Machine Learning

tinyurl.com/ml-intro-pg

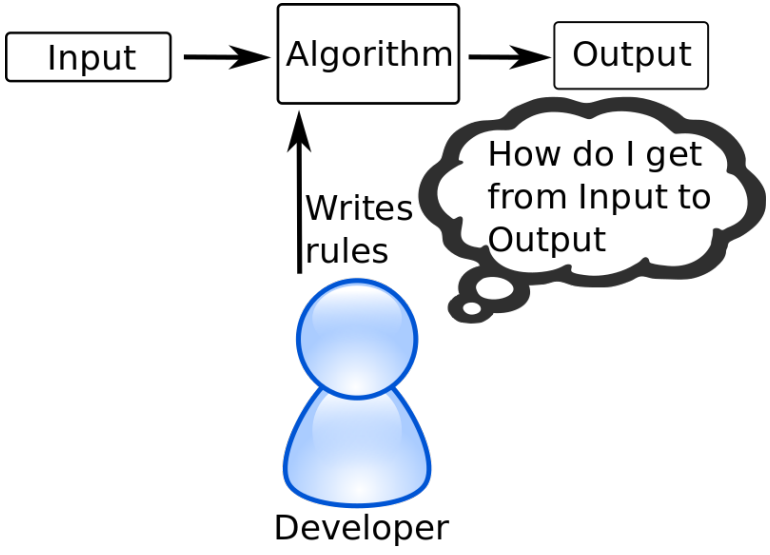
Was ist ML-KA?

- Kurz für **Machine Learning Karlsruhe**
- **Hochschulgruppe** seit 15. Oktober 2015
- **Ziel**: Wissen über ML Verbreiten / Mehren
- **Idee**: Forum für interessierte Studenten bilden, organisation in kleinen Gruppen
- **Umsetzung** bisher
  - Paper Discussion Group (PDG, wöchentlich)
  - Gesellschaftliche Implikationen vom ML (GIML, 2-wöchentlich)
  - Teilnahme (und Preisträger) der Herbsttagung der Gesellschaft für Datenanalyse und Numerische Klassifikation
- Mehr auf https://ml-ka.de

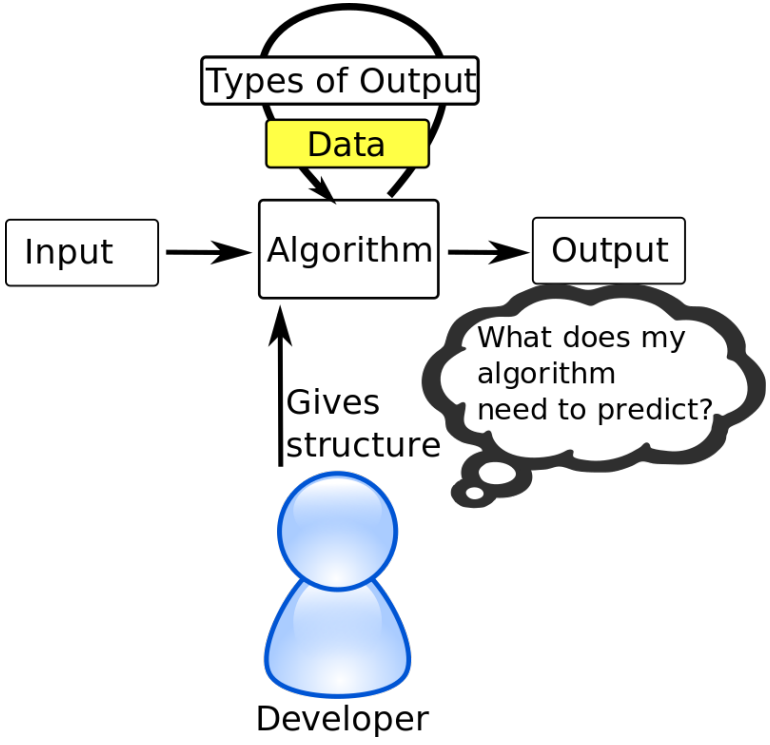
Wer ist ML-KA?

- Vorstand:
  - Martin Thoma (info@martin-thoma.de)
  - Marvin Teichmann
  - Marvin Schweizer
- Mitglieder: Überwiegend aktuelle Studenten des KIT
  - 20 Mitglieder (Stand: 3. Februar 2016)
  - 213 Facebook Mitglieder (Stand: 26. Mai 2016)

Was ist Machine Learning?

