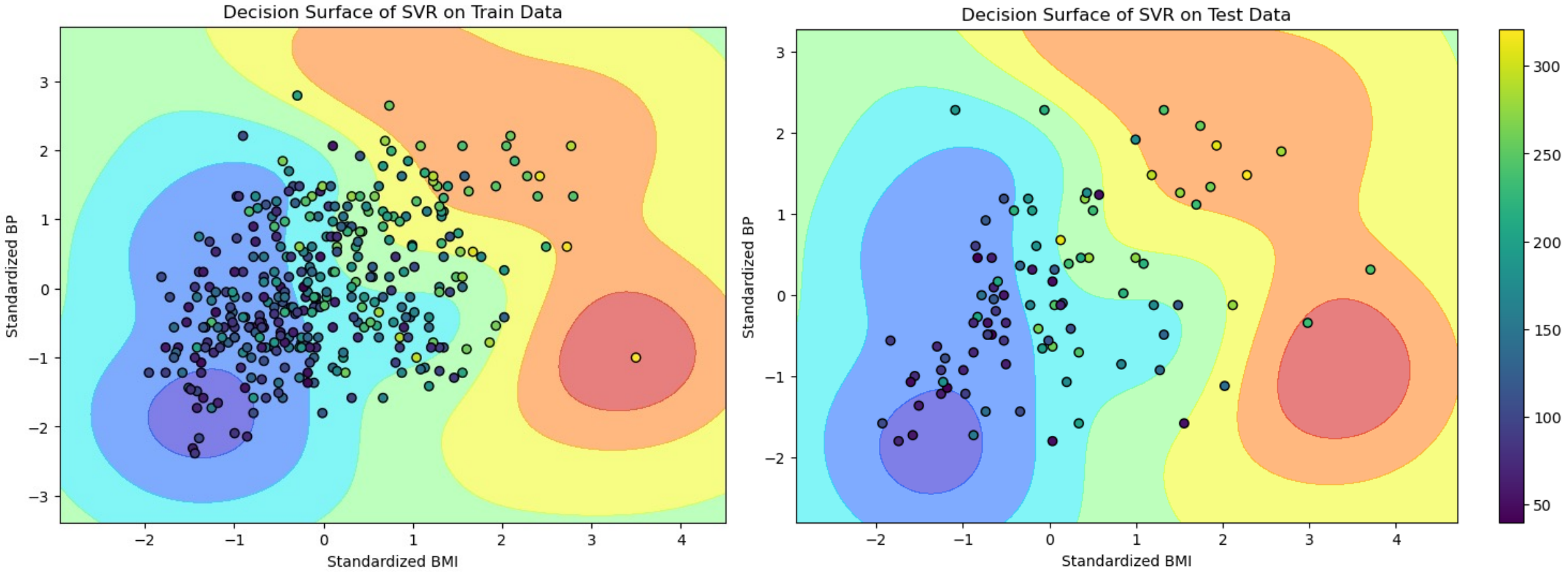
Prognose von Diabetes mit ML



- In der Übung haben wir mit einem Datensatz von Patienten mit Diabetes gearbeitet.

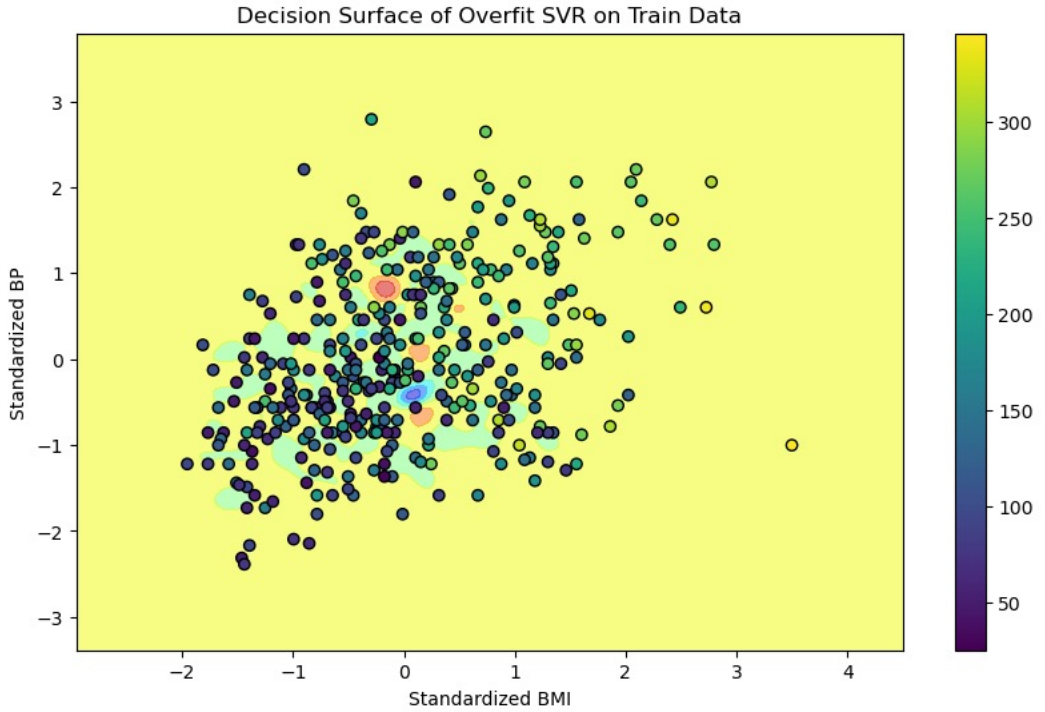
Prognose von Diabetes mit ML

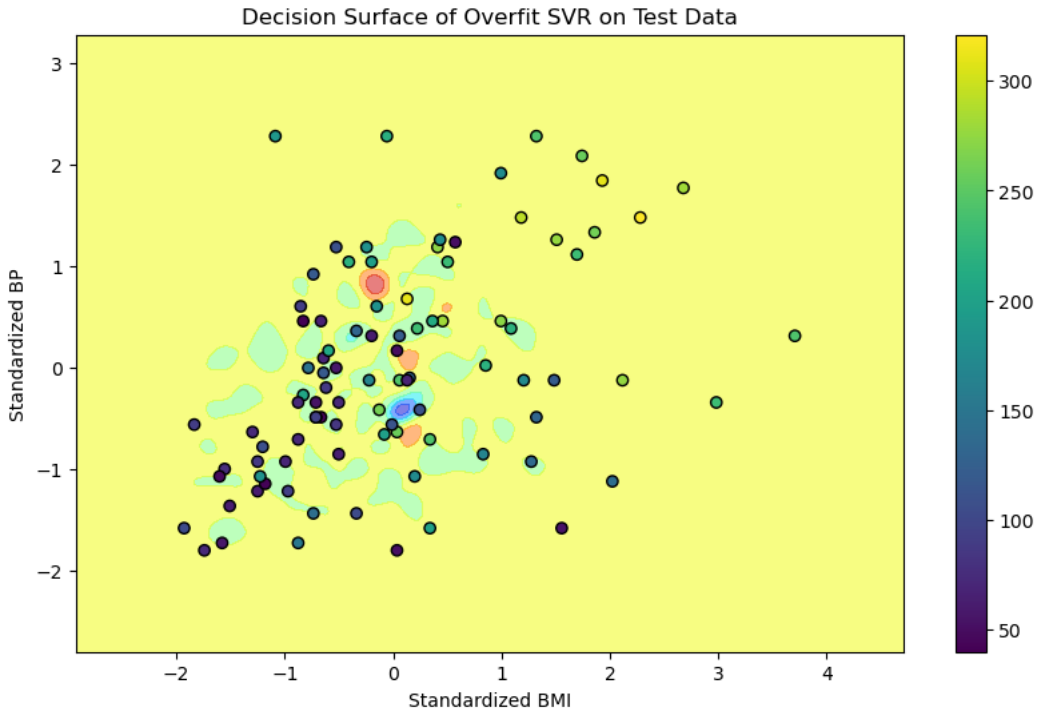
- Mit einfachen Modelle trainiert man Funktionen die gut über andere Daten verallgemeinern.



Prognose von Diabetes mit ML

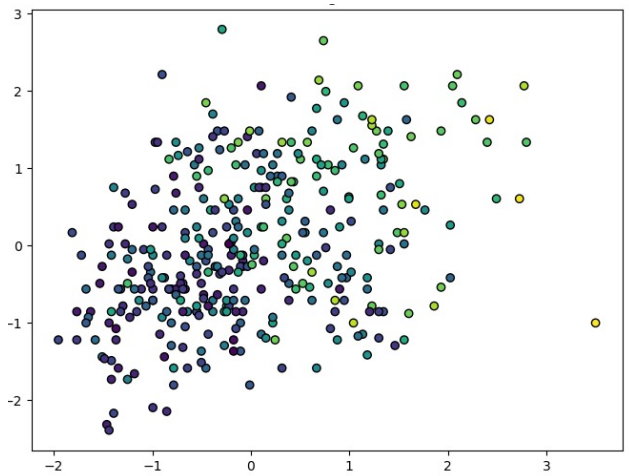
- Mit komplizierten Modellen trainiert man Funktionen die nicht so gut über andere Daten verallgemeinern.

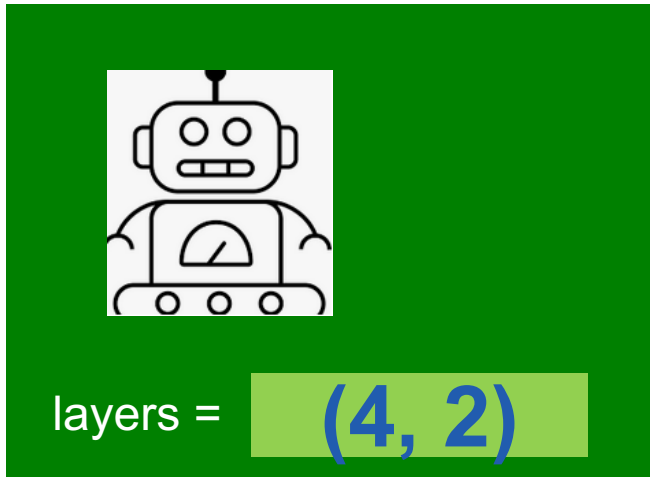


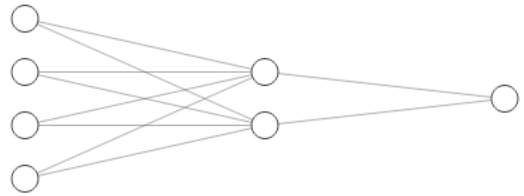


Overfitting und underfitting

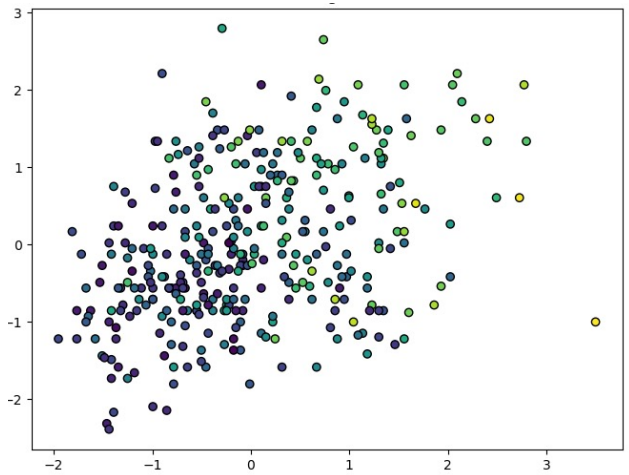
Das Problem der Auswahl eines Modells im maschinellen Lernen

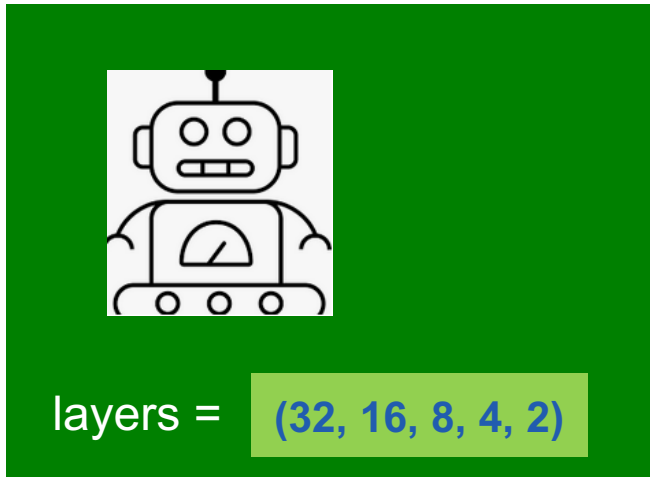


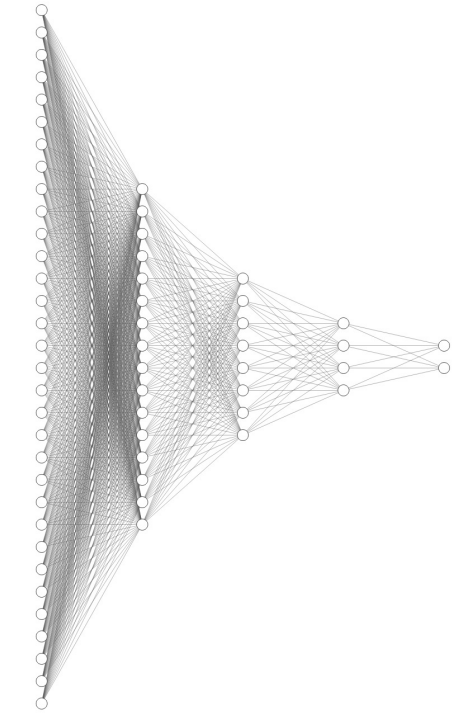




Das Problem der Auswahl eines Modells im maschinellen Lernen

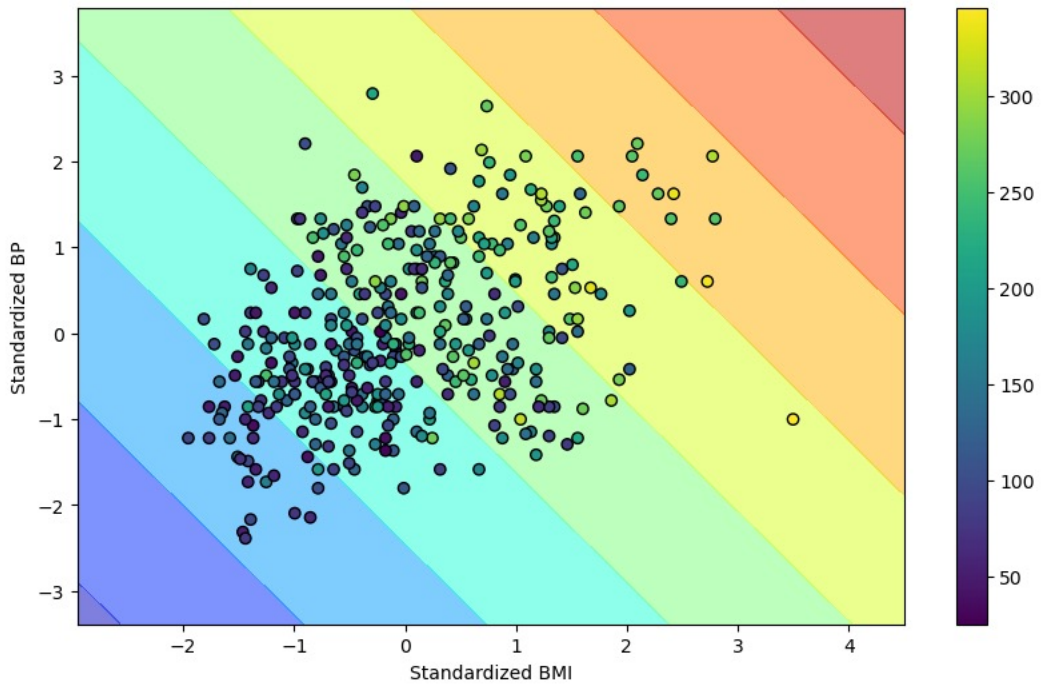


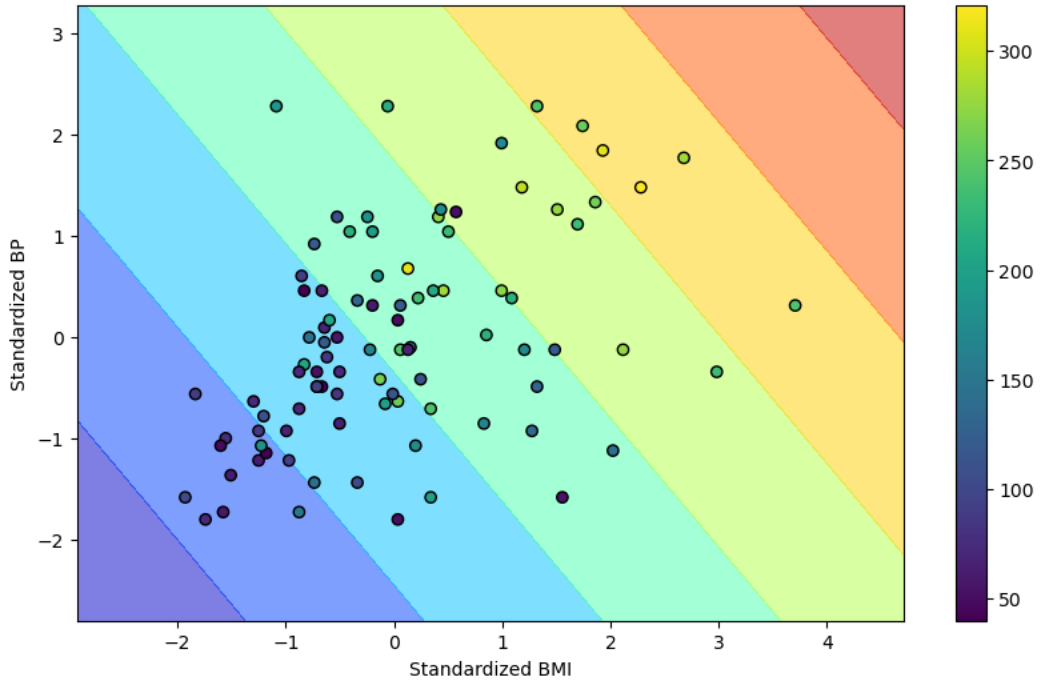




Underfitting

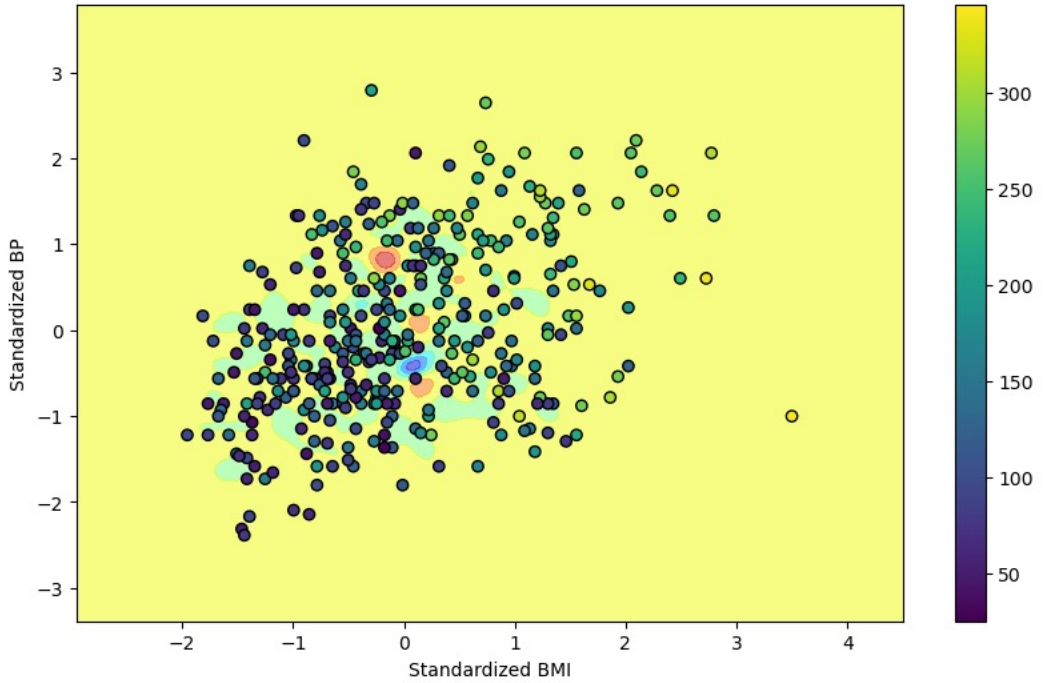
- Underfitting tritt auf, wenn ein Modell zu einfach ist und nicht genug von den Datenmuster erfasst, was zu schlechter Leistung führt.

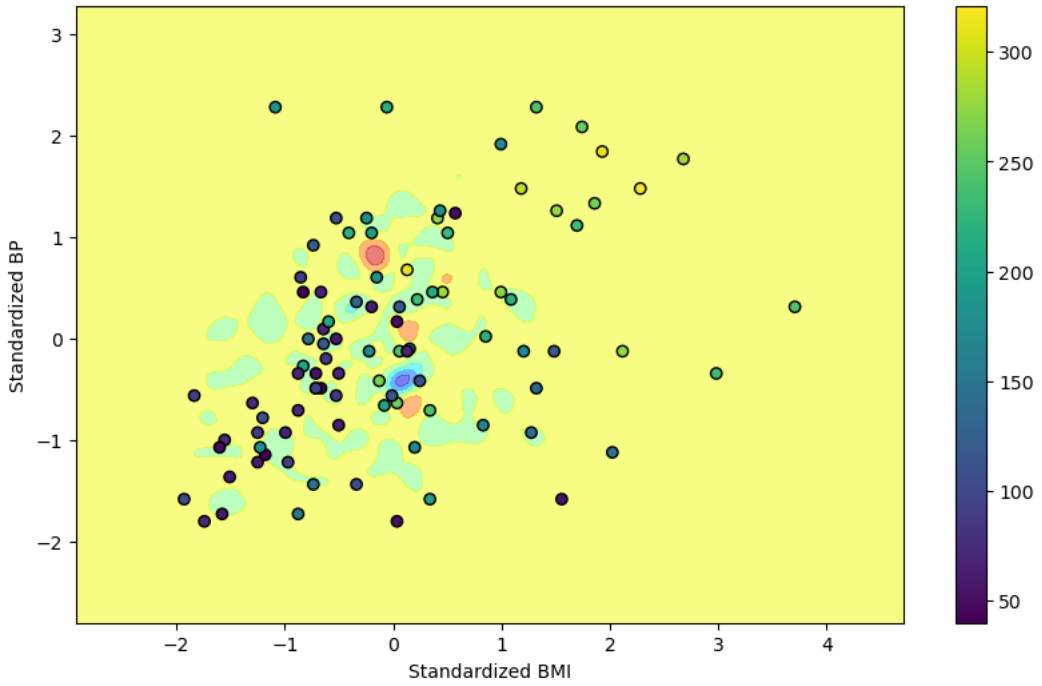




Overfitting

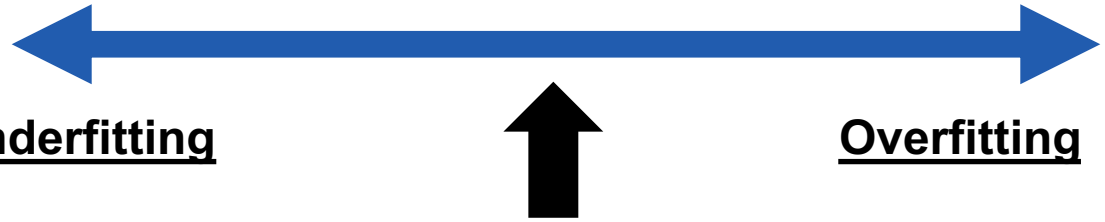
- Overfitting geschieht, wenn ein Modell zu komplex ist und zu sehr auf die Trainingsdaten zugeschnitten ist, wodurch es neue Daten schlecht generalisiert.