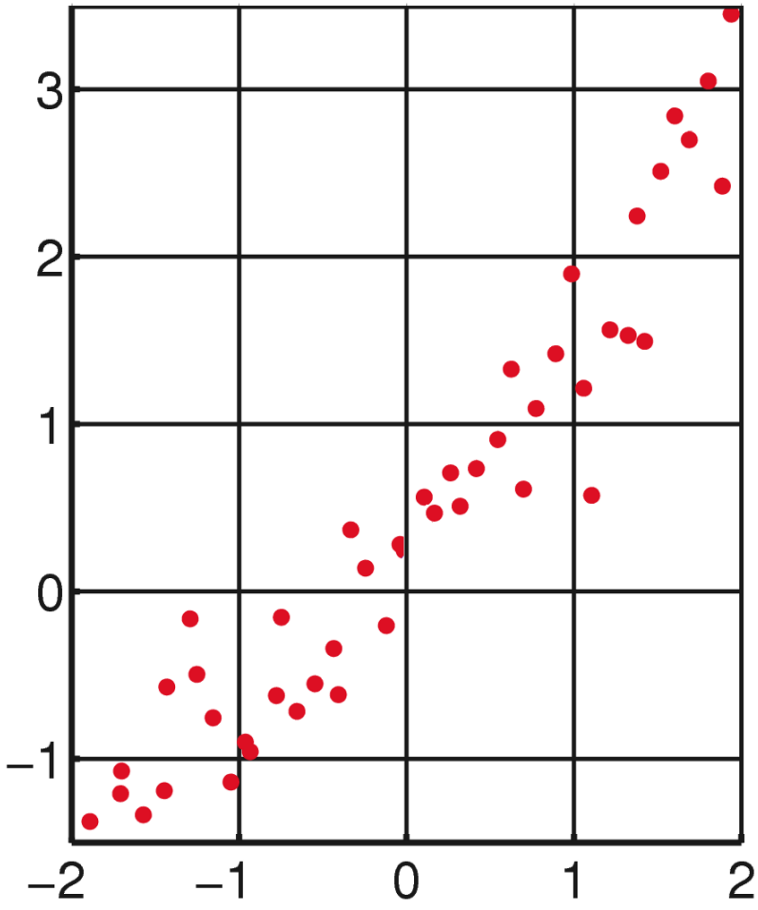


Traditional Development Model



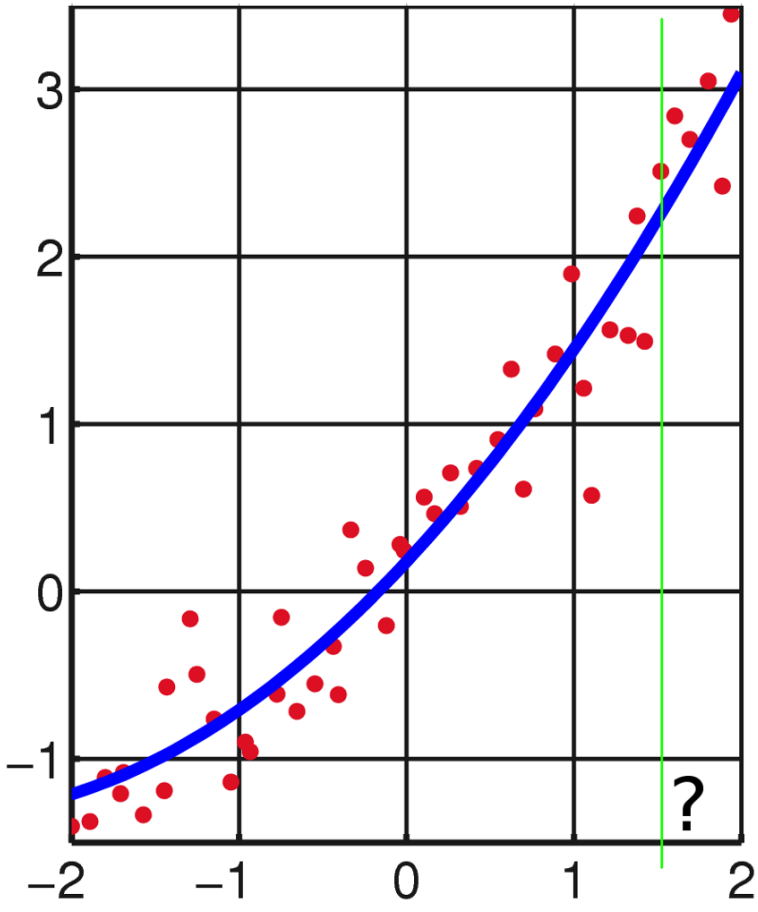
ML Model