Auswahl des richtigen Modells

Fehlerhafte Schätzer



Schätzer, die die Daten auswendig lernen

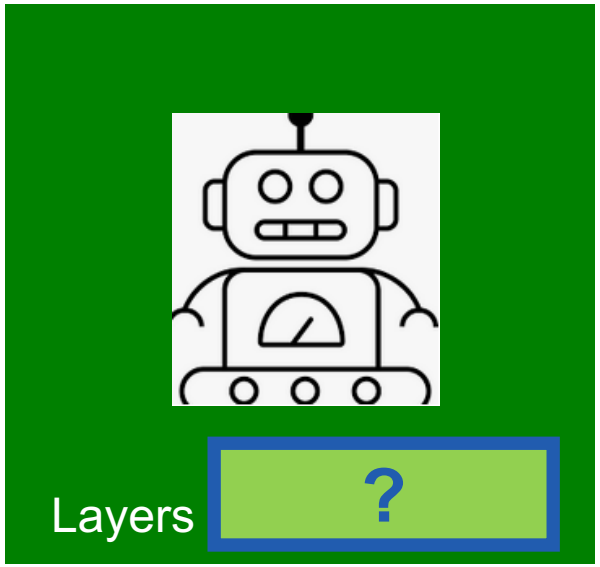
Underfitting

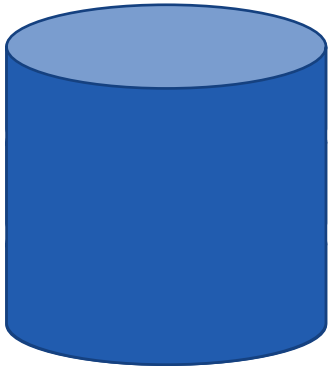
Overfitting

Schätzer, die die richtige Muster erkennen

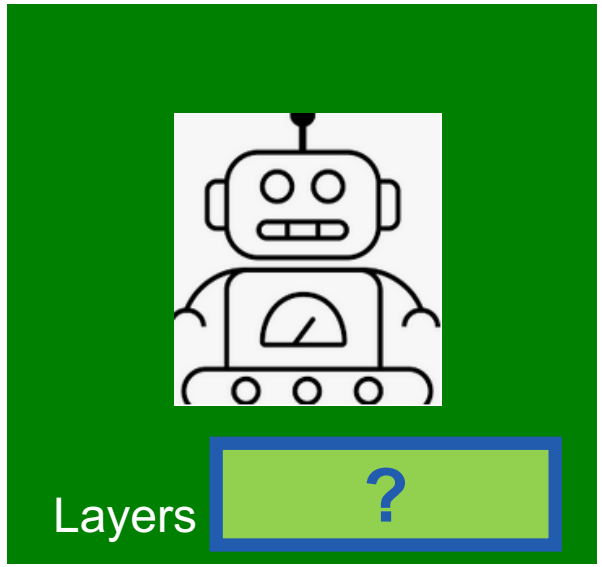
Modellauswahl mit Cross-Validation

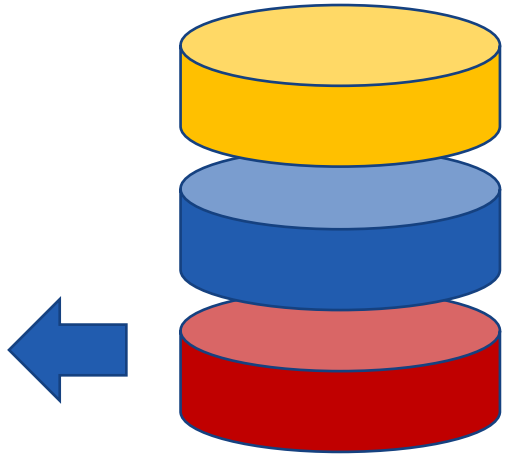
Grid search via K-fold cross-validation





Grid search via K-fold cross-validation





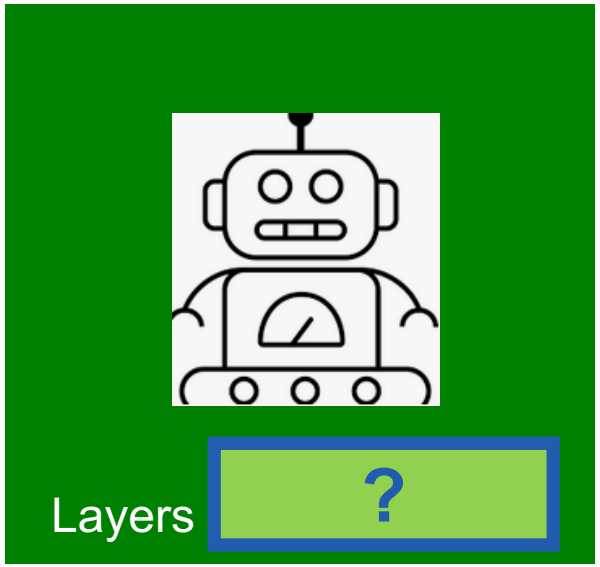
Training set

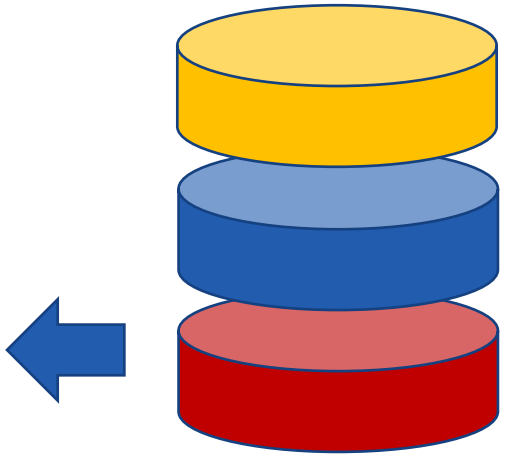
(2, 2) (8, 4) (32, 16, 8, 4, 2)

Validation set

% mistakes in 

Grid search via K-fold cross-validation





Training set

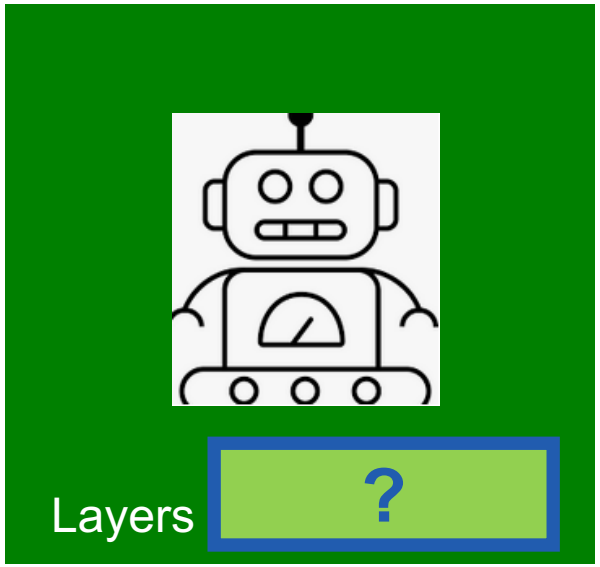
Validation set

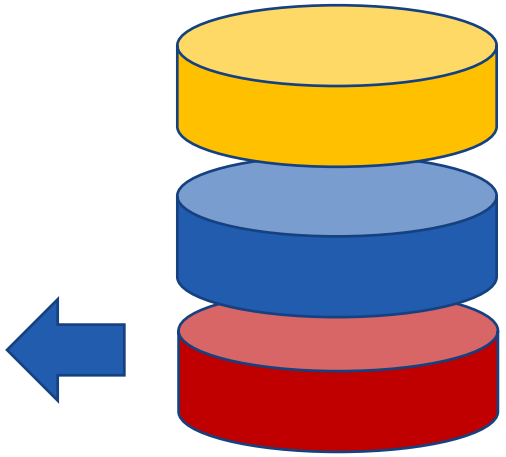
(2, 2)	(8, 4)	(32, 16, 8, 4, 2)
20	5	50

% mistakes in 

% mistakes in 

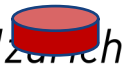
Grid search via K-fold cross-validation



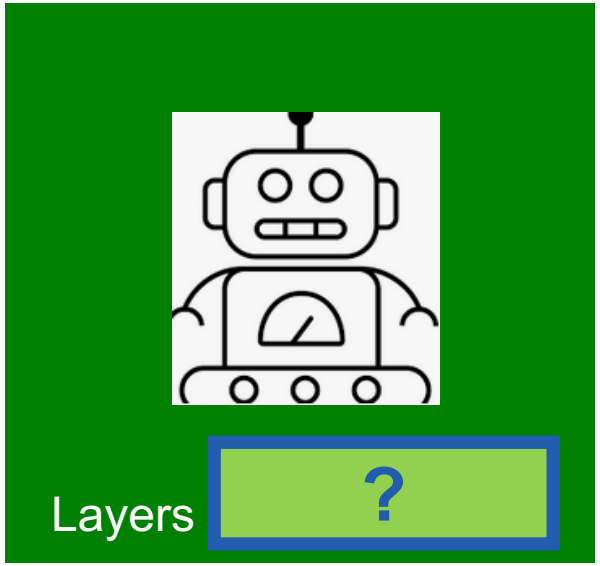


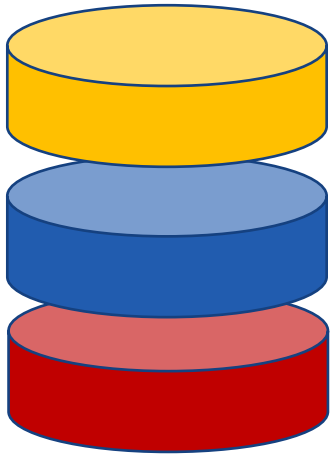
Training set

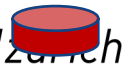
Validation set

	(2, 2)	(8, 4)	(32, 16, 8, 4, 2)
% mistakes in 	20	5	50
% mistakes in 	15	3	80
% mistakes in 			

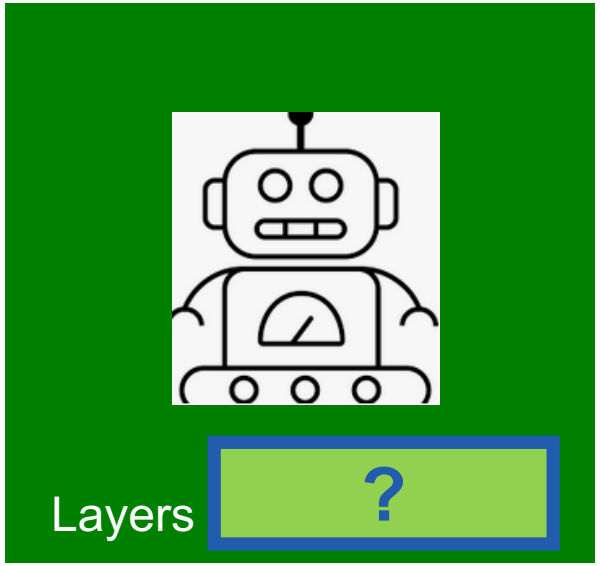
Grid search via K-fold cross-validation

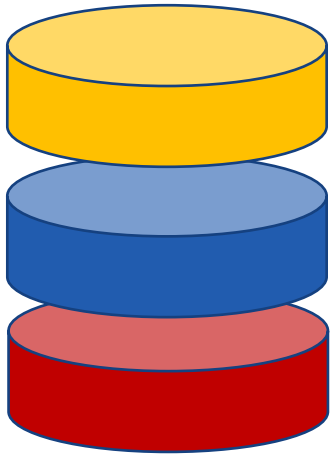


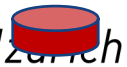


	(2, 2)	(8, 4)	(32, 16, 8, 4, 2)
% mistakes in 	20	5	50
% mistakes in 	15	3	80
% mistakes in 	13	4	70

Grid search via K-fold cross-validation





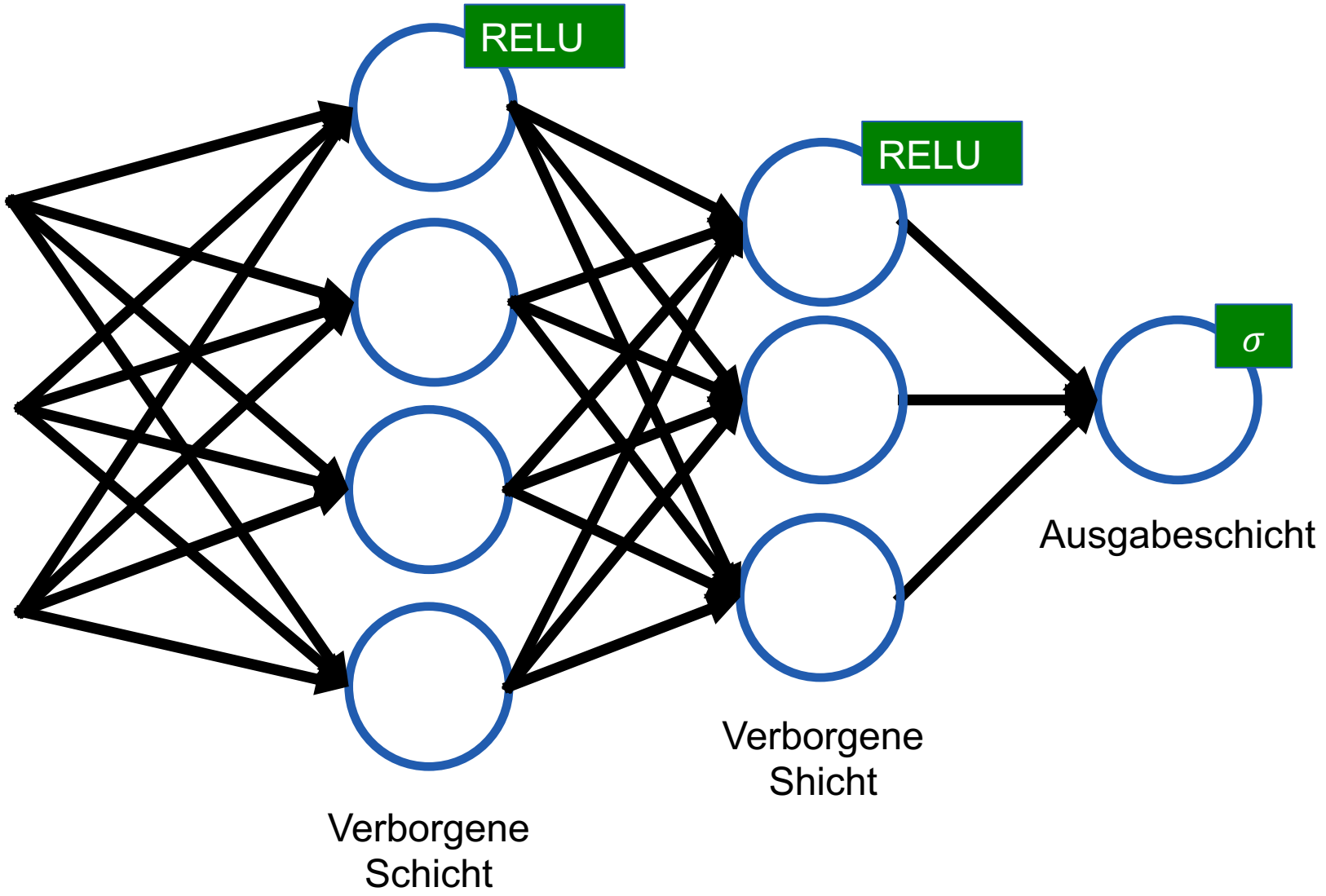
	(2, 2)	(8, 4)	(32, 16, 8, 4, 2)
% mistakes in 	20	5	50
% mistakes in 	15	3	80
% mistakes in 	13	4	70

Code Beispiel

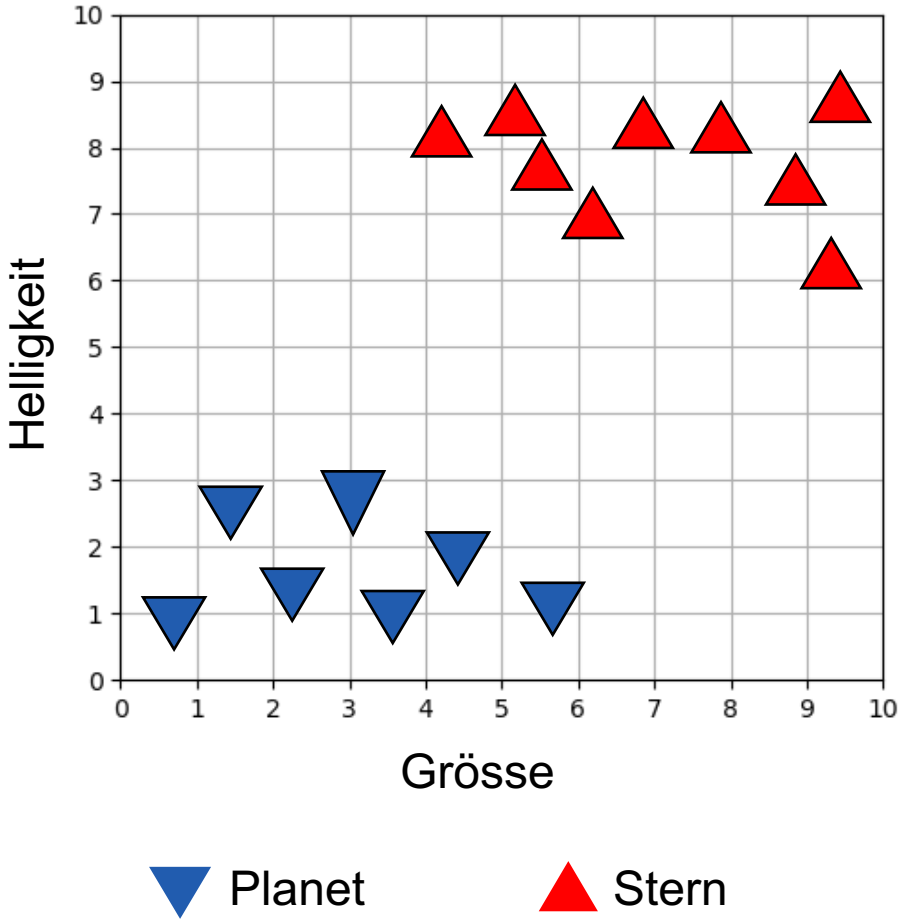
1. https://colab.research.google.com/drive/1gjlUMEHSw-ZSypZutha0fV8FM5R563Gx?usp=sharing

ETH zürich

Neuronale Netze


RELU
RELU
σ
Ausgabeschicht
Verborgene
Shicht
Verborgene
Schicht

Beispiel: Planeten und Sterne



Neuronale Netze

- Ein *Neuron* ist das Ergebnis einer Anwendung einer *Aktivierungsfunktion* zu einer Lineare Funktion.

Neuronale Netze

- Ein *Neuron* ist das Ergebnis einer Anwendung einer *Aktivierungsfunktion* zu einer Lineare Funktion.
- Beispiel: ReLU (Rectified linear unit)

$$ReLU(x)\begin{cases} x & falls\ x > 0 \\ 0 & andernfalls \end{cases}$$

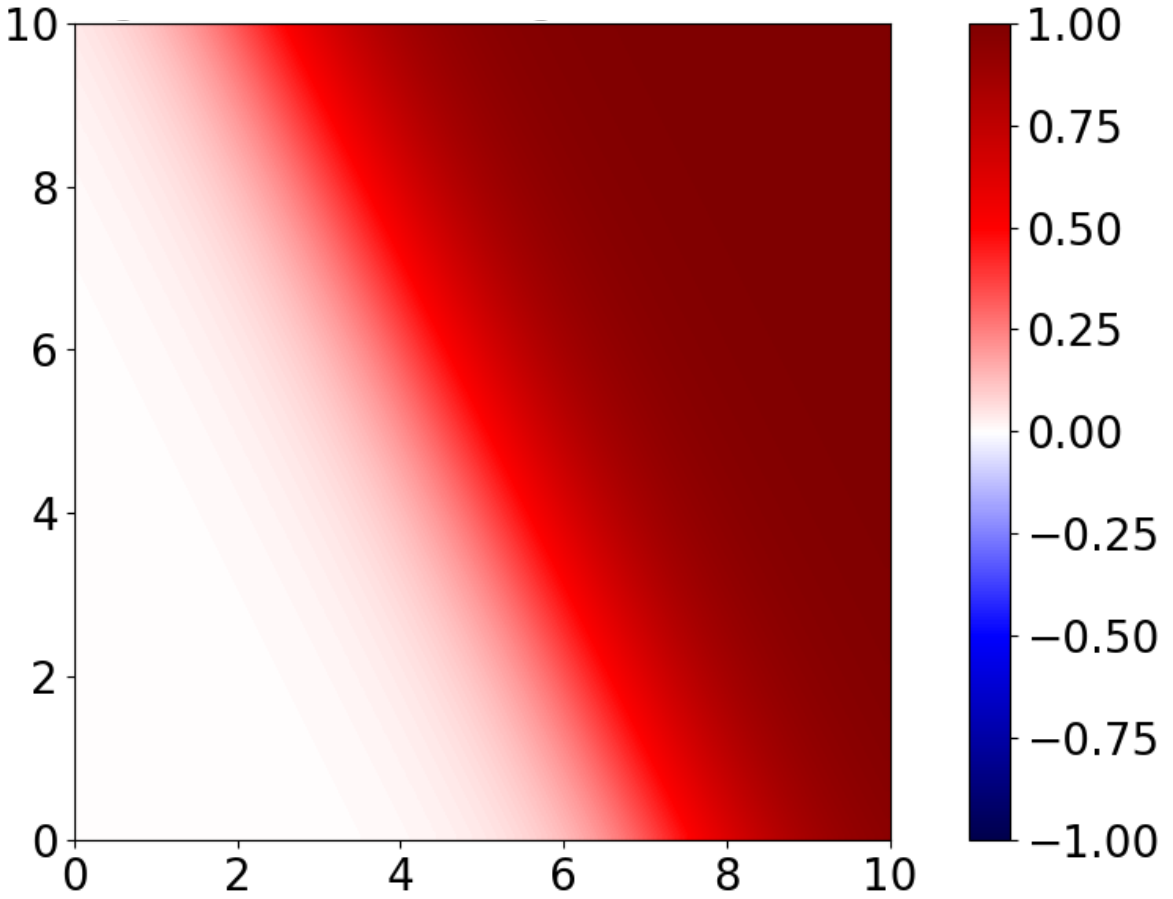
$$f(Grösse, Hell) = ReLU(-7 + Grösse + Hell)$$

- $f(3{,}2) = ?$
- $f(5{,}4) = ?$

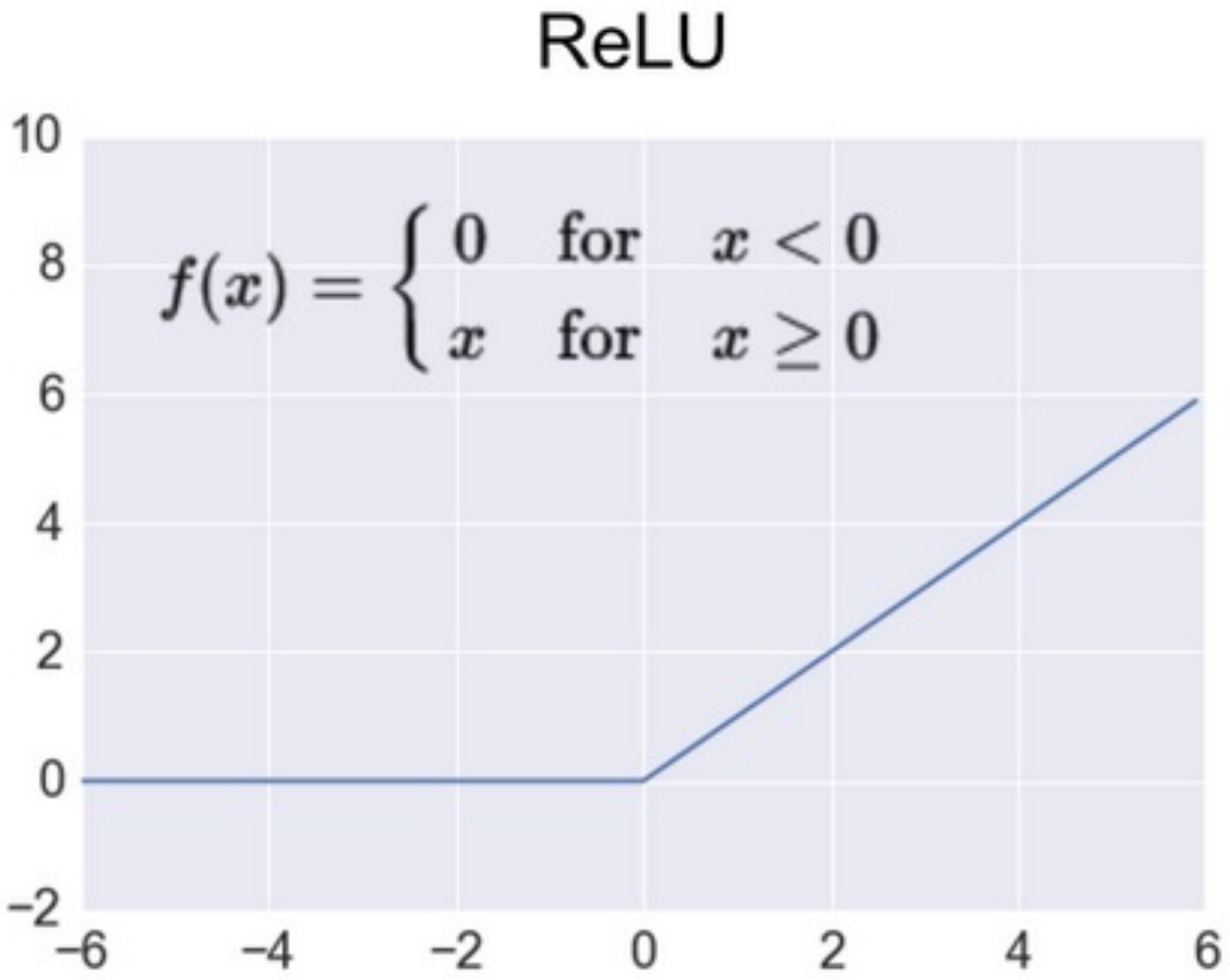
Neuronen

- Ein anderes Beispiel einer Aktivierungsfunktion: Sigmoid
- Example:

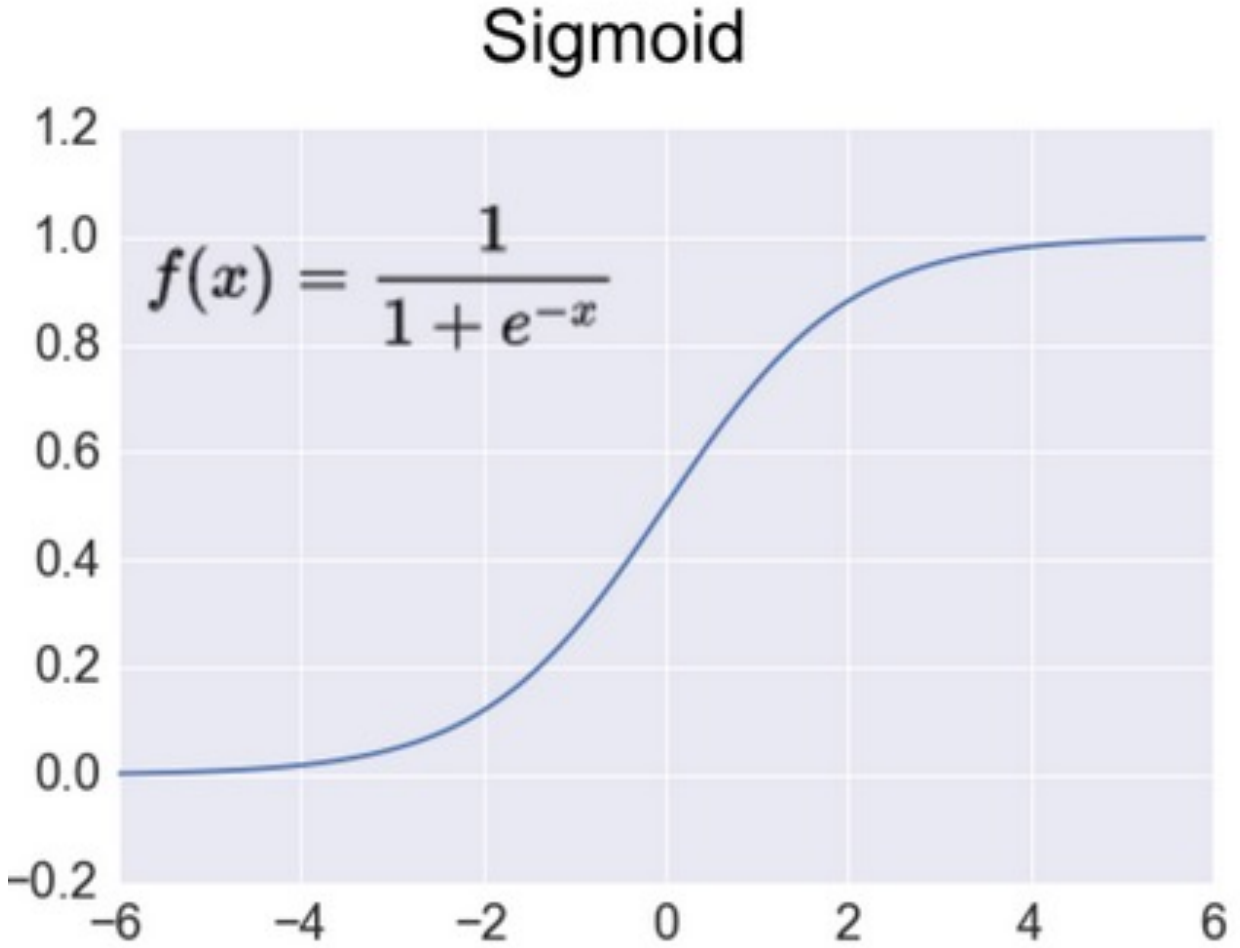
$$f(\,Grösse, Hell)$$
$$= \sigma(-10 + 2 * Grösse + 1 * Hell)$$



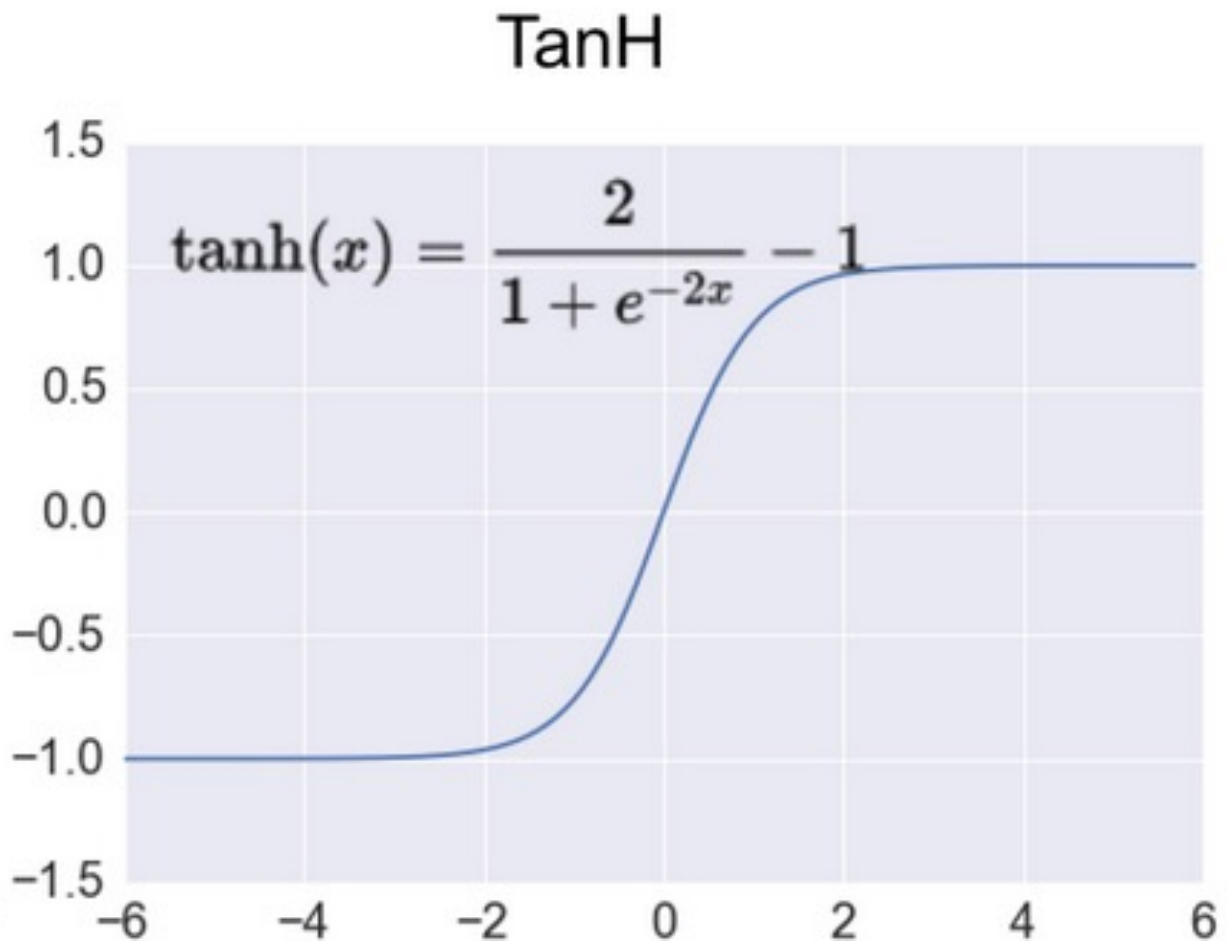
Häufig gebrauchte Aktivierungsfunktionen



Häufig gebrauchte Aktivierungsfunktionen



Häufig gebrauchte Aktivierungsfunktionen

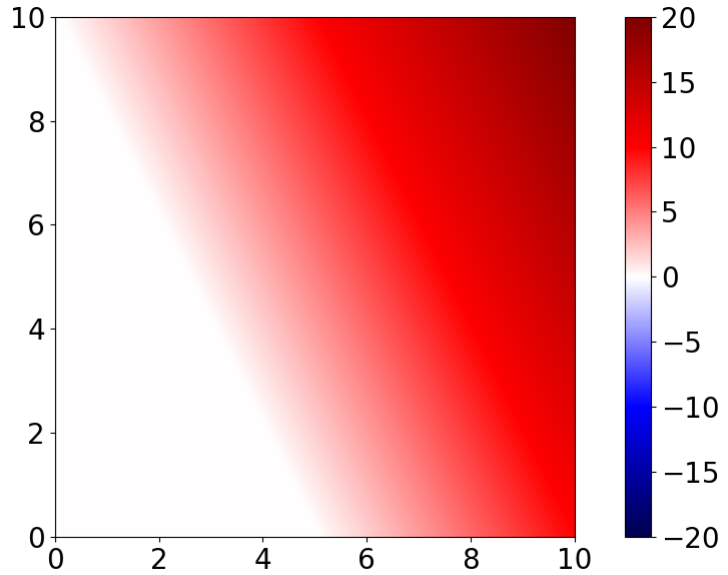


Darstellung von Neuronen

- Das Neuron

$$f(\,Grösse, Hell) = RELU(-10 + 2 * Grösse + 1 * Hell)$$

kann man wie folgt darstellen:

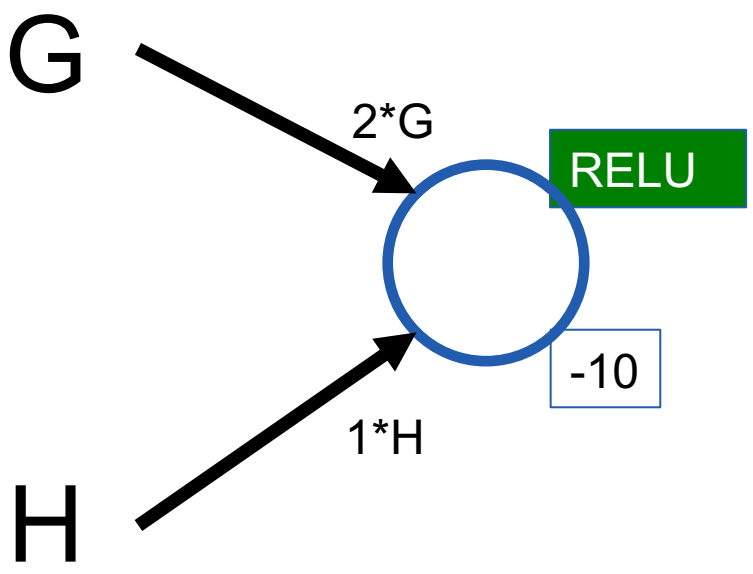


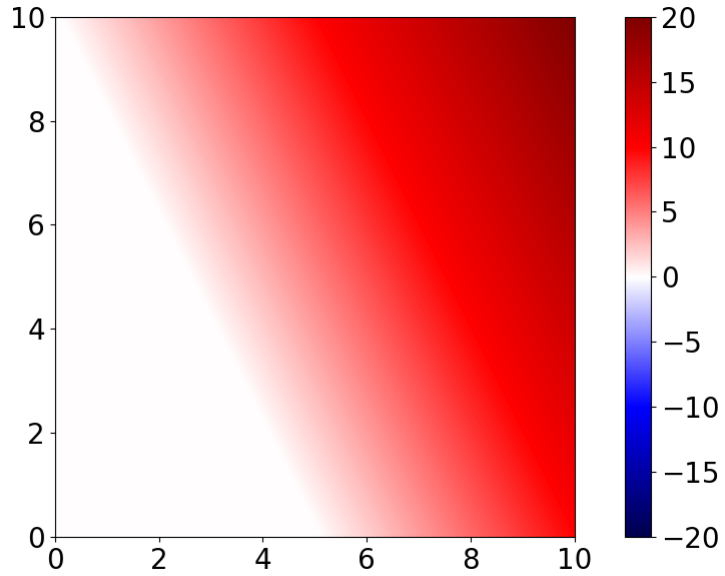
Darstellung von Neuronen

- Das Neuron

$$f(\,Grösse, Hell) = RELU(-10 + 2 * Grösse + 1 * Hell)$$

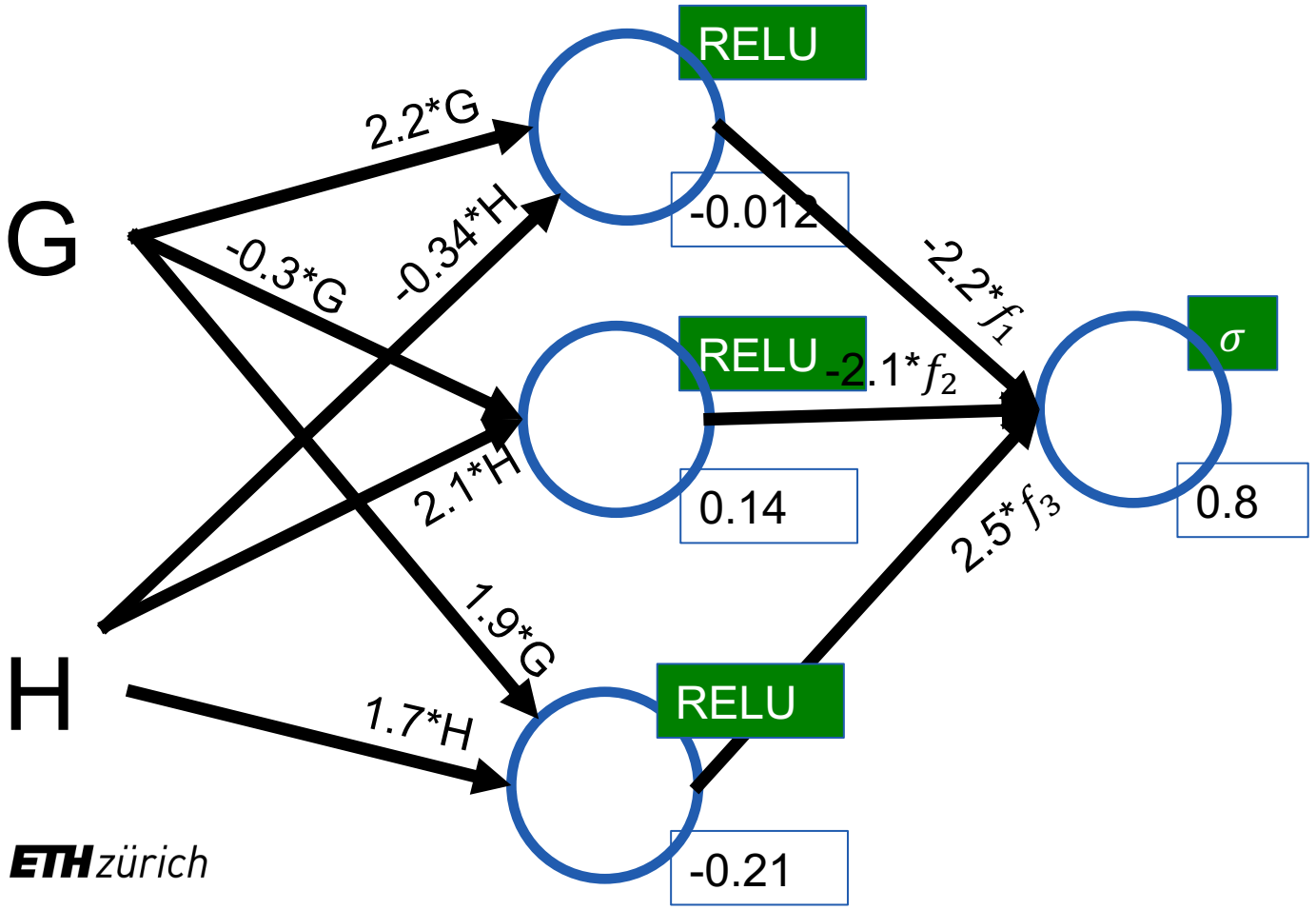
kann man wie folgt darstellen:

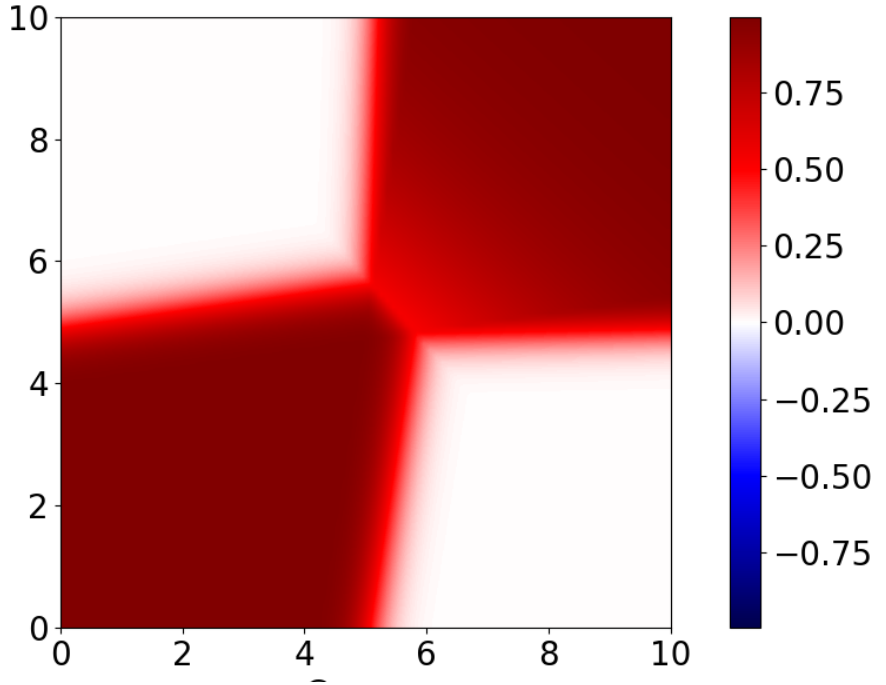




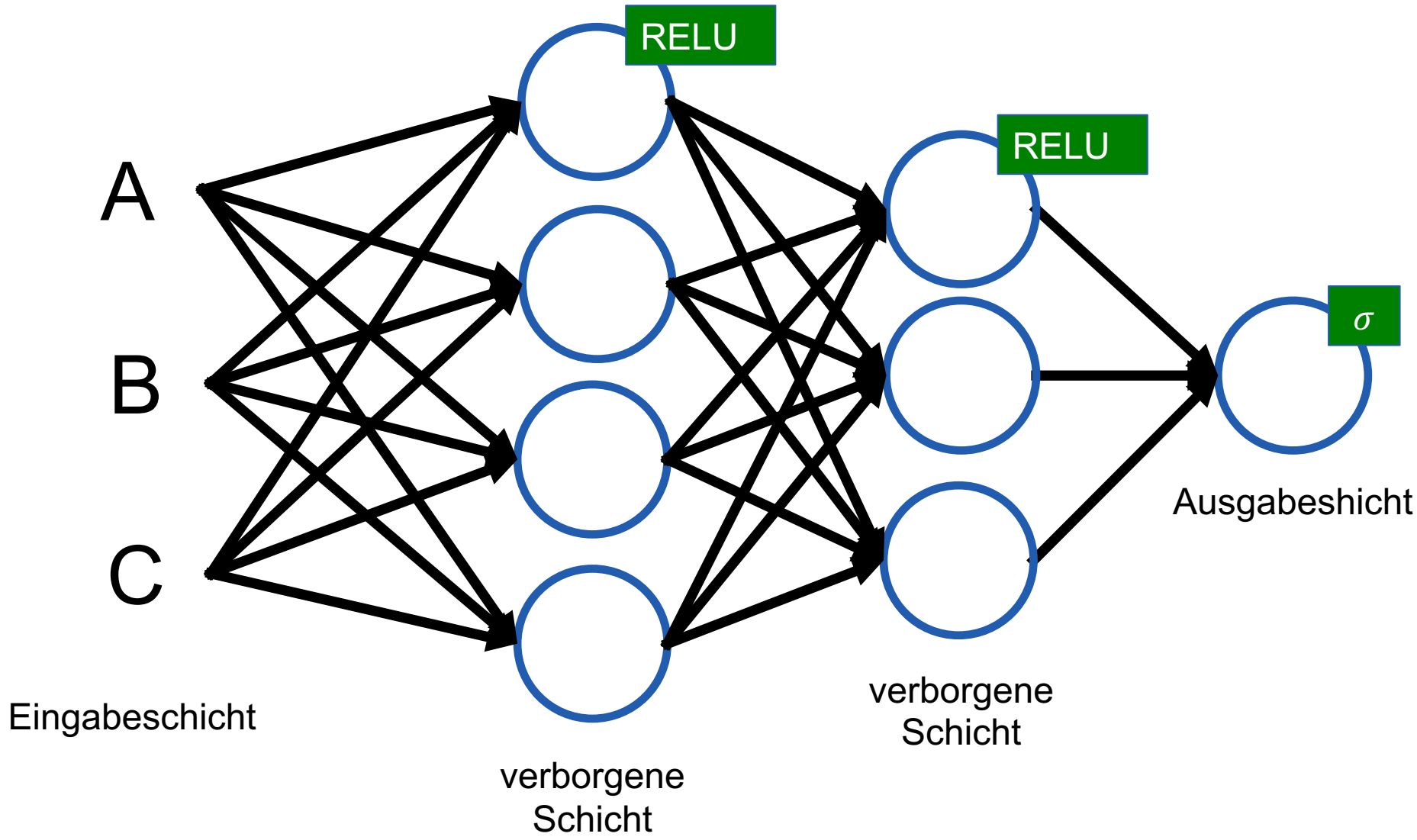
Unser erstes neuronale Netz

Unser erstes neuronale Netz





Die allgemeine Architektur eines neuronalen Netzes



Bemerkungen

- Beim Klassifizierung von Objekten in zwei Klassen, ist die Aktivierungsfunktiondes Neurons in Ausgabeschicht die Sigmoid. Das ist weil die Ausgabe zwischen 0 und 1 sein sollte.
- Bei Regression (vorhersage reeller Zahlen), gibt es keine Aktivierungsfunktion in Ausgabeschicht.
- Die Aktivierungsfunktion für die Neuronen in verborgenen Schichten ist normalerweise die ReLU.

ETH zürich

Warum sind neuronale Netze im Allgemein besser?

Warum sind neuronale Netze so beliebt?

Viele ML Algorithmen können keine neue Erkenntnisse gewinnen

Full house

Royal flush

- Logistische Regression

49%

Viele ML Algorithmen können keine neue Erkenntnisse gewinnen

Full house

Royal flush

Rank 1	Suit 1	…	Rank 5	Suit 5	Type
2	♦	…	8	♥	FH
10	♥	…	A	♥	RF

- Logistische Regression

49%

Viele ML Algorithmen können keine neue Erkenntnisse gewinnen

Full house

Royal flush

- Logistische Regression

99.5%

Rank 1	Suit 1	…	Rank 5	Suit 5	# A	#2	…	#♠	#♥	Type
2	♦	…	8	♥	0	3		1	2	FH
10	♥	…	A	♥	1	0		0	5	RF

Neuronale Netze können neue Erkenntnisse gewinnen!

Full house

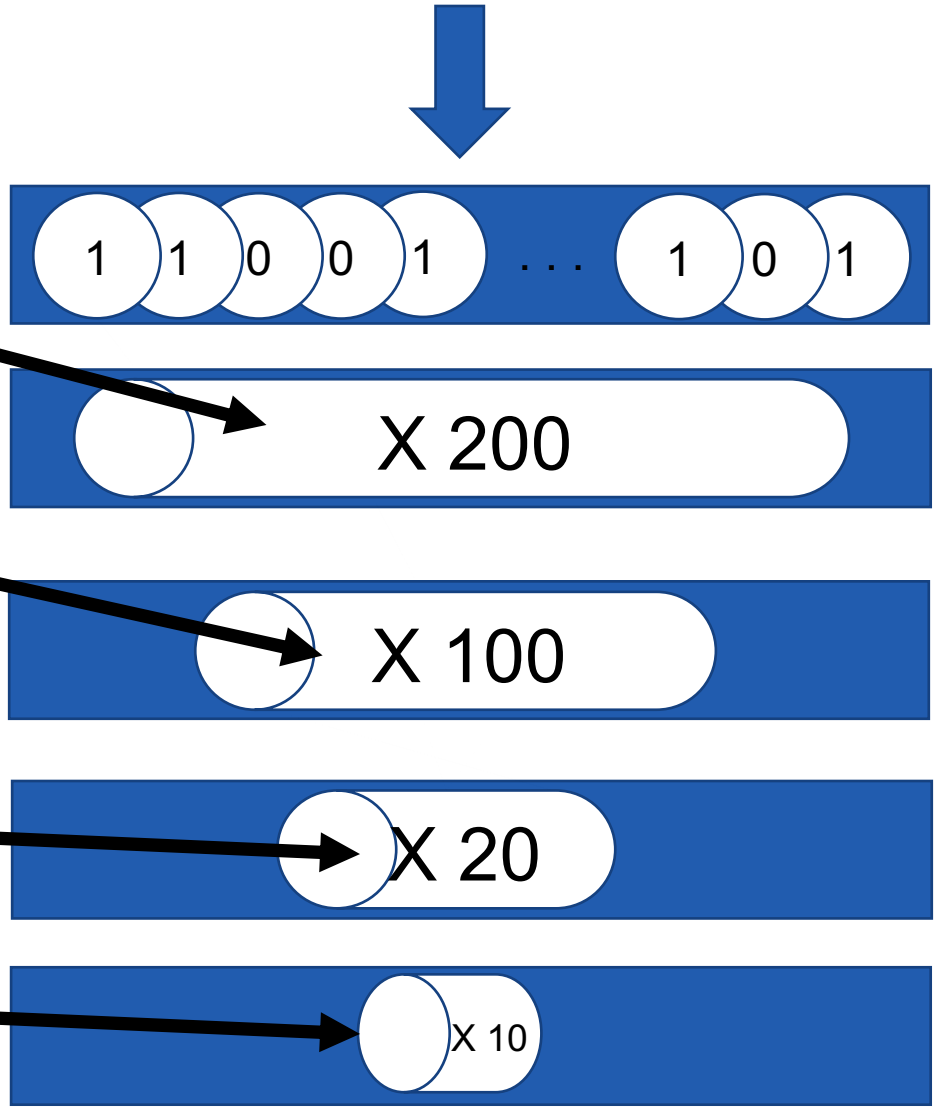
Royal flush

Rank 1	Suit 1	…	Rank 5	Suit 5	Type
2	♦	…	8	♥	FH
10	♥	…	A	♥	RF

- Neuronale Netze

99%

Handerkennung



Gibts mehr als einen König?

Sind alle Karte der selben Farbe?

Bilden die Karten eine Reihe?

Jedes Neuron in der Ausgabeschicht erkennt eine Kombination

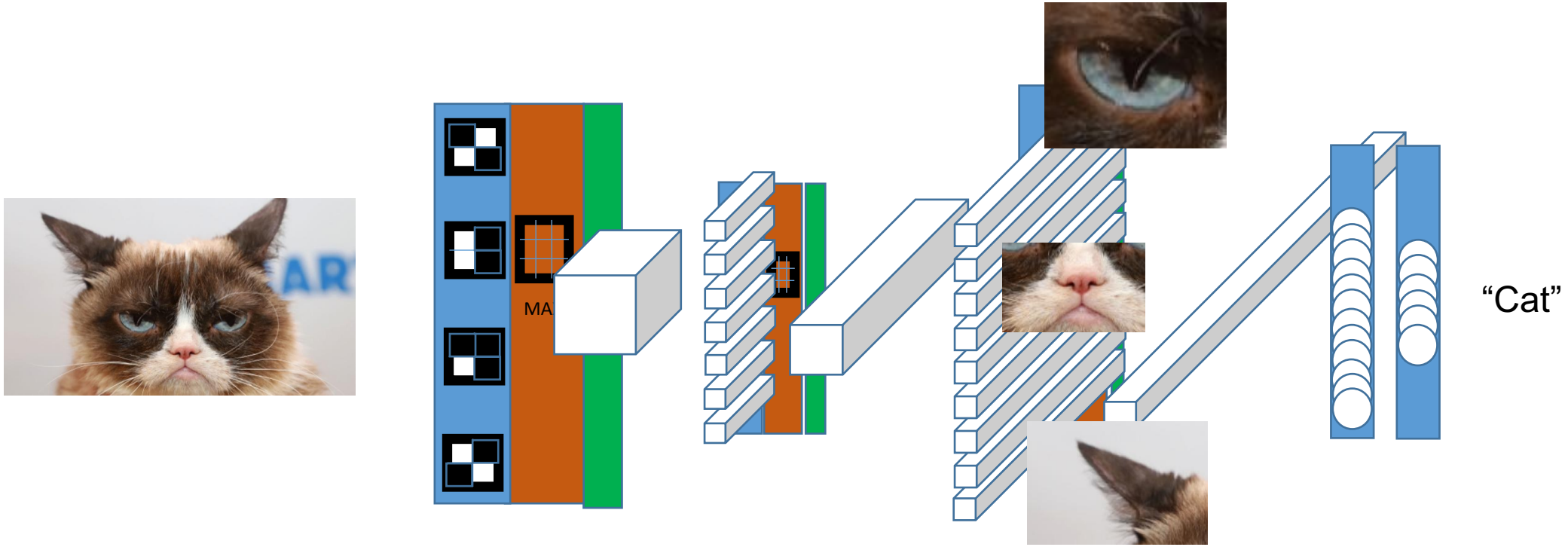
Interaktive Webseite

- http://playground.tensorflow.org/

ETH zürich

Convolutional neural networks

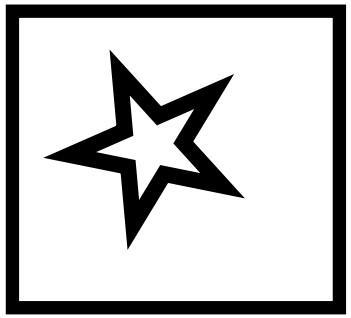
Bilderkennung mit CNNs



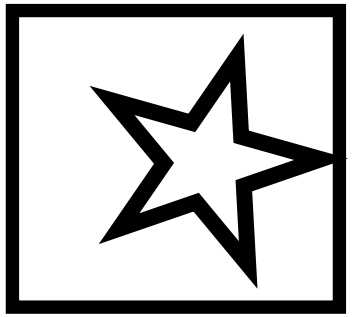
Convolutional neural networks

- Bilderkennung mit allgemeinen Neuronalen Netzen würde eine Unmenge von Neuronen benötigen.
- Für Bildverarbeitung es ist besser CNNs zu benutzen.