Problemtypen



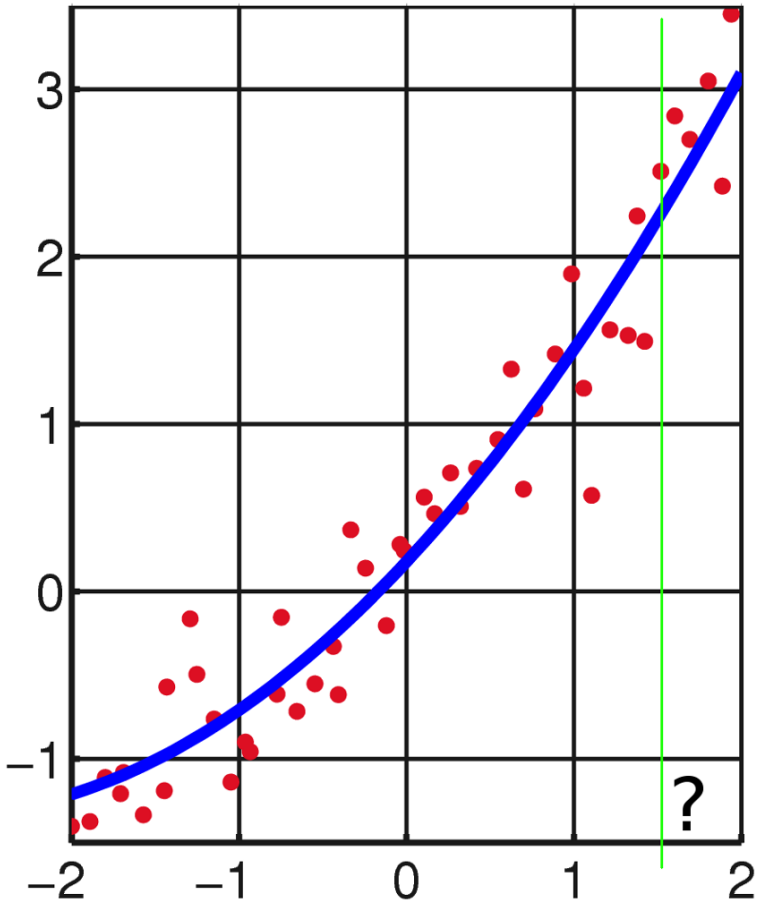
Regression

Problemtypen

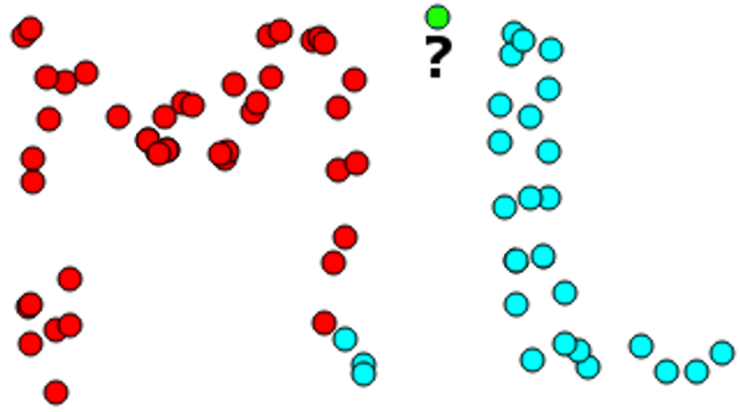


Regression

Problemtypen

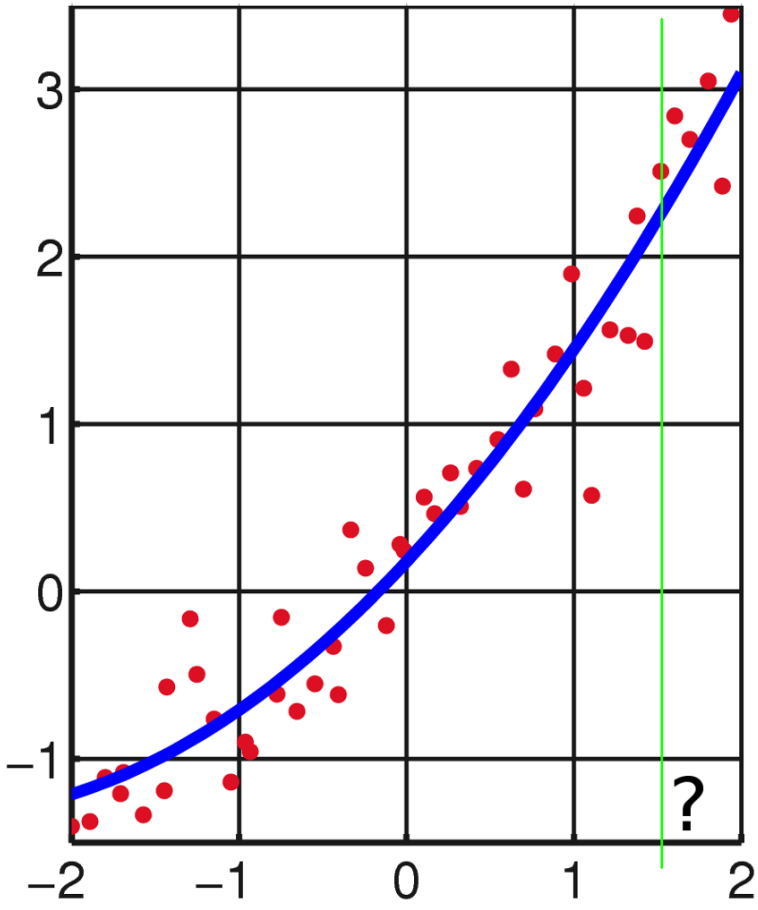


Regression

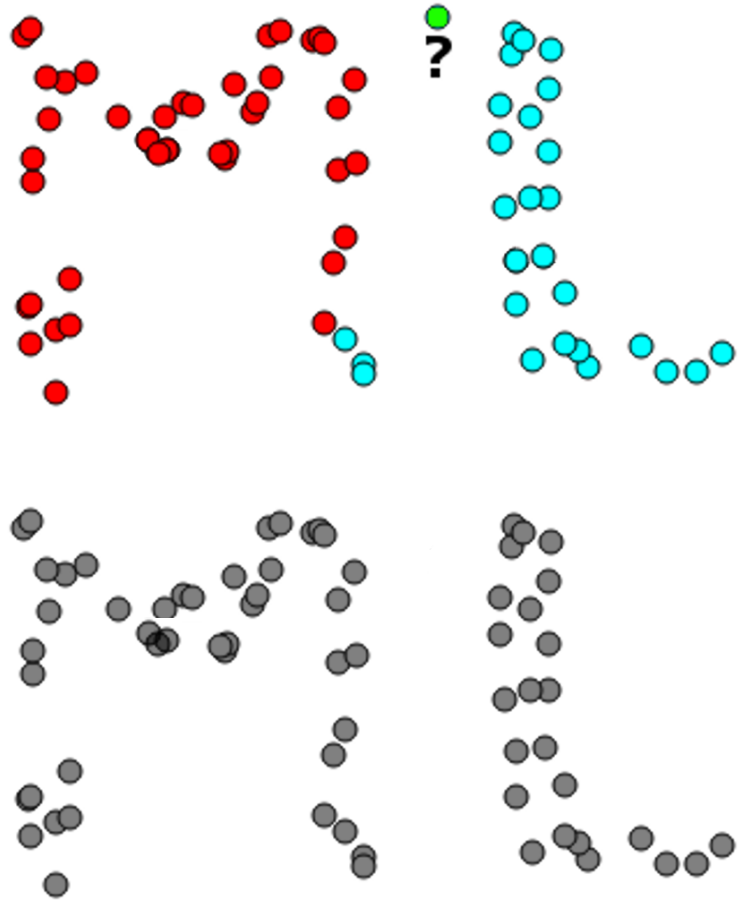


Klassifikation (überwacht)