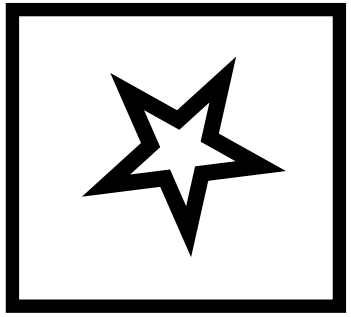
Bilderkennung



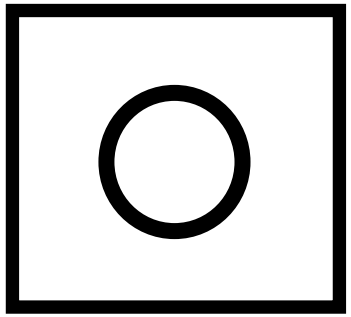
Stern



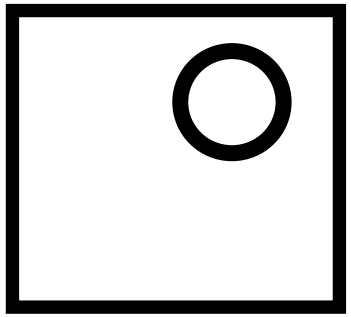
Stern



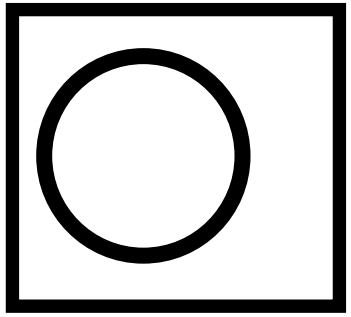
Stern



Planet

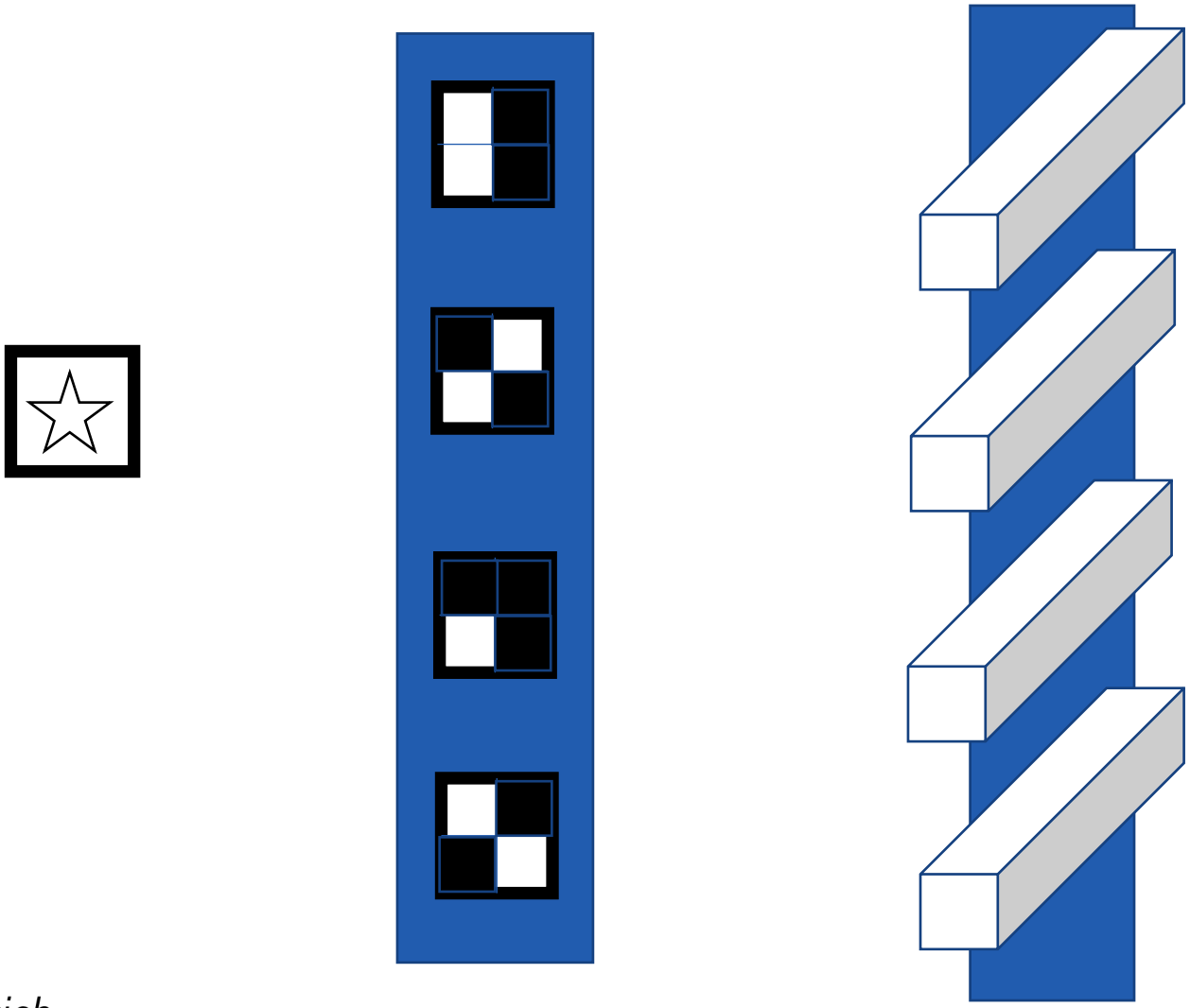


Planet

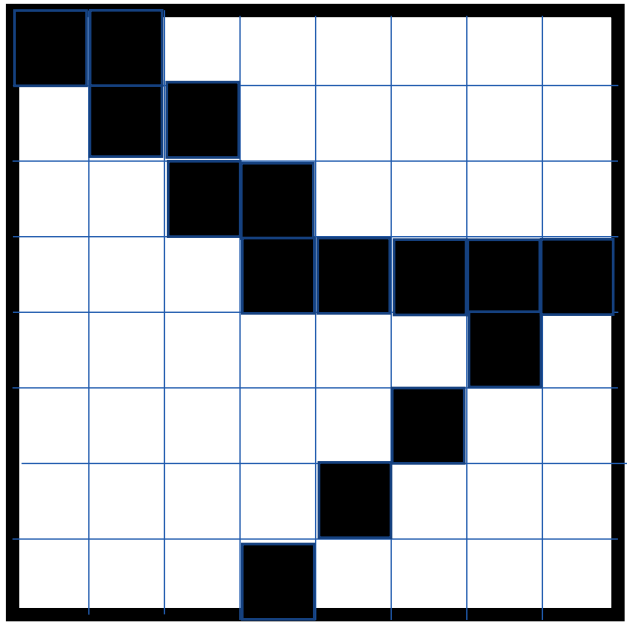


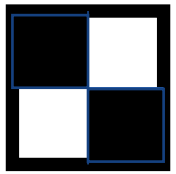
Planet

Convolutional neural networks

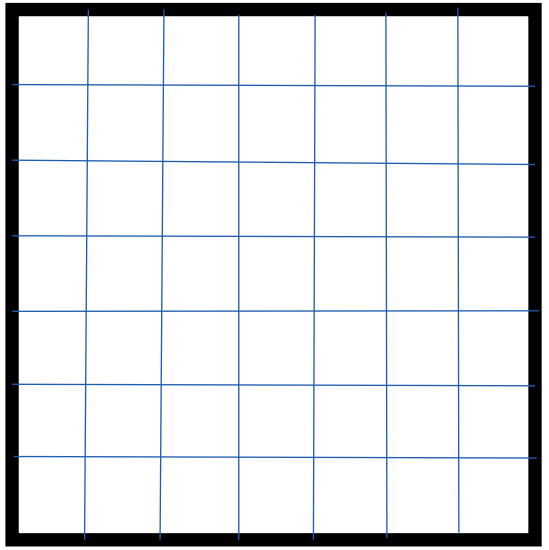


Convolutional neural networks

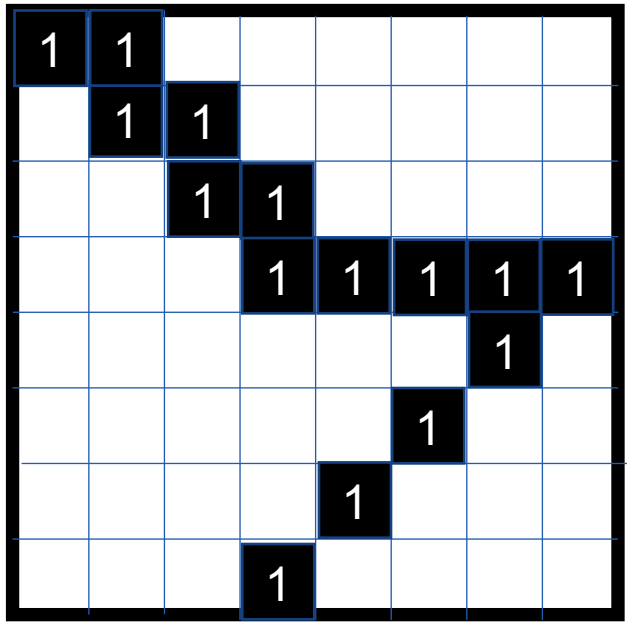


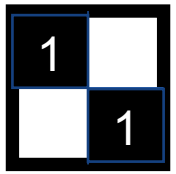


Faltungsmatrix

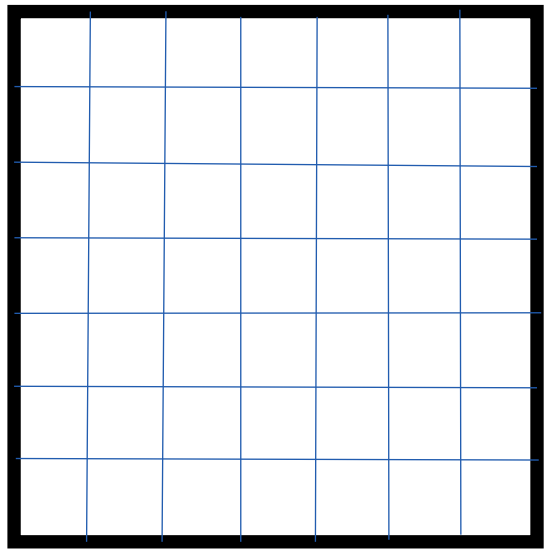


Convolutional neural networks

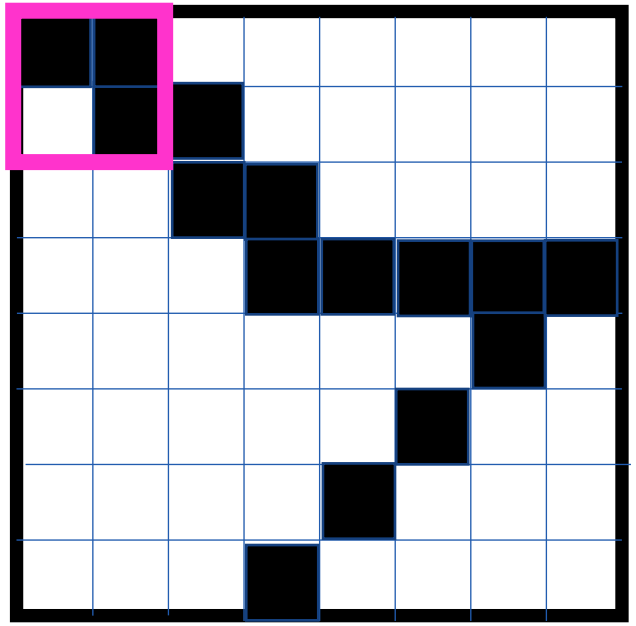


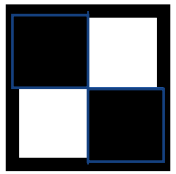


Faltungsmatrix

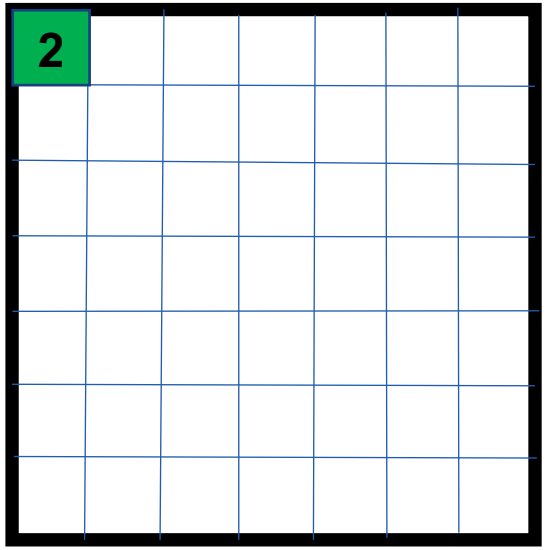


Convolutional neural networks

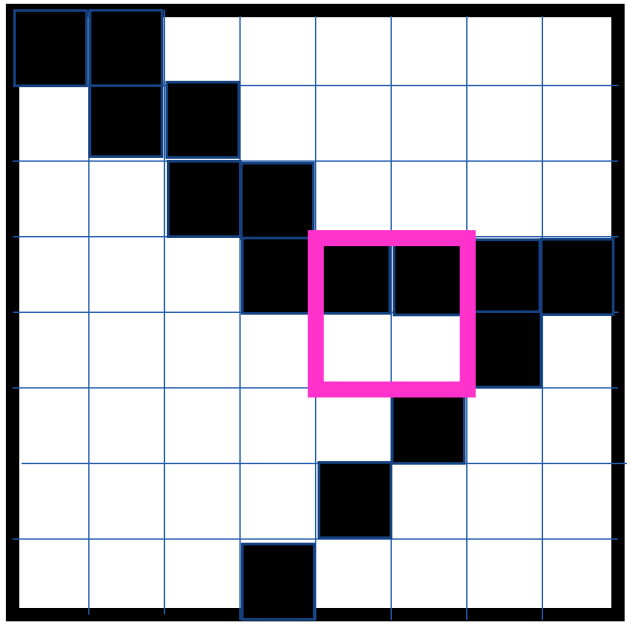


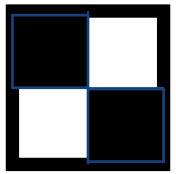


Faltungsmatrix

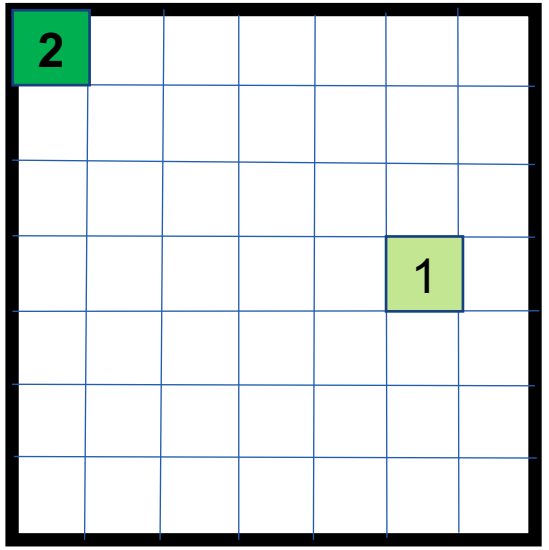


Convolutional neural networks

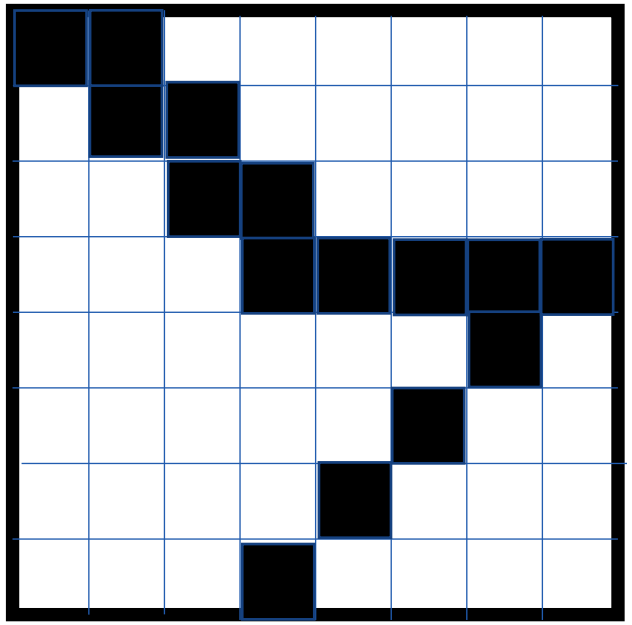


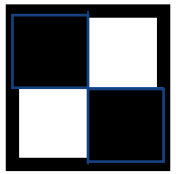


Faltungsmatrix

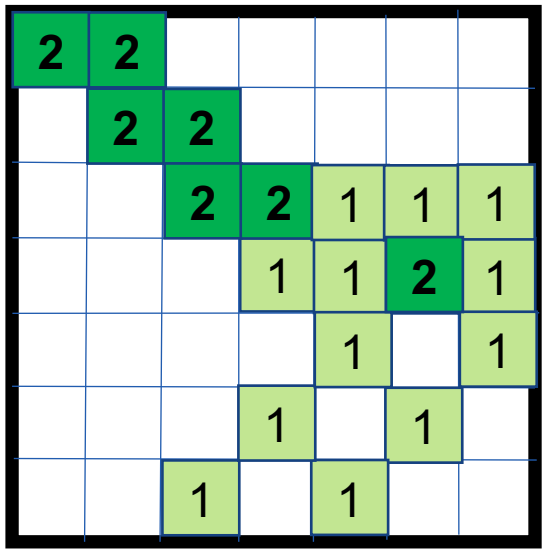


Convolutional neural networks

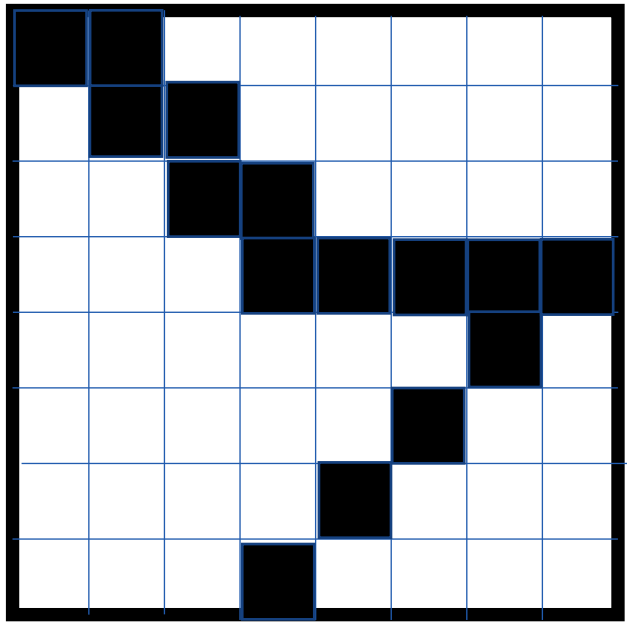


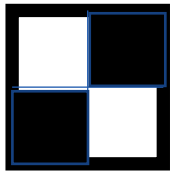


Faltungsmatrix

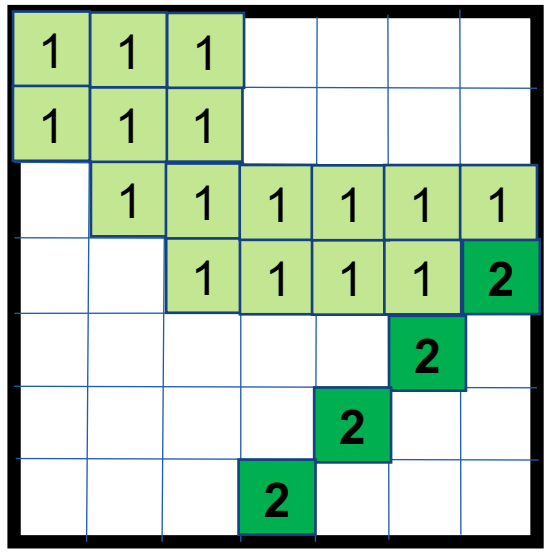


Convolutional neural networks





Faltungsmatrix



Filters

1	3	5
1	2	4
6	3	2

Eingabe

1	2
-5	0

Faltungsmatrix

ReLU

Ausgabe

Filters

1	3	5
1	2	4
6	3	2

Eingabe

1	2
-5	0

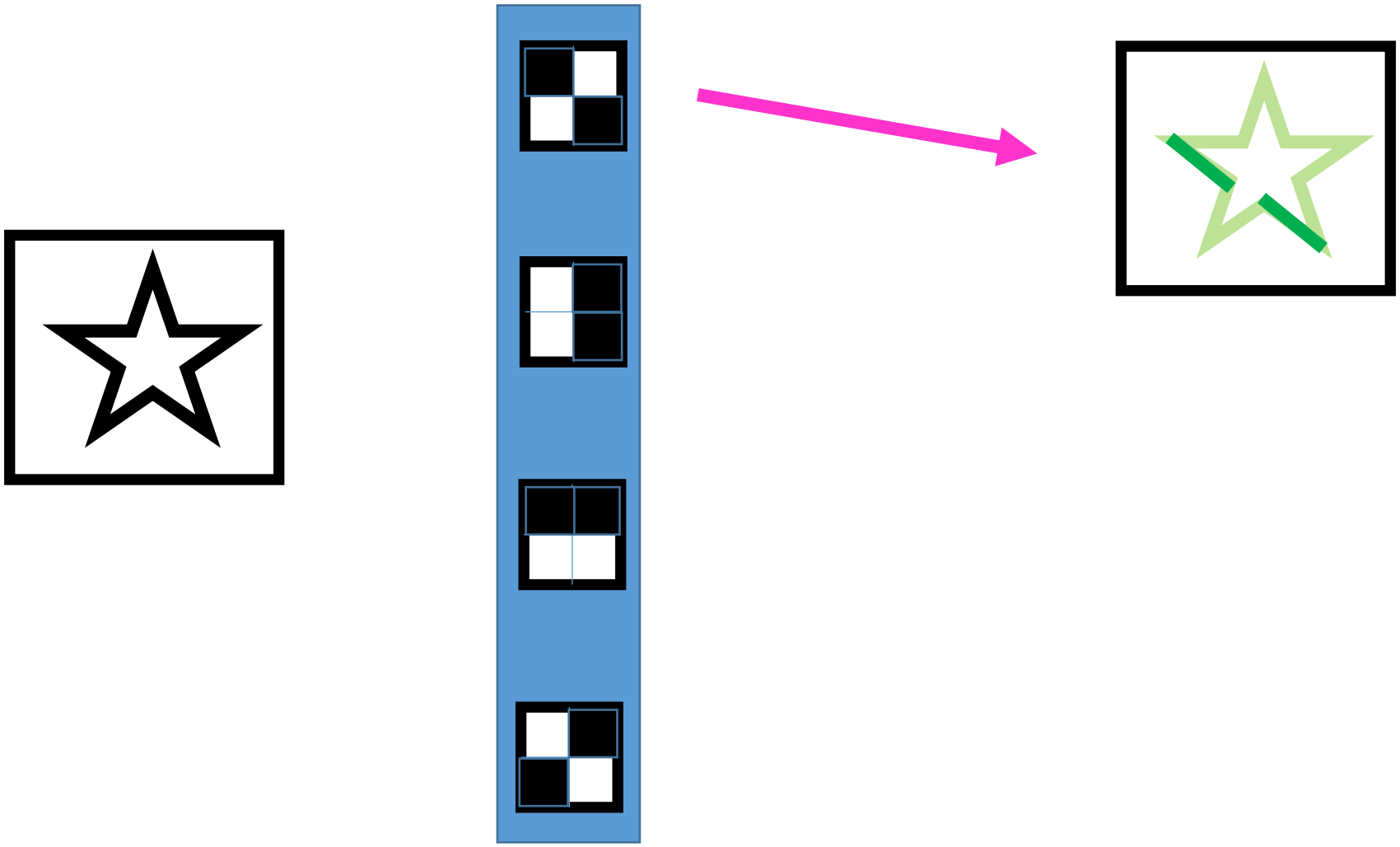
ReLU

Faltungsmatrix

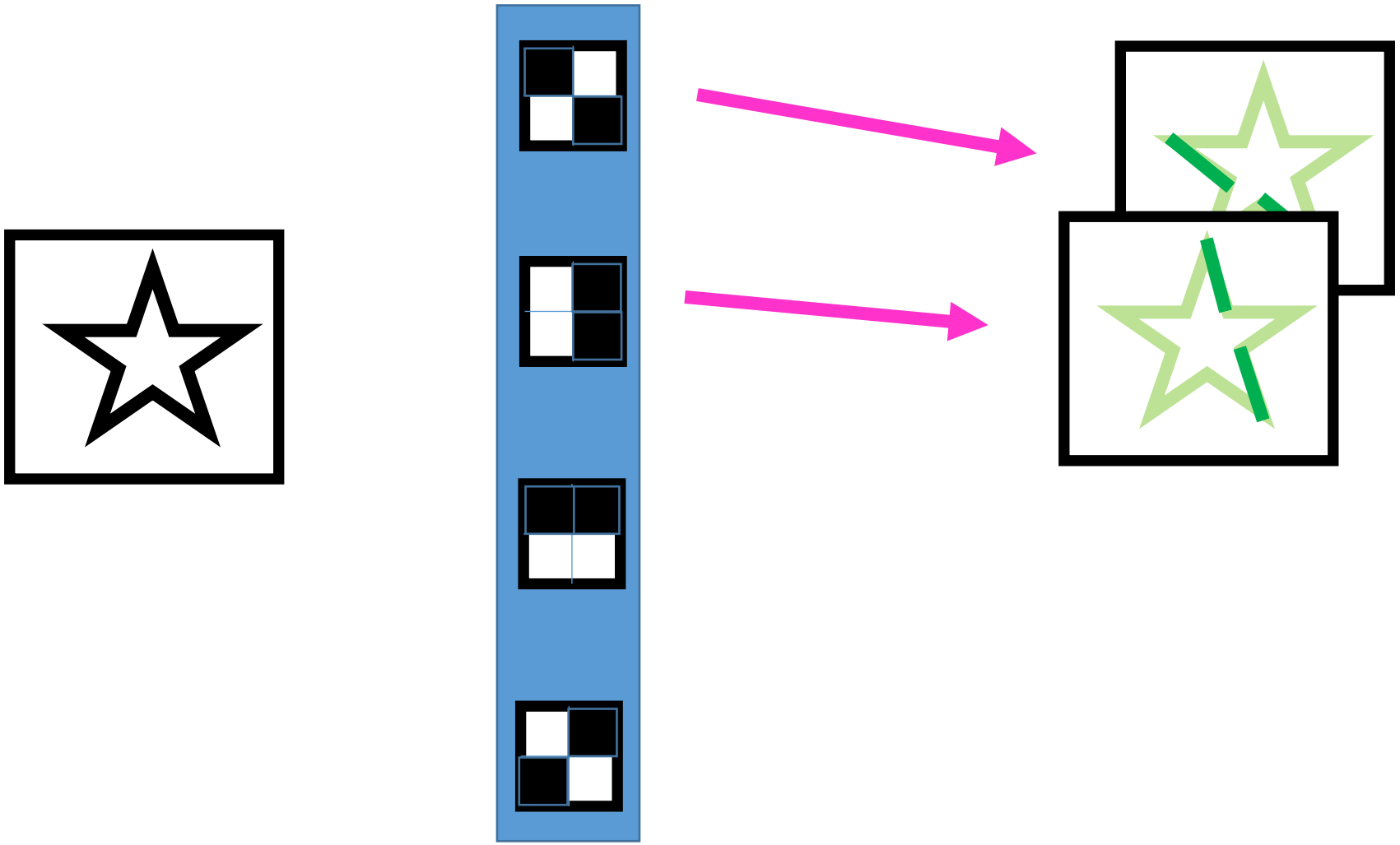
2	3
0	0

Ausgabe

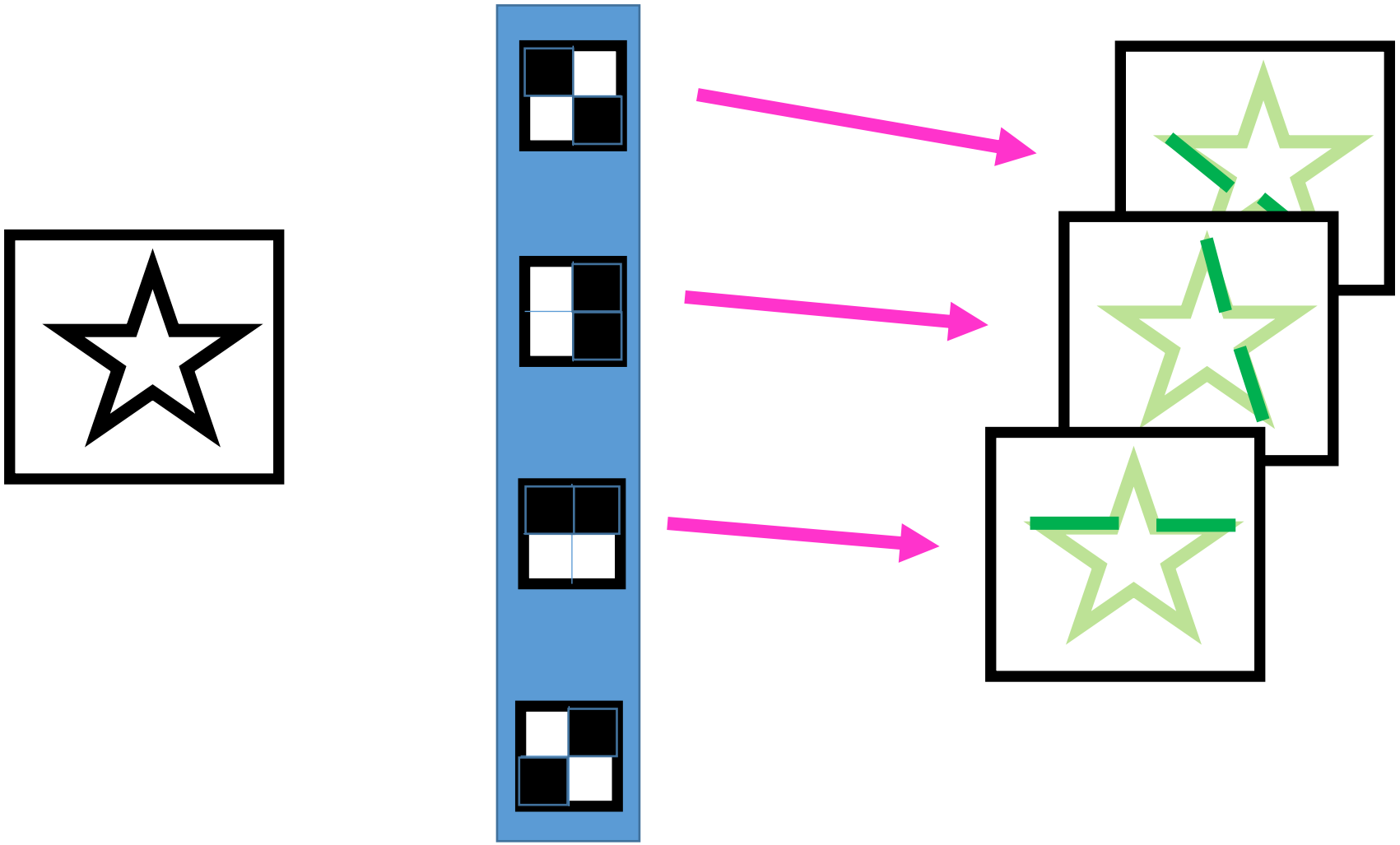
Convolutional neural networks



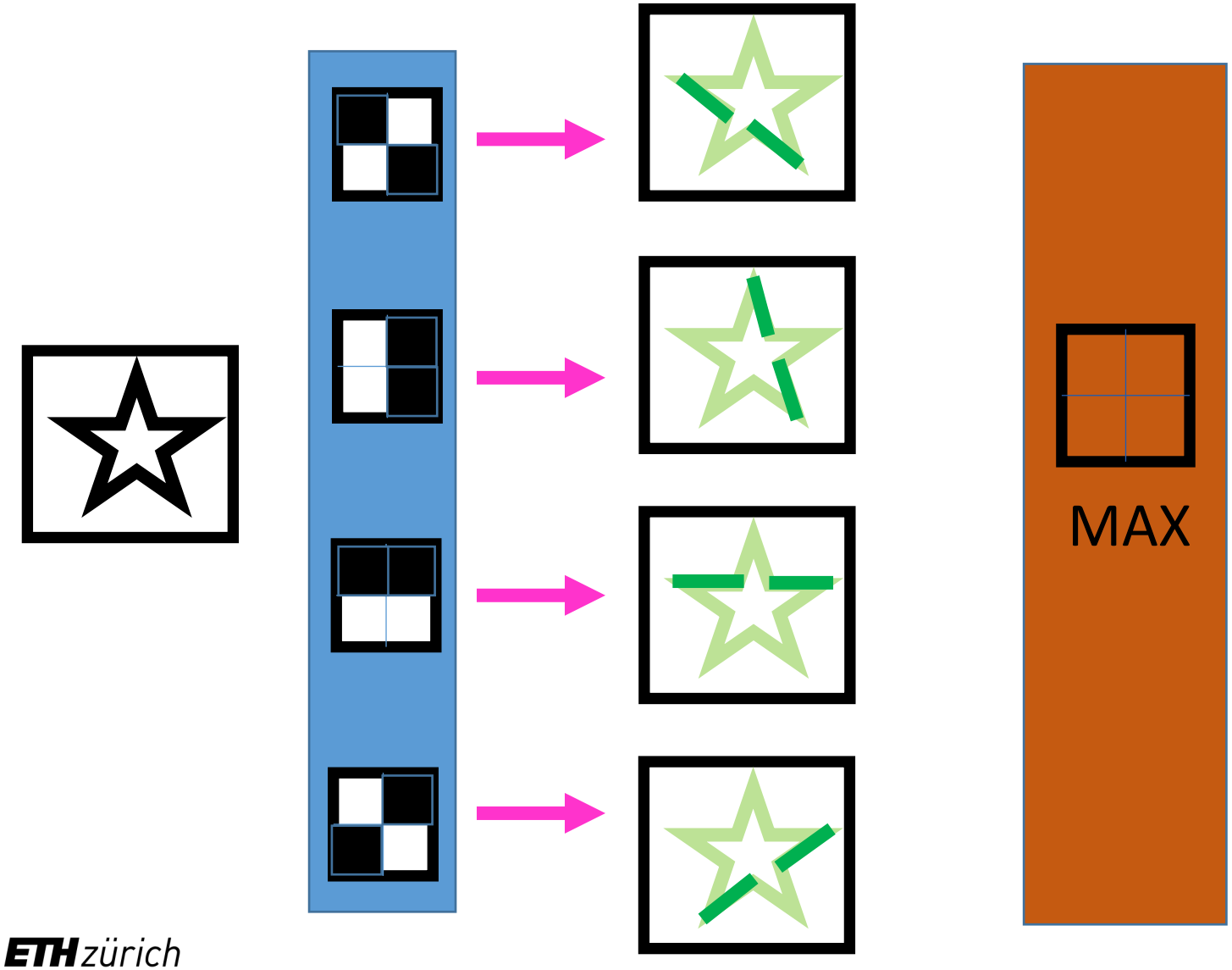
Convolutional neural networks



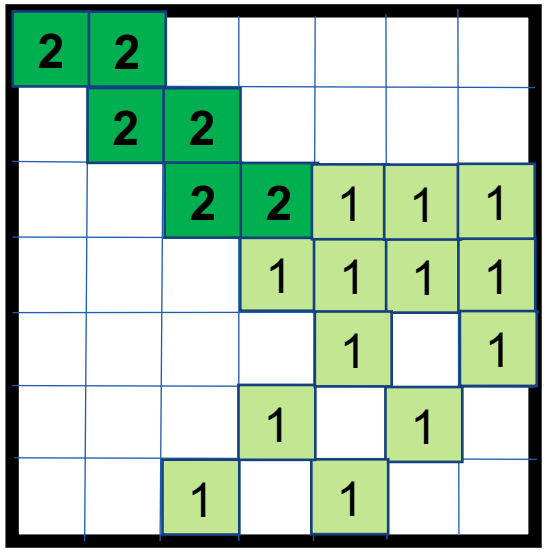
Convolutional neural networks

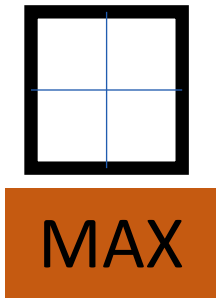


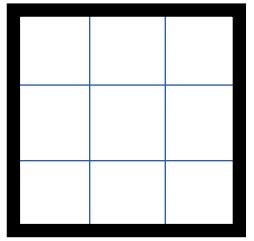
Convolutional neural networks



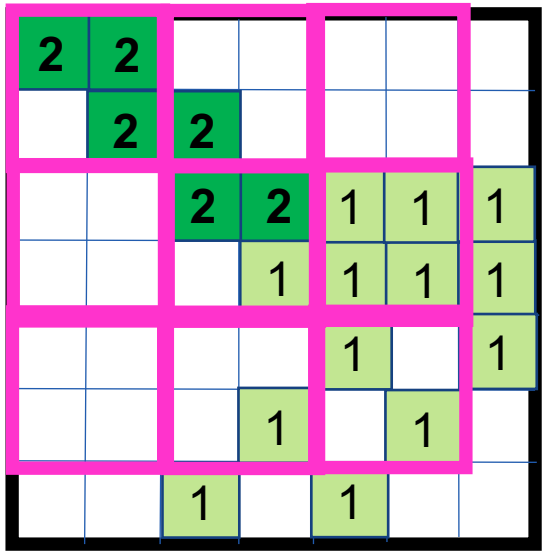
Convolutional neural networks

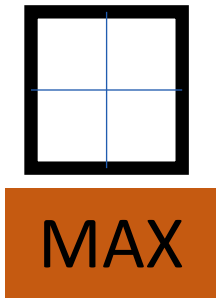


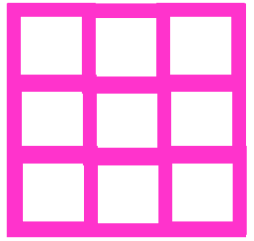




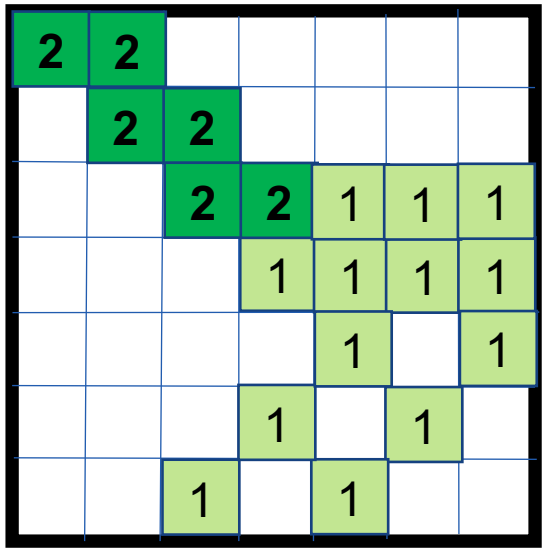
Convolutional neural networks

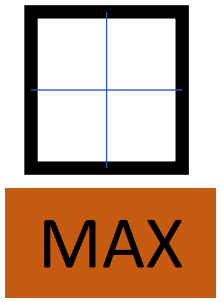


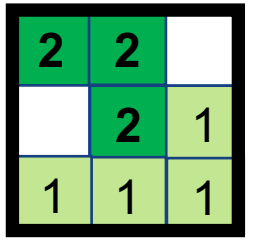




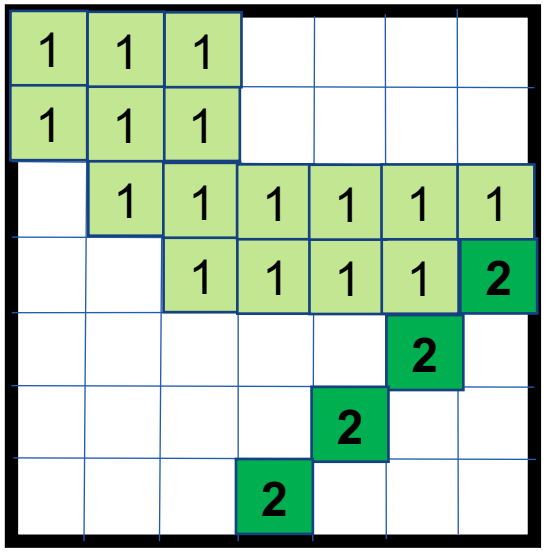
Convolutional neural networks

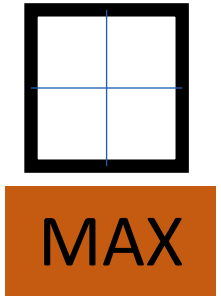


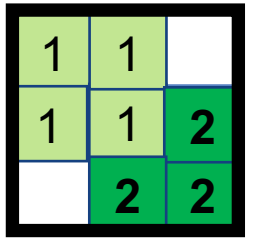




Convolutional neural networks



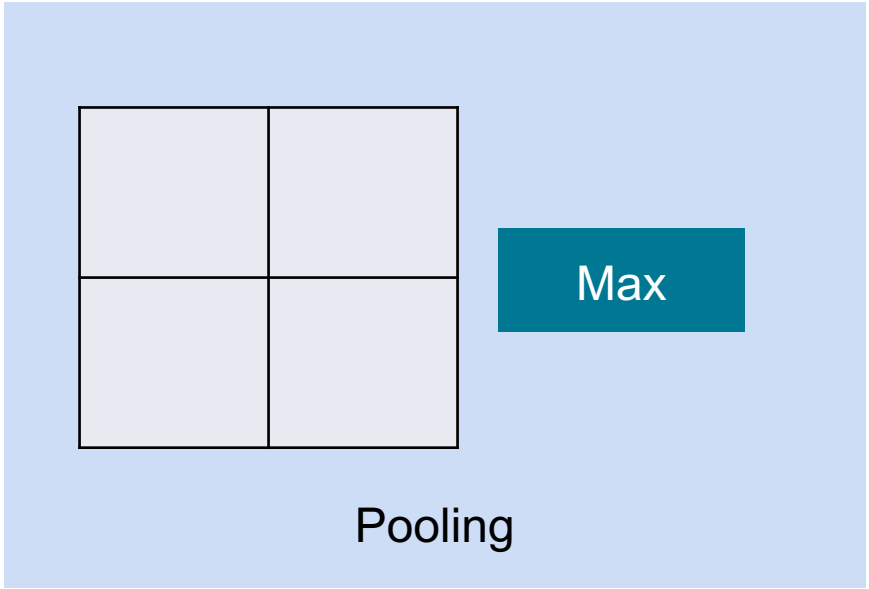




Pooling

1	0	5	3	1
4	0	0	2	2
6	1	0	1	0
0	0	0	2	4
3	0	1	2	1

Gefiltertes Bild

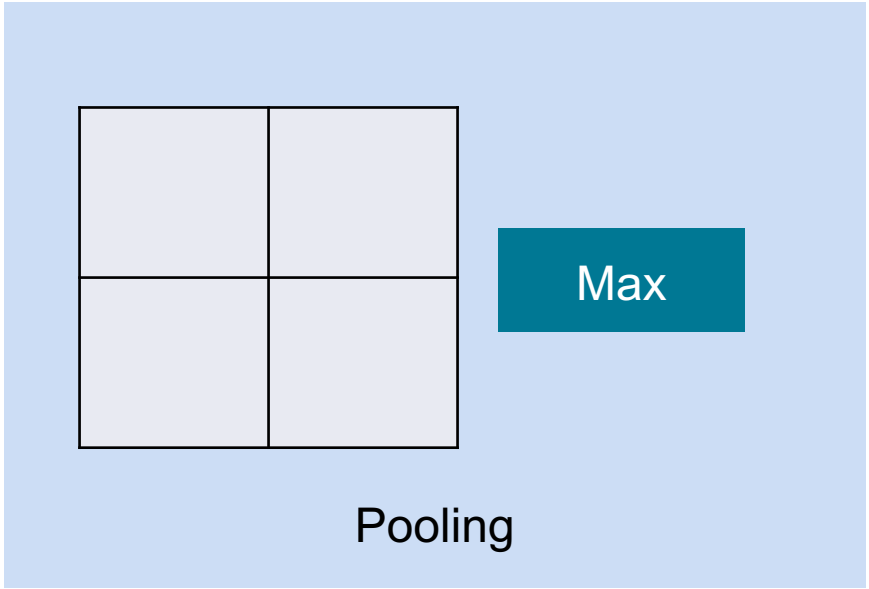


Ausgabe

Pooling

1	0	5	3	1
4	0	0	2	2
6	1	0	1	0
0	0	0	2	4
3	0	1	2	1

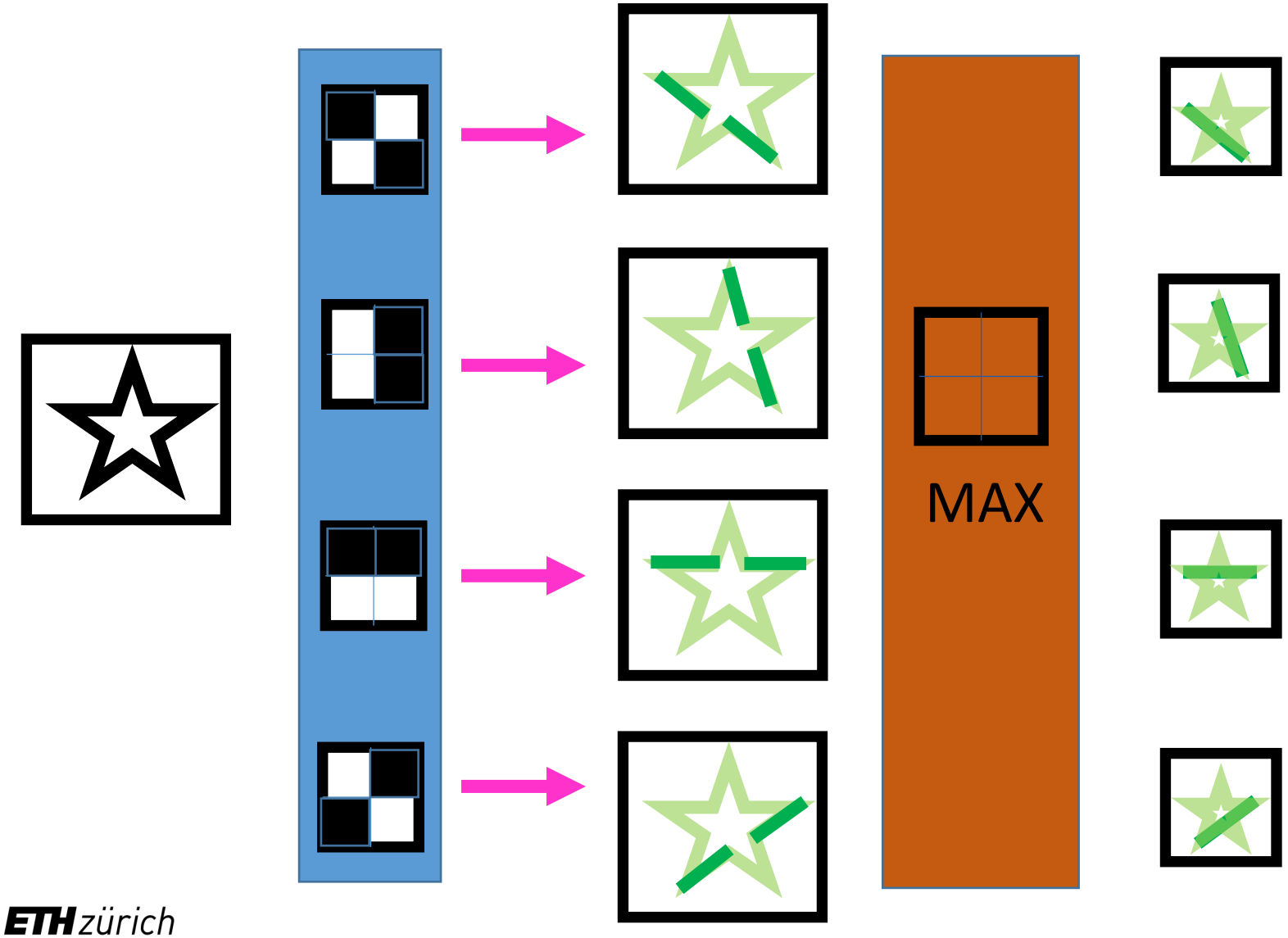
Gefiltertes Bild



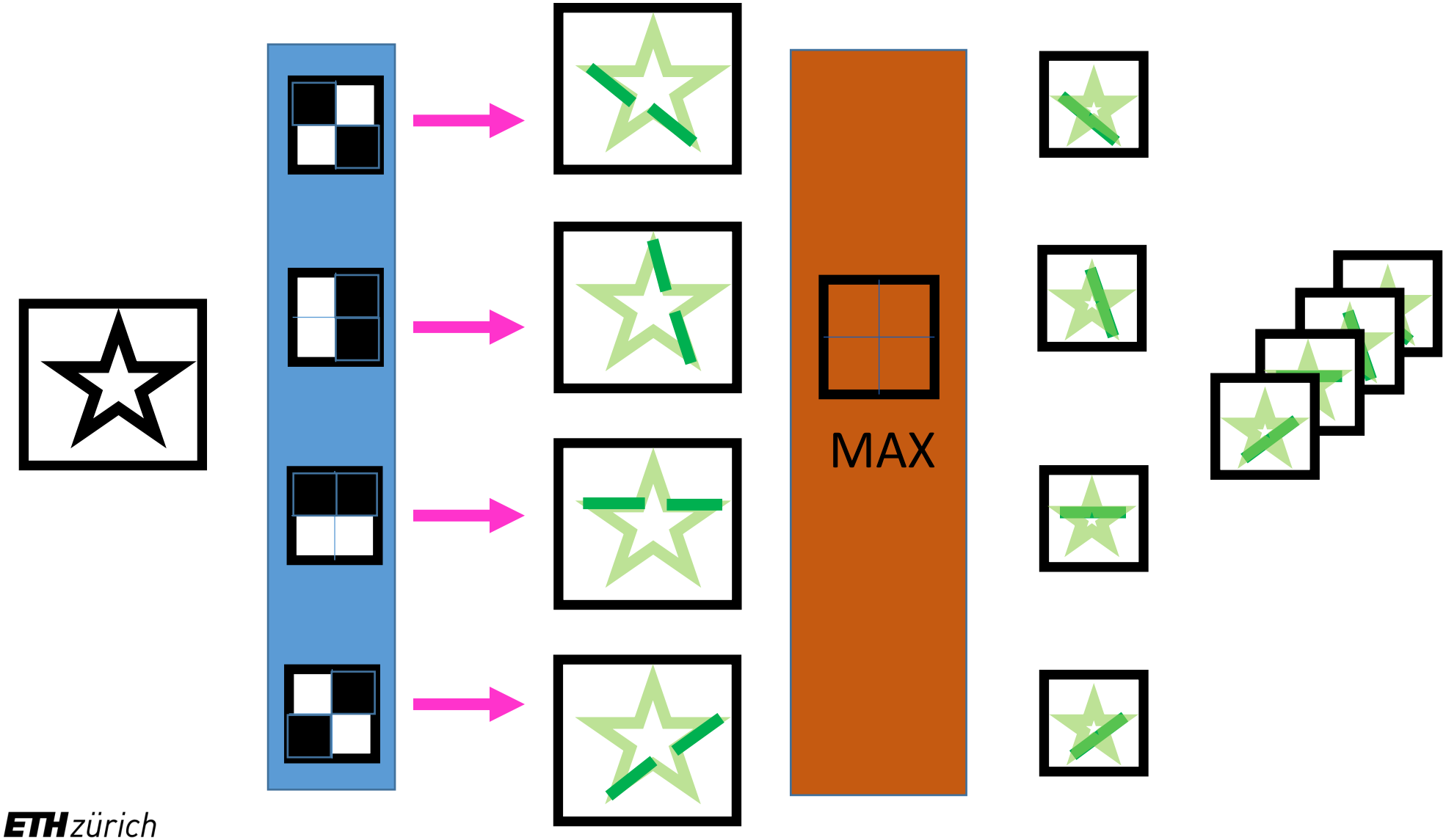
4	5
6	2

Ausgabe

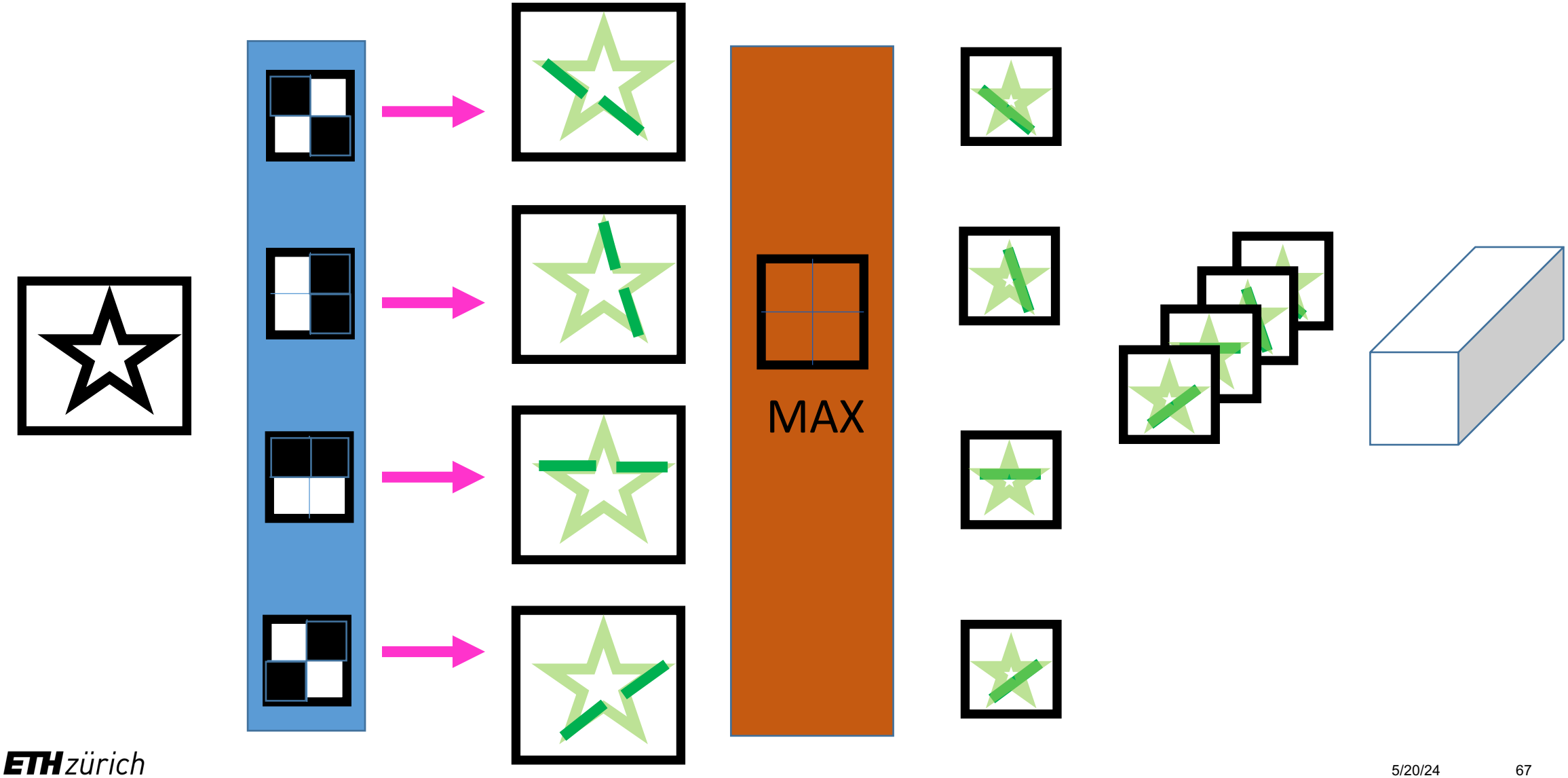
Convolutional neural networks



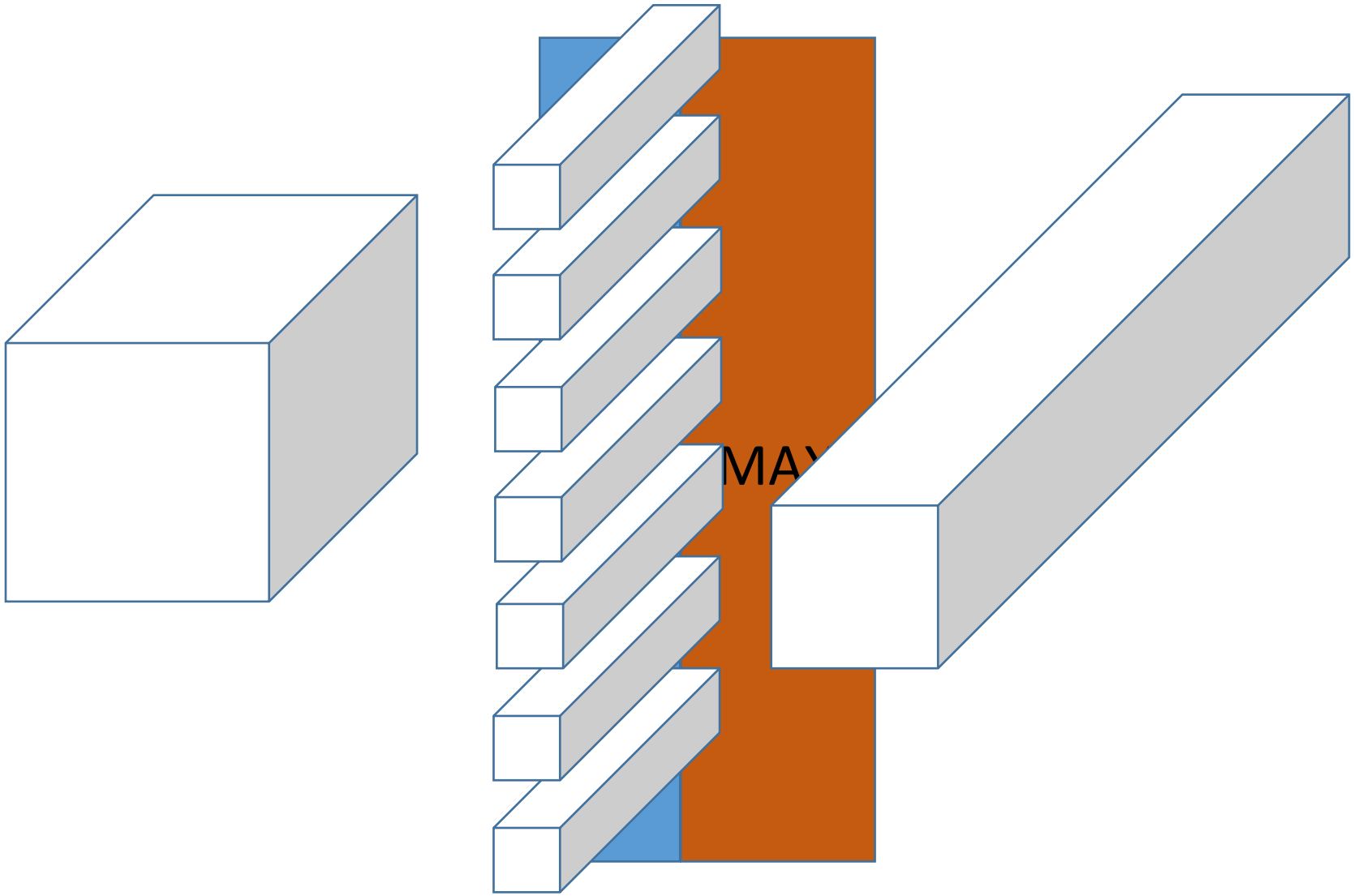
Convolutional neural networks



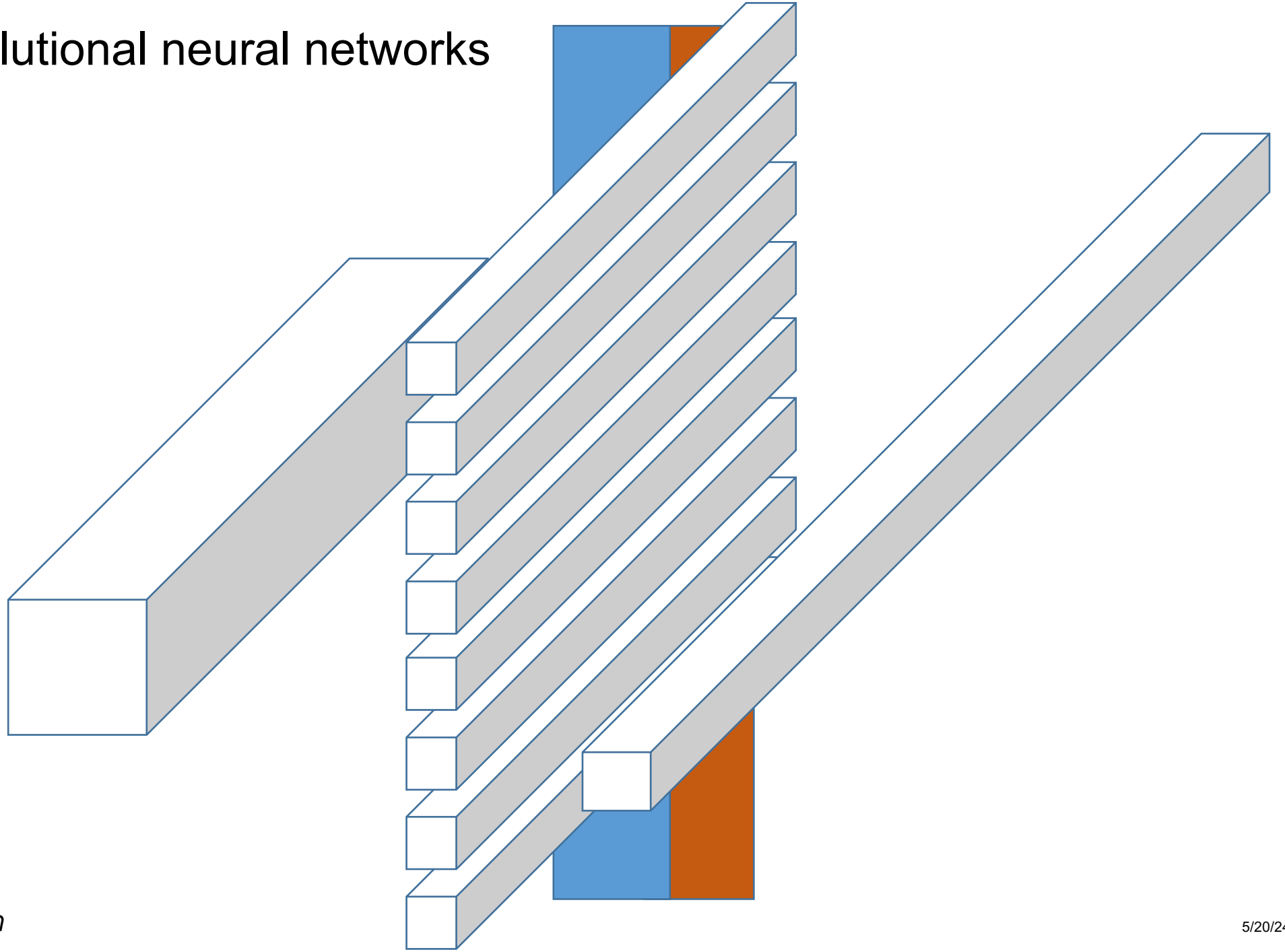
Convolutional neural networks



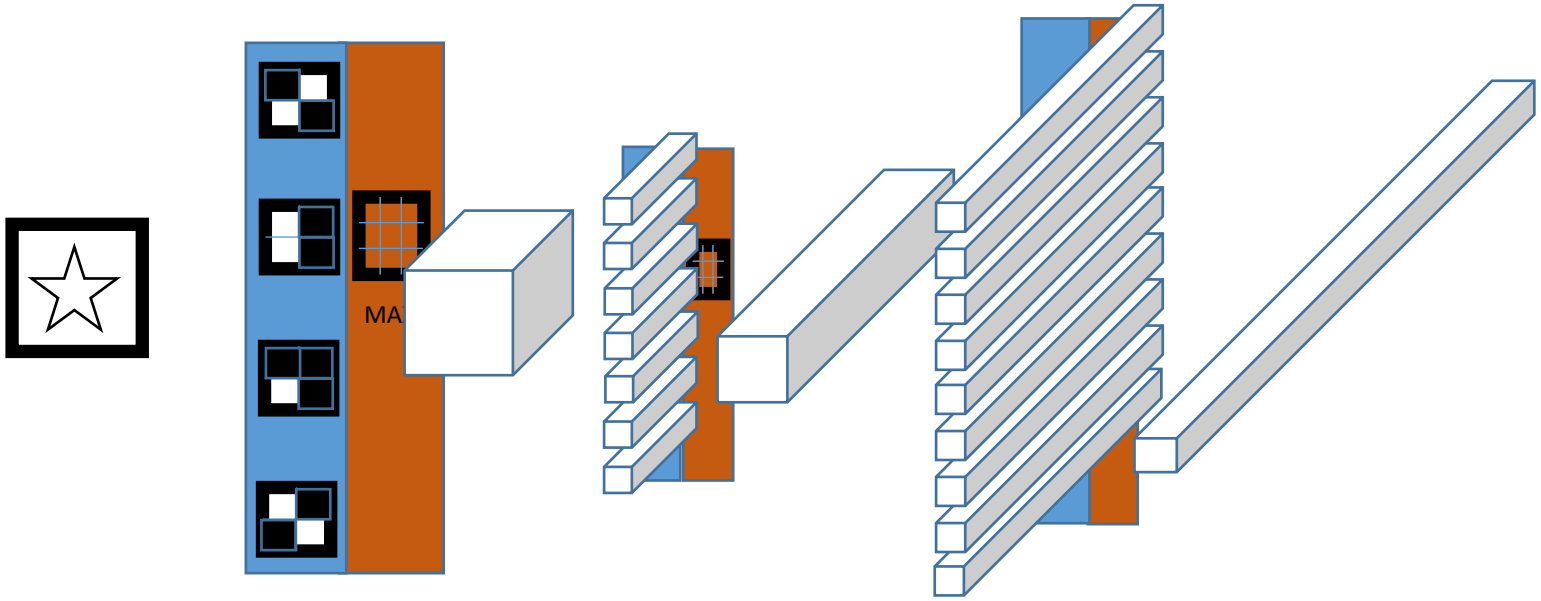
Convolutional neural networks



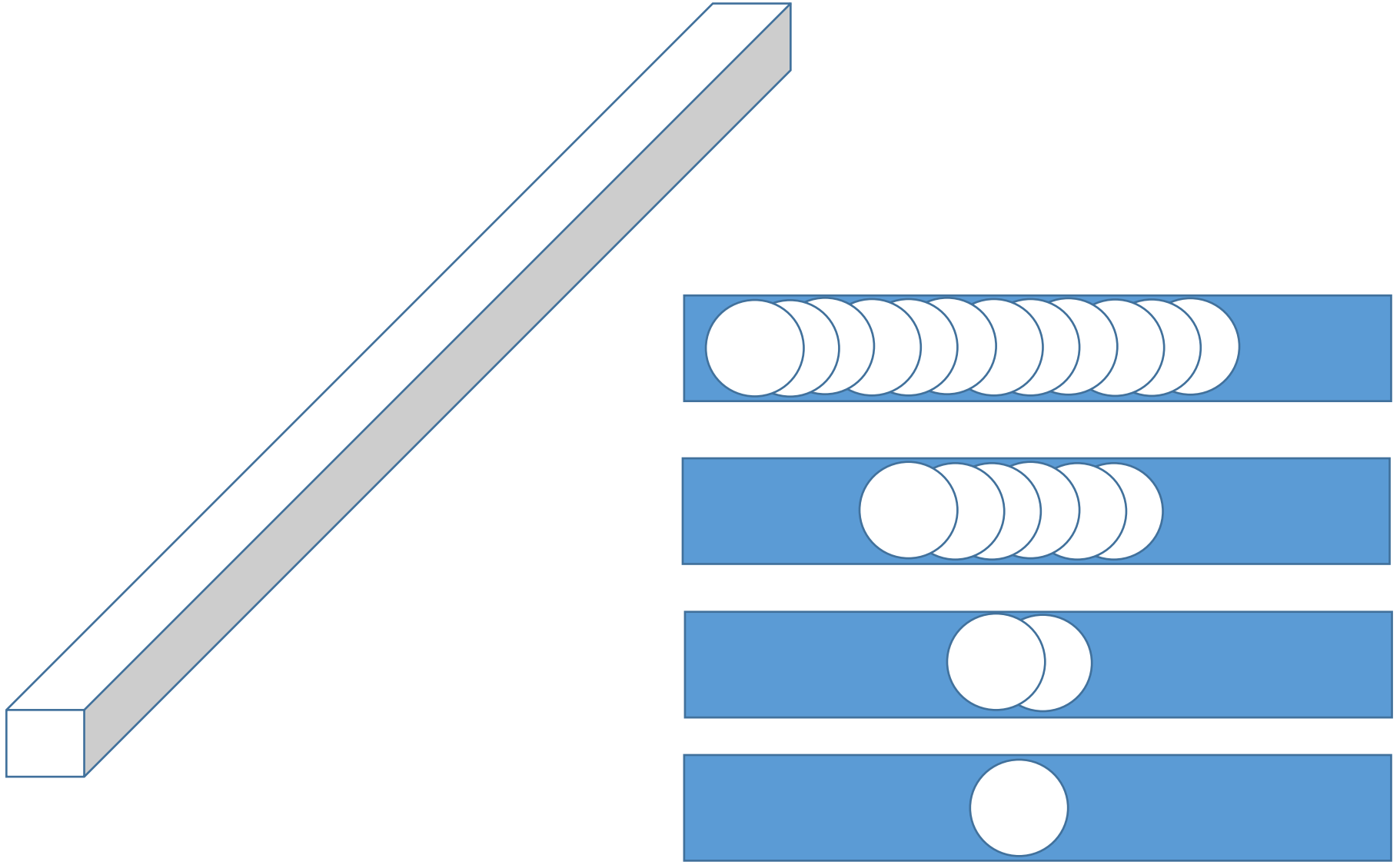
Convolutional neural networks



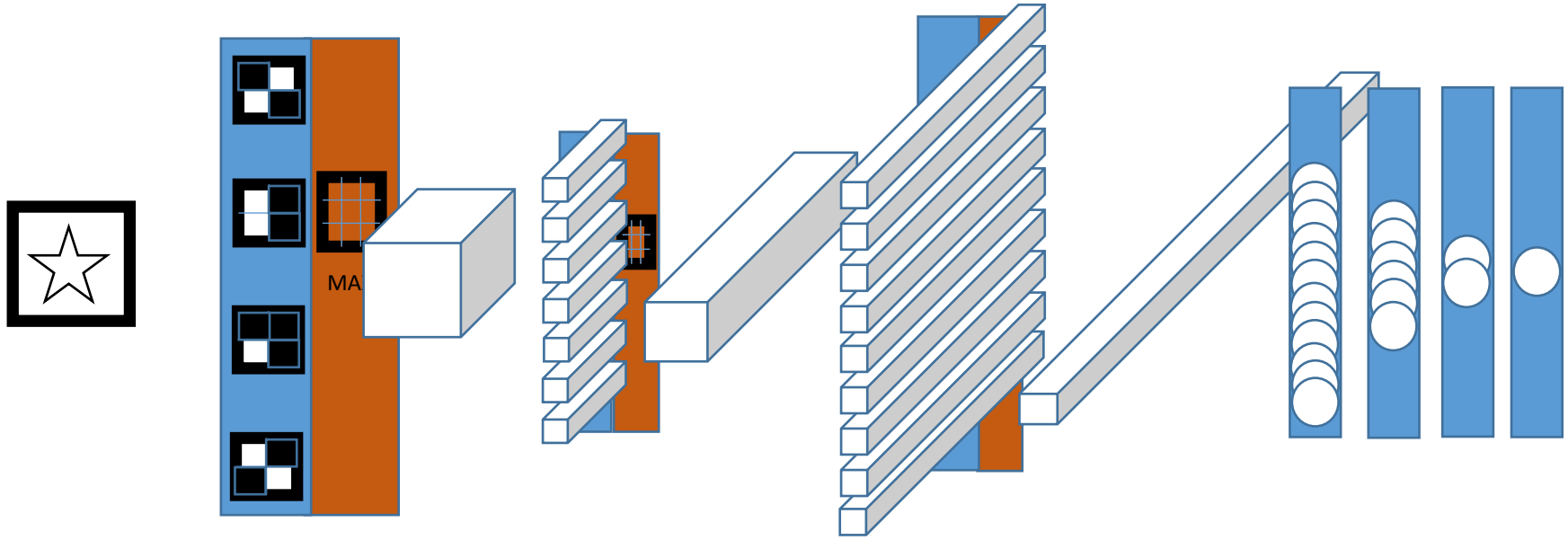
Convolutional neural networks



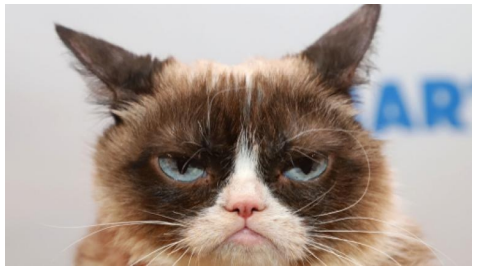
Convolutional neural networks

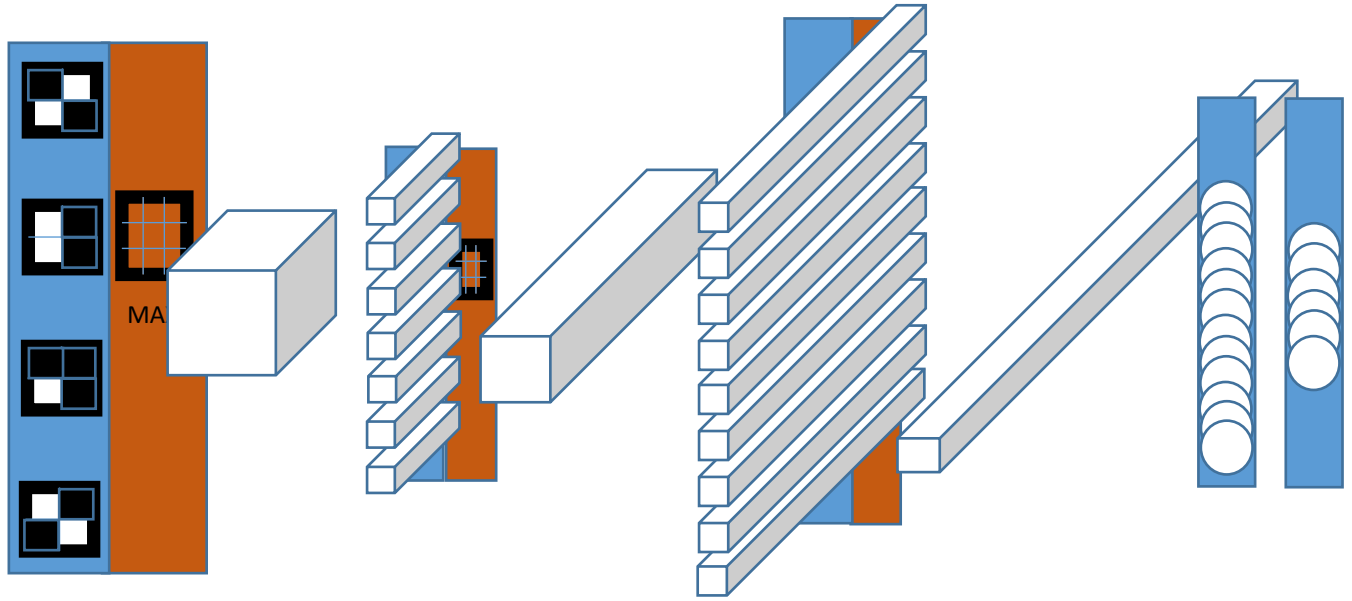


Convolutional neural networks

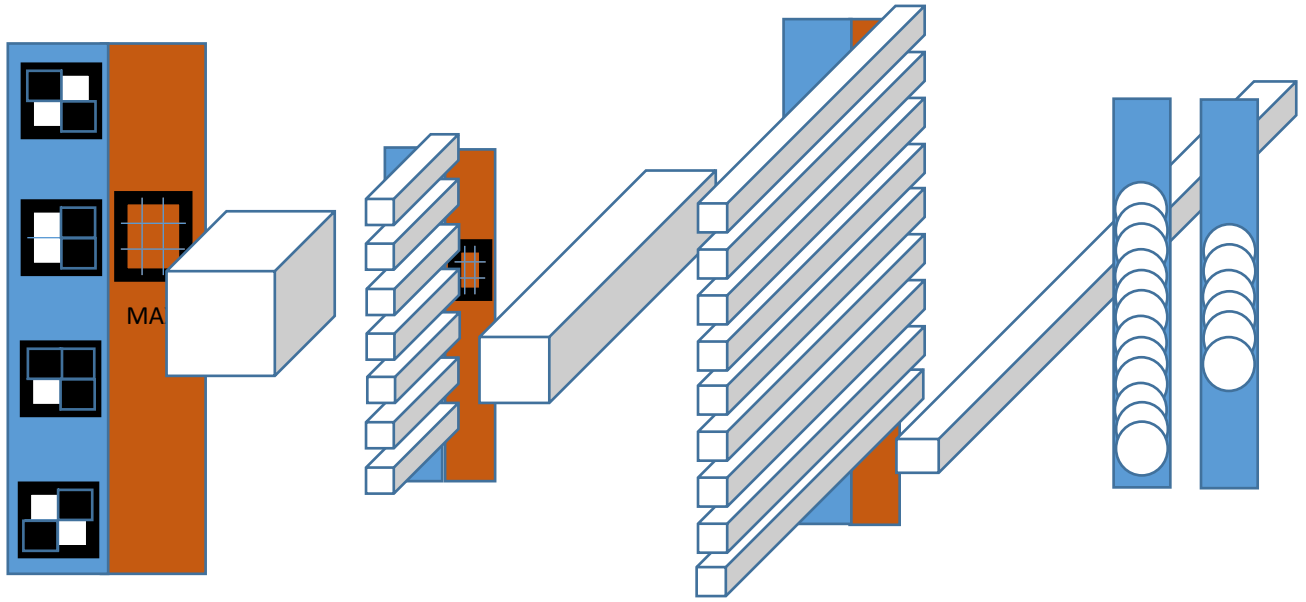


Bilderkennung





Softmax



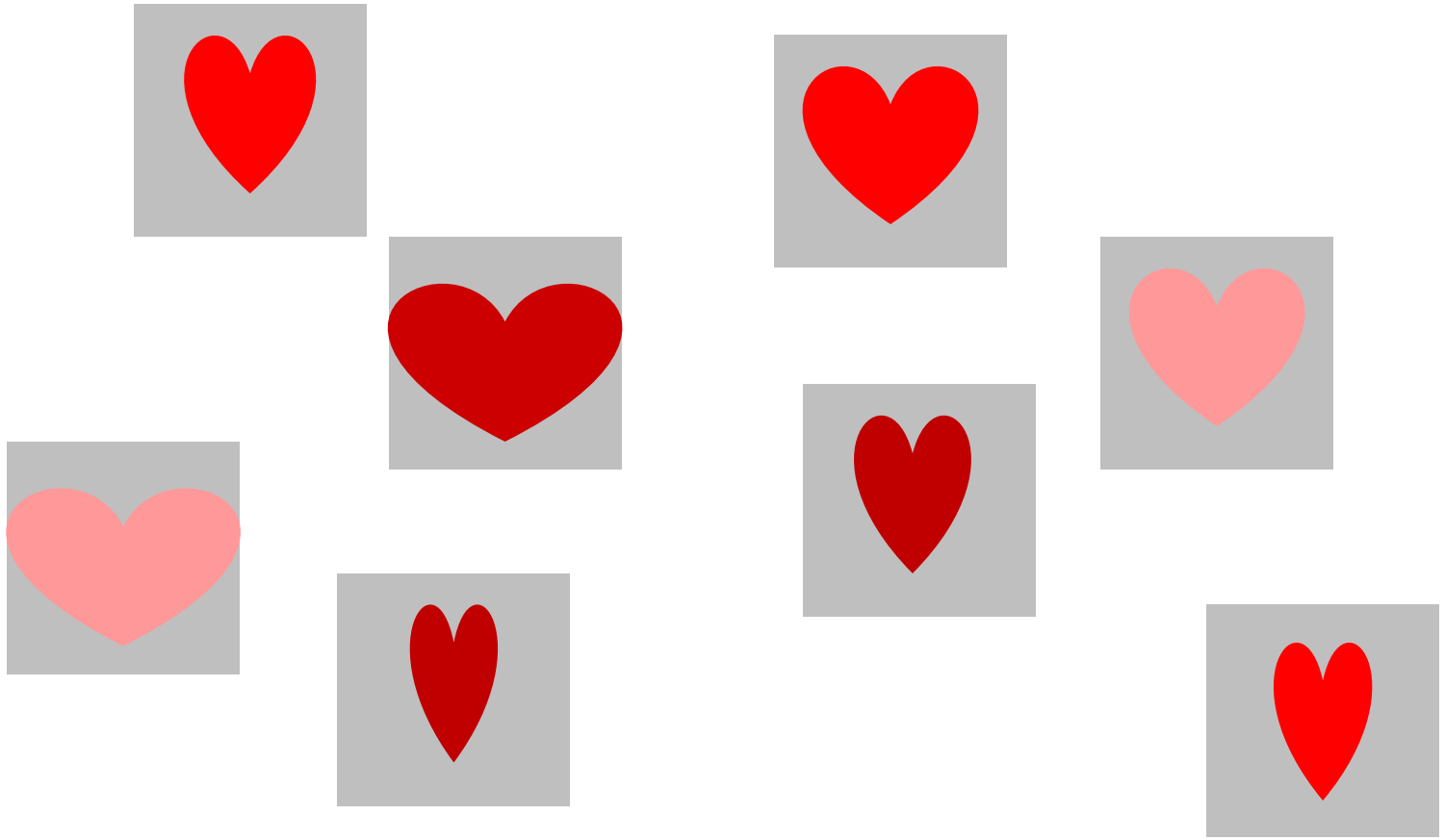
Softmax

Andere Architekturen: Autoencoders

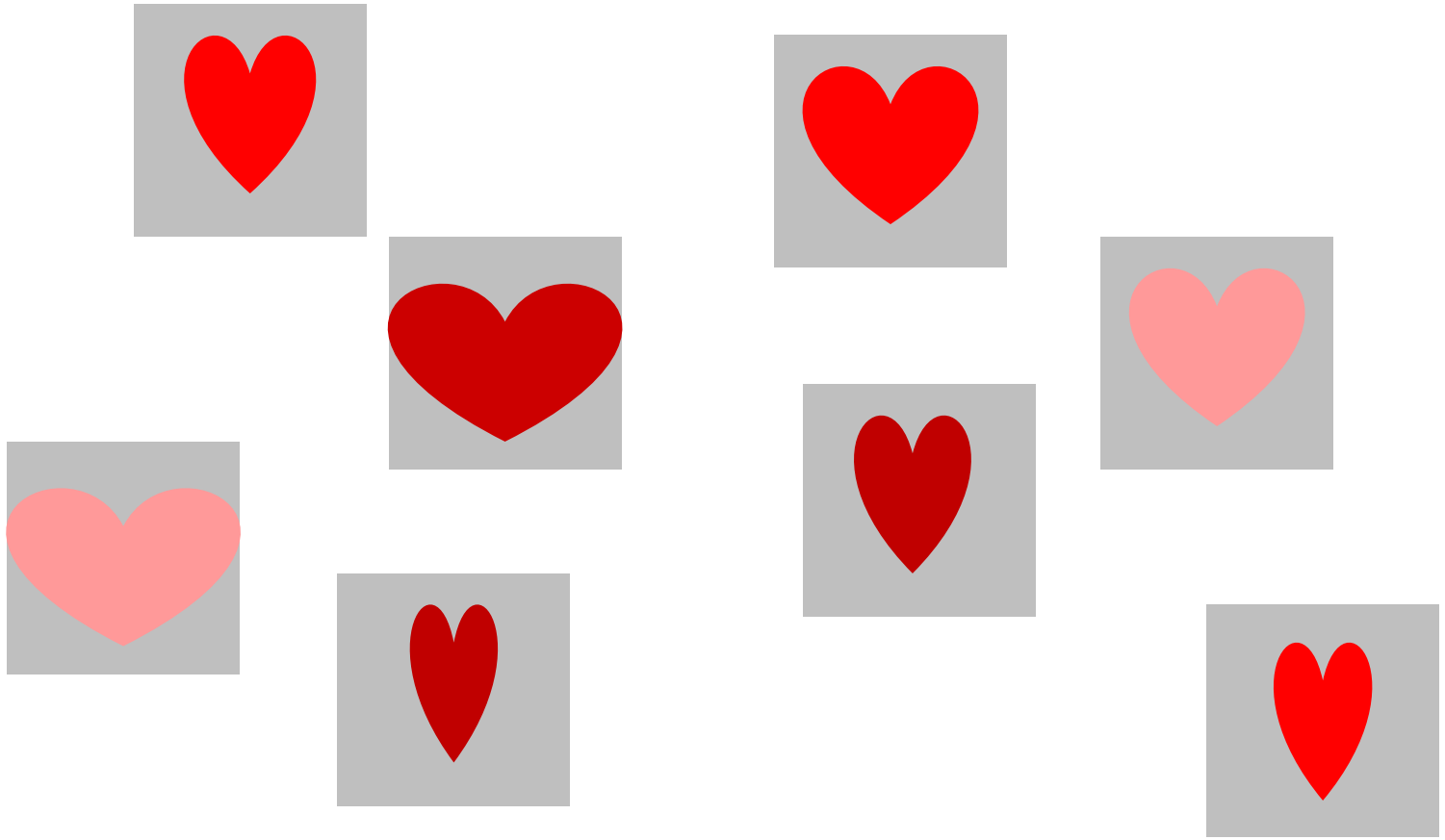
Autoencoders

- Ein Autoencoder ist eine neuronale Architektur, die *numerische Darstellungen* berechnet.

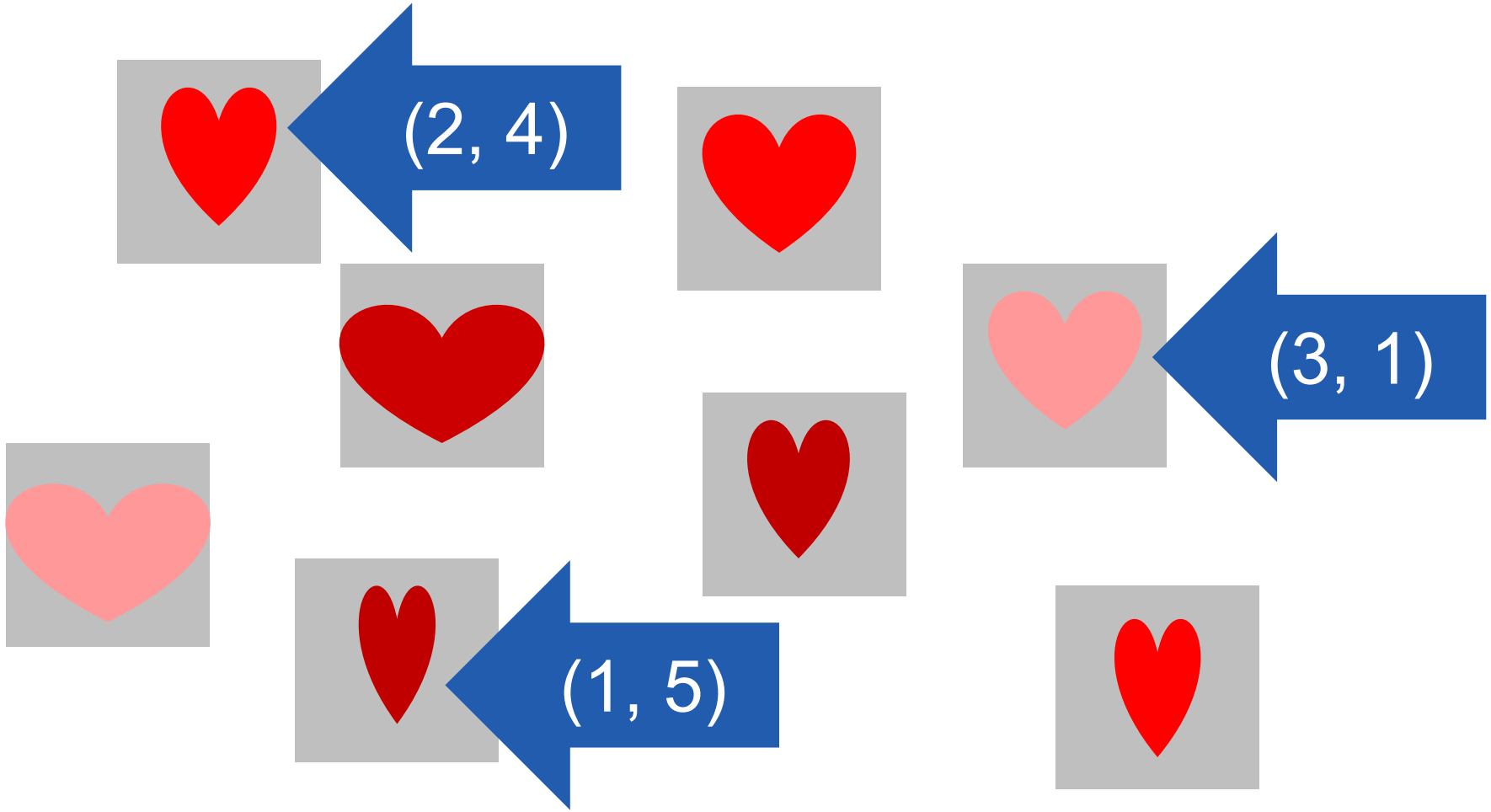
Beispiel



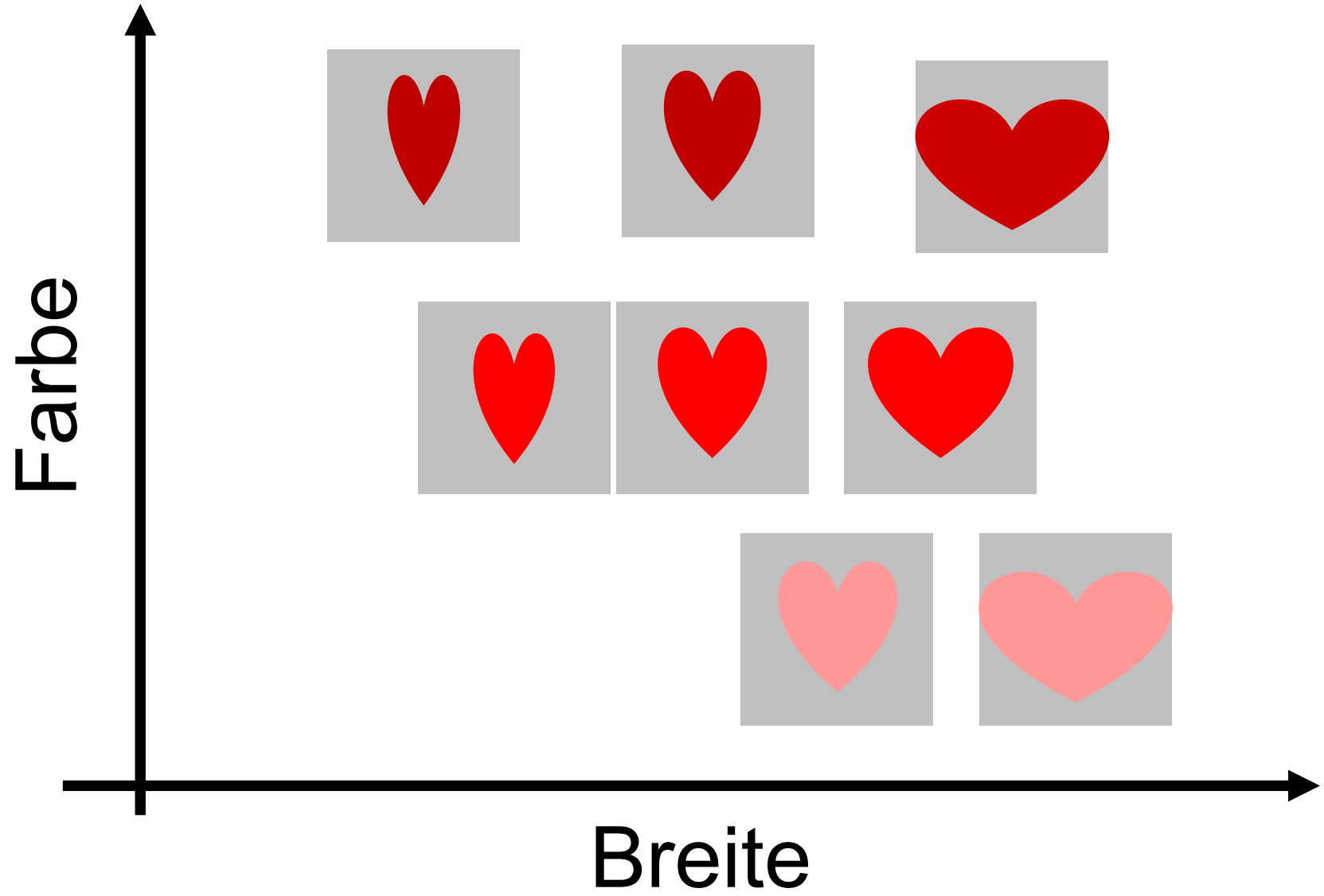
Die Herzen unterscheiden sich nur in deren Farben und Breite



Man kann die Herzen mit Zahlenpaaren darstellen



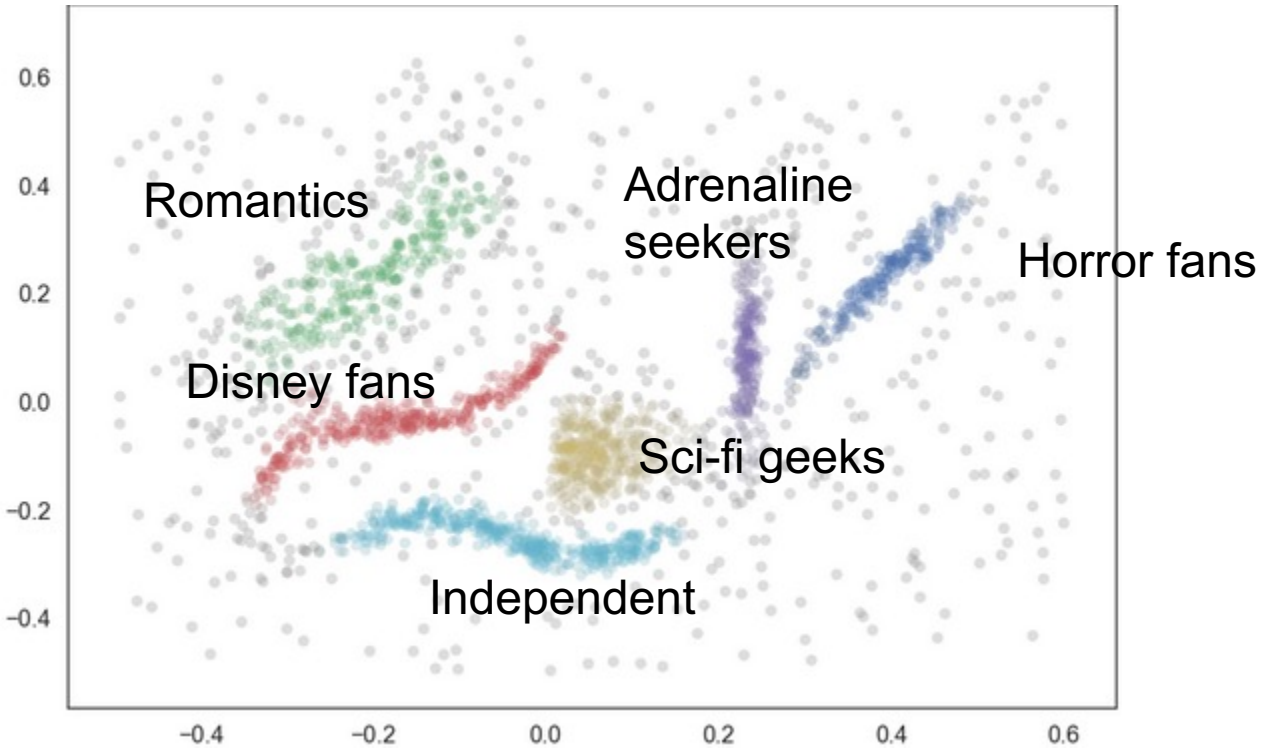
Man kann die Herzen in einem zweidimensionalen Raum darstellen



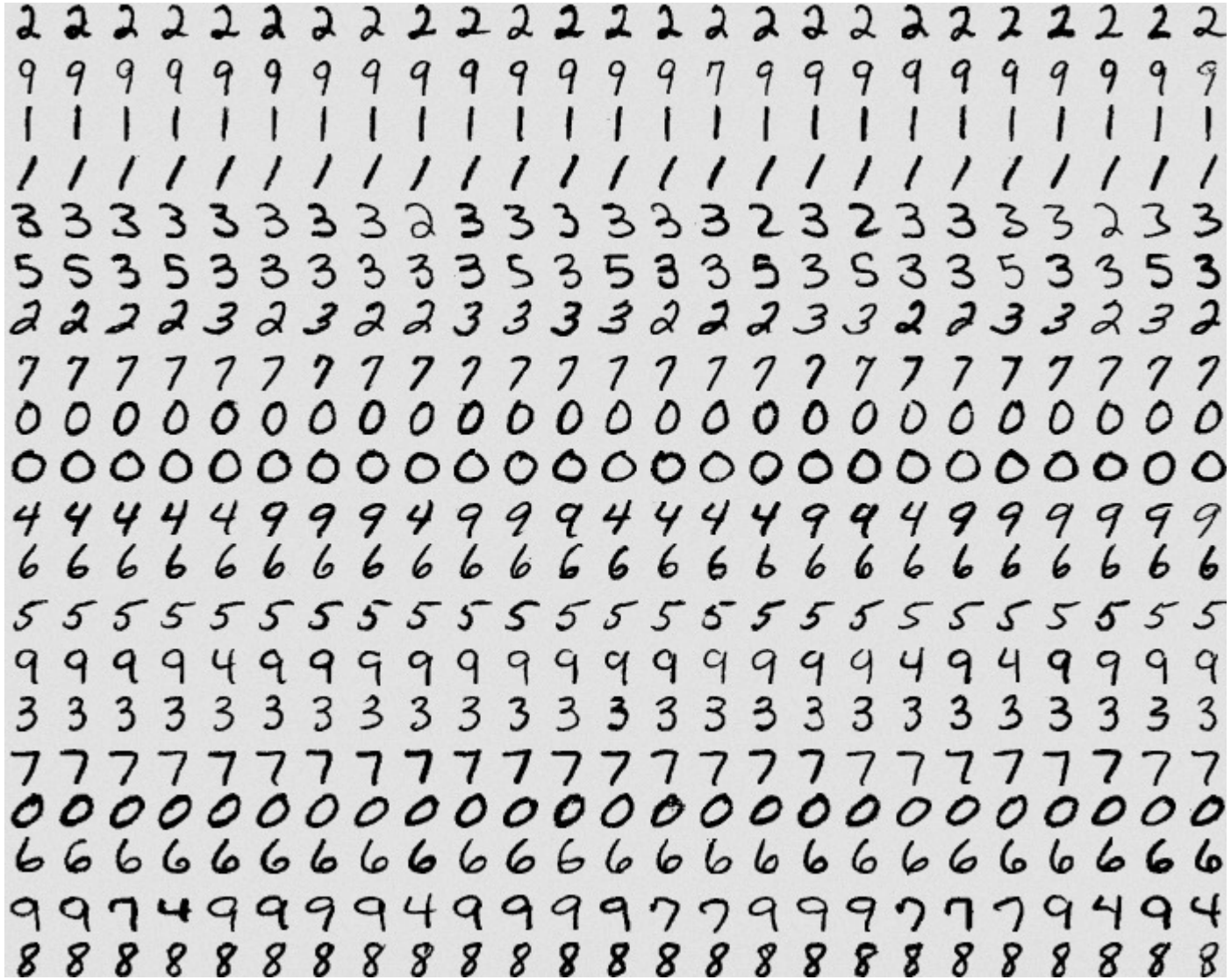
Vorteile einer Darstellung

- Eine Darstellung veranschaulicht welche Features ein Beispiel charakterisiert.
- Man kann Clusters im Datensatz entdecken.

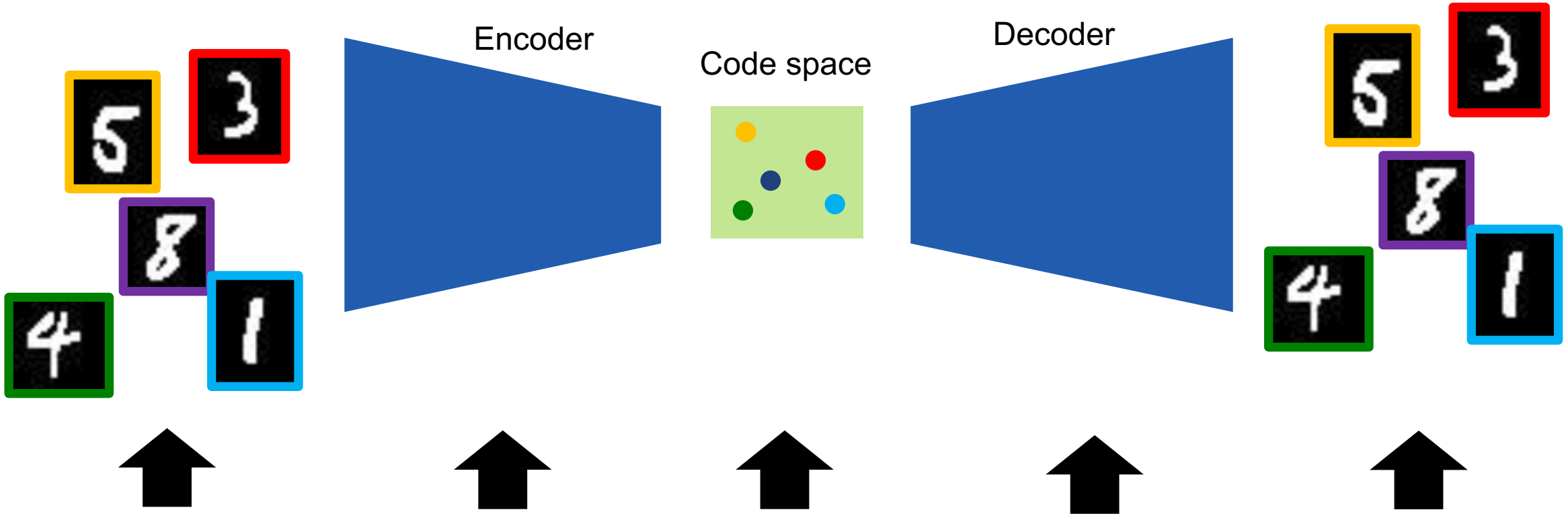
Clusters


Romantics
Adrenaline seekers
Horror fans
Disney fans
Sci-fi geeks
Independent
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.4
−0.2
0.0
0.2
0.4
0.6

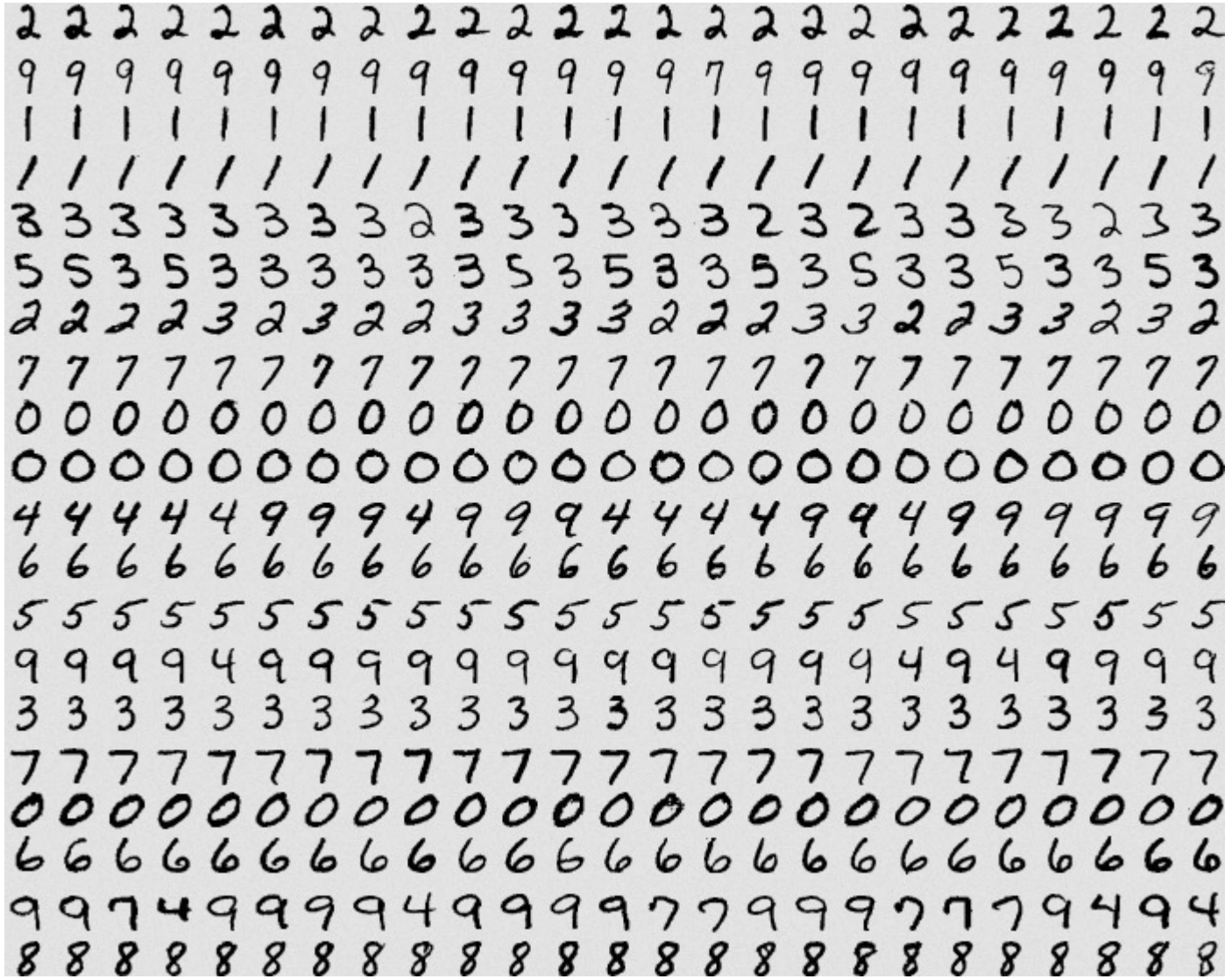
Autoencoders: Handgeschriebene Ziffer

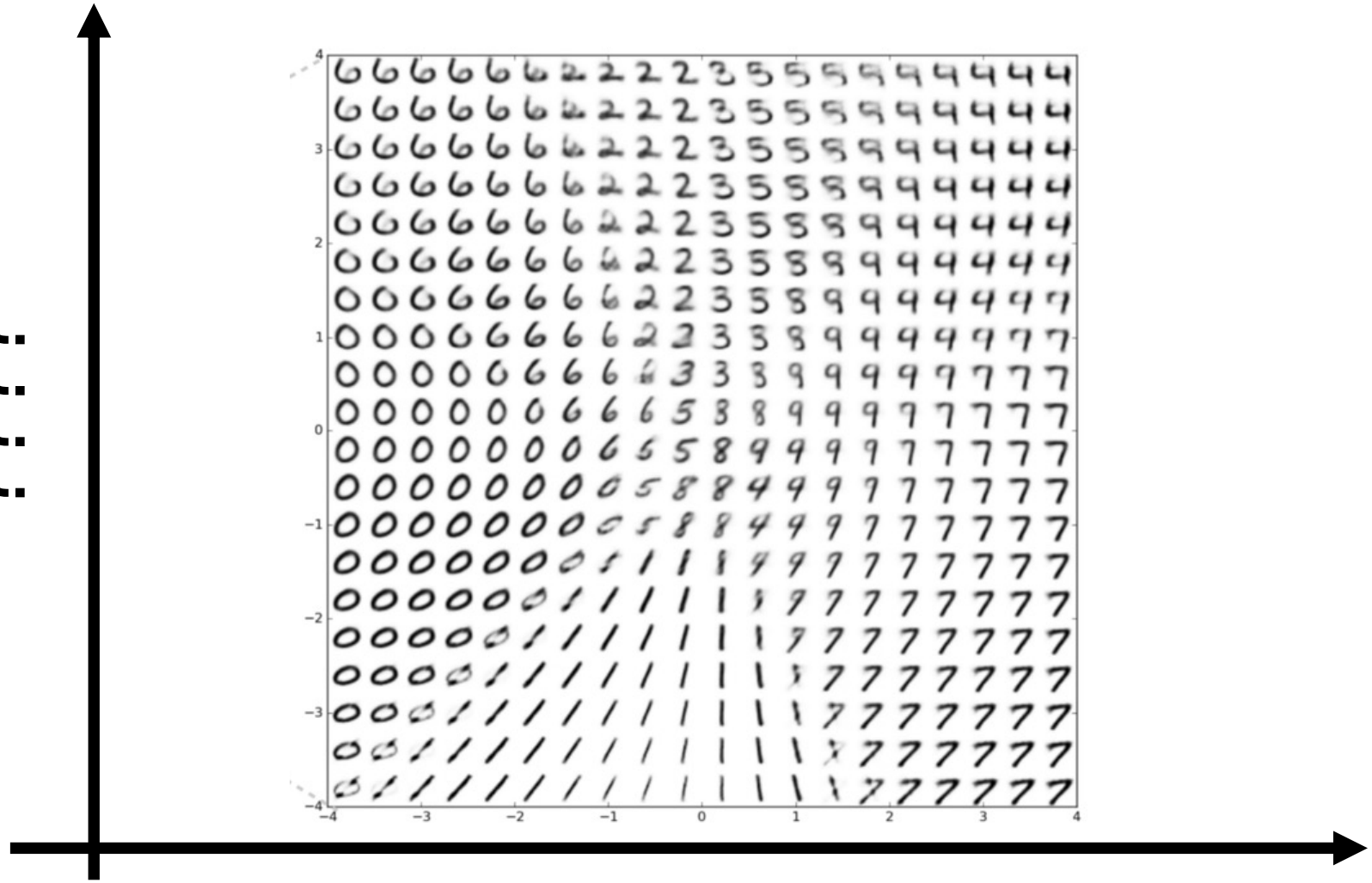


Autoencoders



Beispiel: Handgeschriebene Ziffer





Verlustfunktion

Verlustfunktion

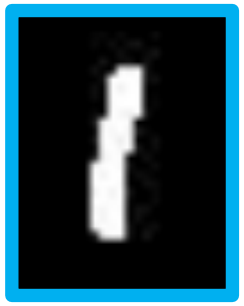
- Sei Enc der Encoder und Dec der Decoder.

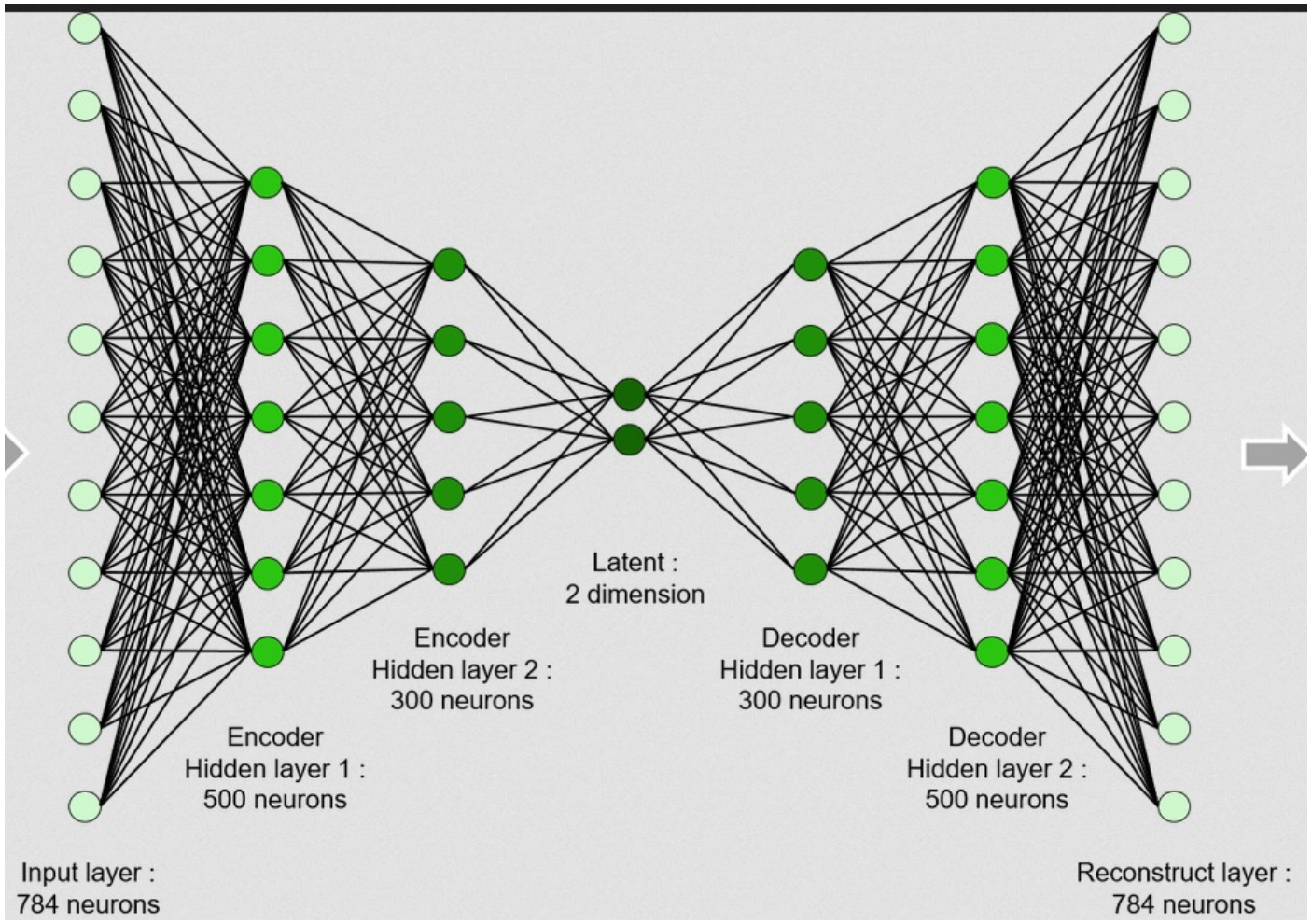
$$L(D, Enc, Dec) = \sum_{x \in D} \left\| x - Dec\big(Enc(x)\big) \right\|^2.$$

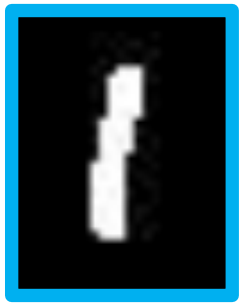
Die Sanduhr

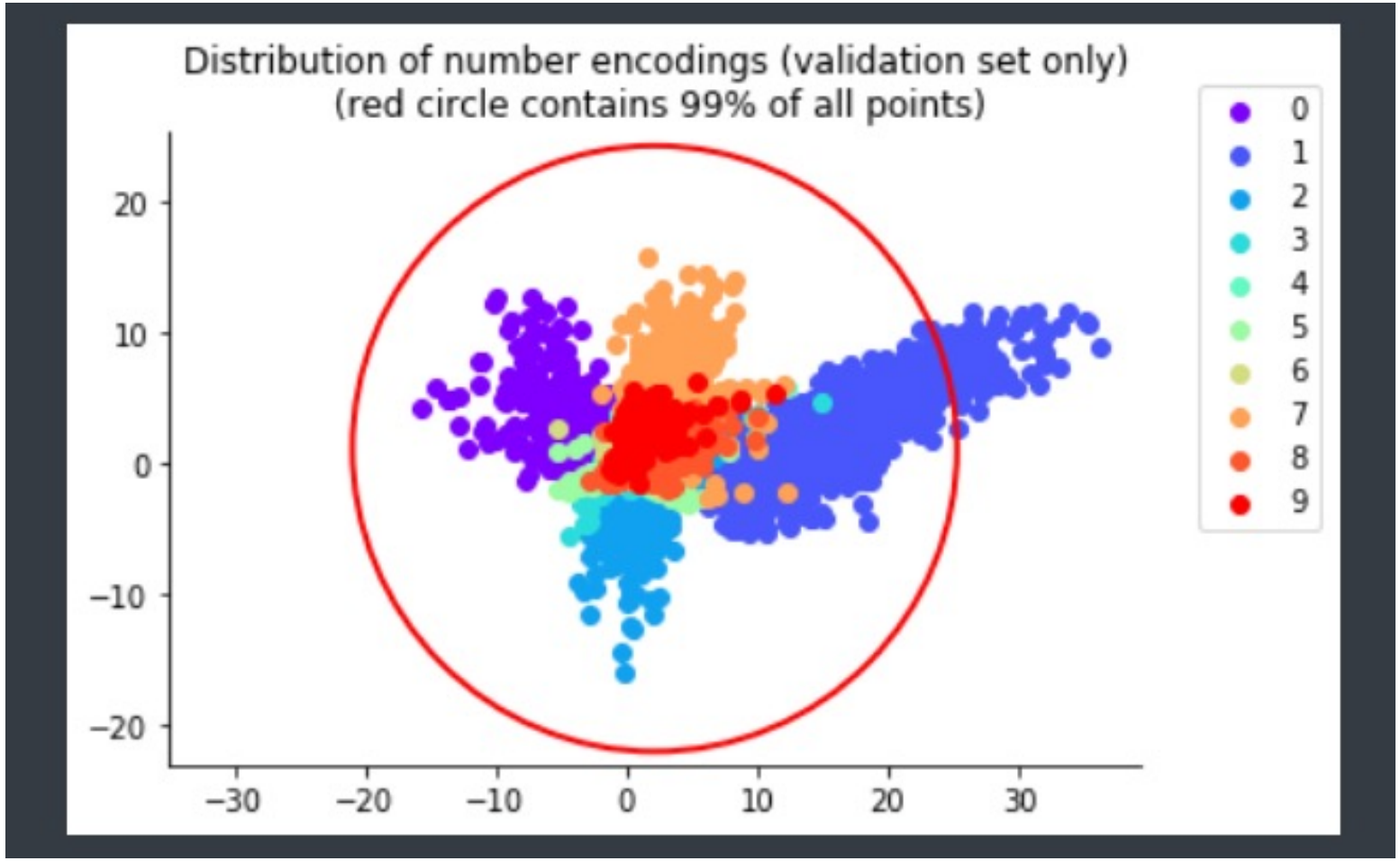
- Tiefere Schichte extrahieren die wichtigste Information.
- Die Sanduhrform zwingt der Encoder nur die wesentliche Information zu extrahieren.
- Der Decoder lernt, das ursprüngliche Bild nur anhand der Darstellung zu rekonstruieren.

Encoder und Decoder







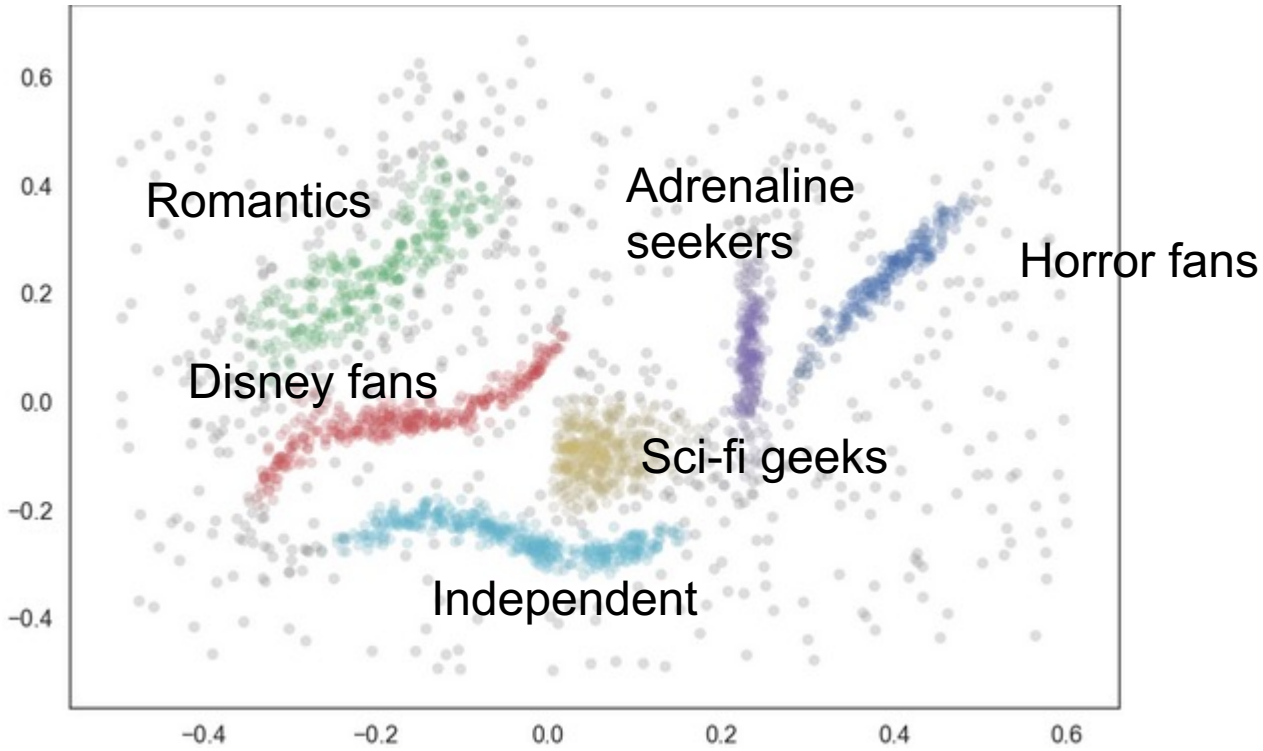

Distribution of number encodings (validation set only)
(red circle contains 99% of all points)
20
10
0
−10
−20
−30
−20
−10
0
10
20
30
0
1
2
3
4
5
6
7
8
9

Was bedeuten die Achsen?

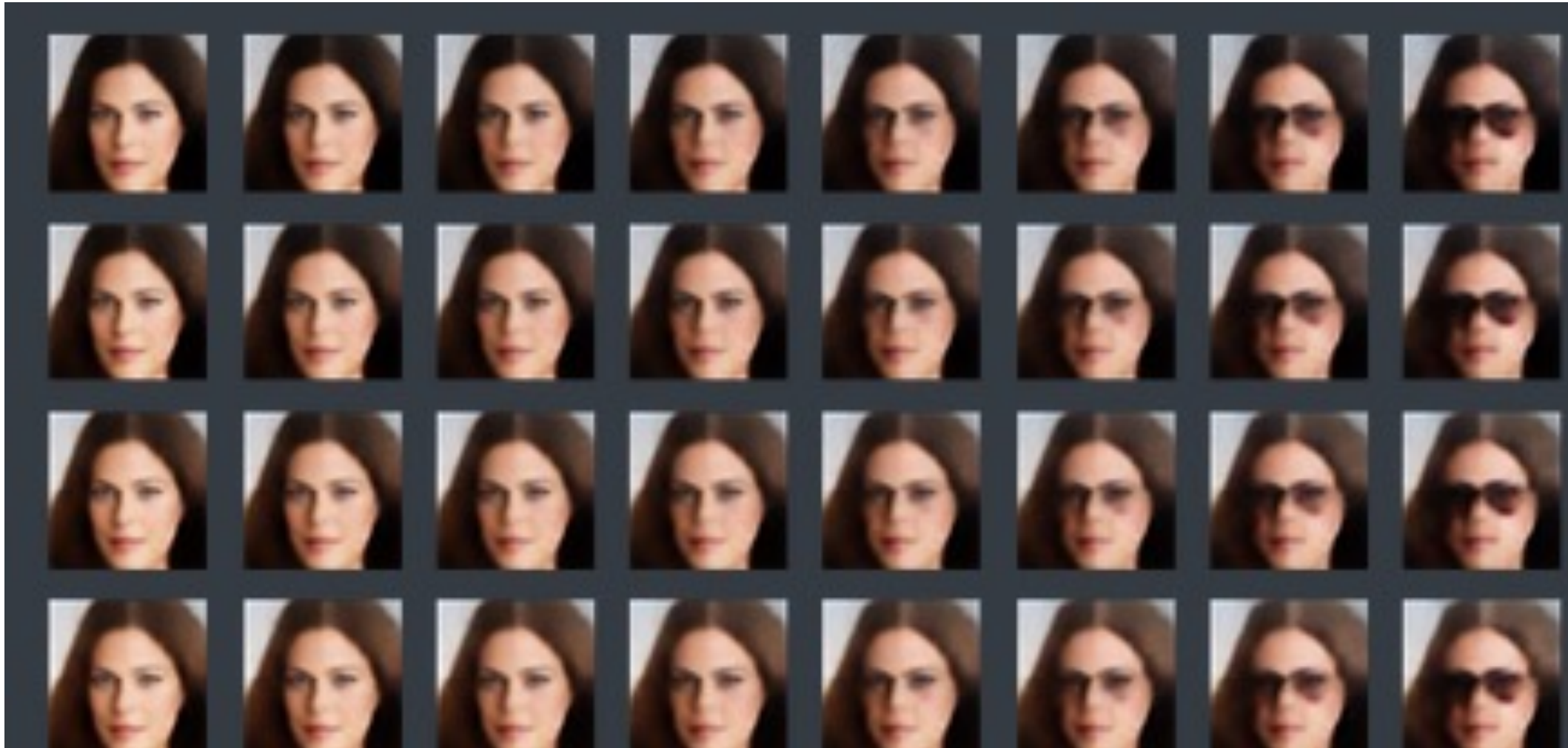
- Das weisst nur der Autoencoder…

So warum sind Autoencoders hilfreich?

- Empfehlungsdienste



Interpolation mit Autoencoders



Credits

- Images taken from:
  - https://www.compthree.com/blog/autoencoder/
  - https://hdbscan.readthedocs.io/en/latest/soft_clustering_explanation.html

Generative adversarial networks (GAN)

GANs

Diskriminator

Echt oder künstlich

$\mathbb{R}^d$

Generator

Anwendungen

- Erzeugen von Daten
- Fictitious characters in movies and games
- Text-to-Image Translation
- Deepfakes
- Face Aging
- Photo inpainting

Was haben wir heute gelernt?

- Neuronale Netze
- Warum sind neuronale Netze im Allgemein besser als andere Modelle?
- Convolutional neural networks.
- Autoencoders.
- GANs.