Problemtypen

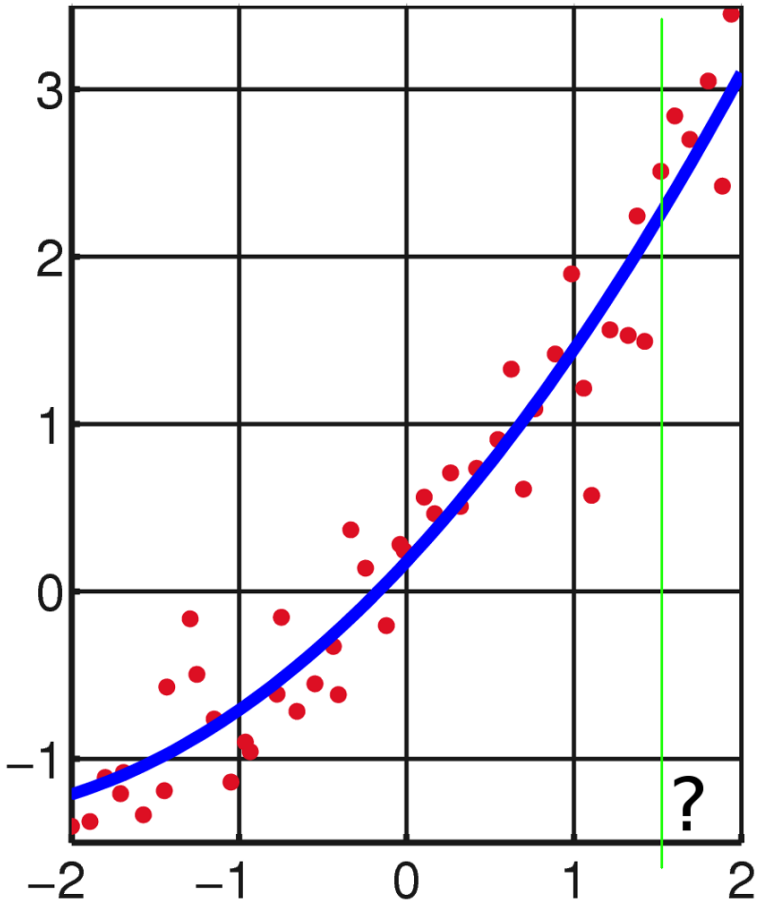


Regression

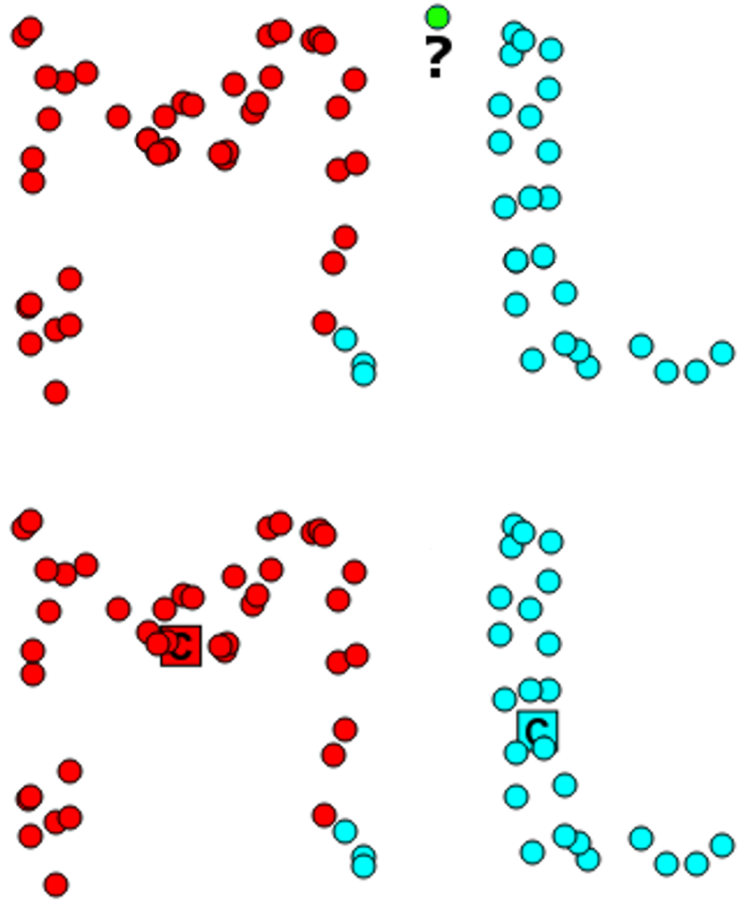


und Clustering (unüberwacht)

Problemtypen

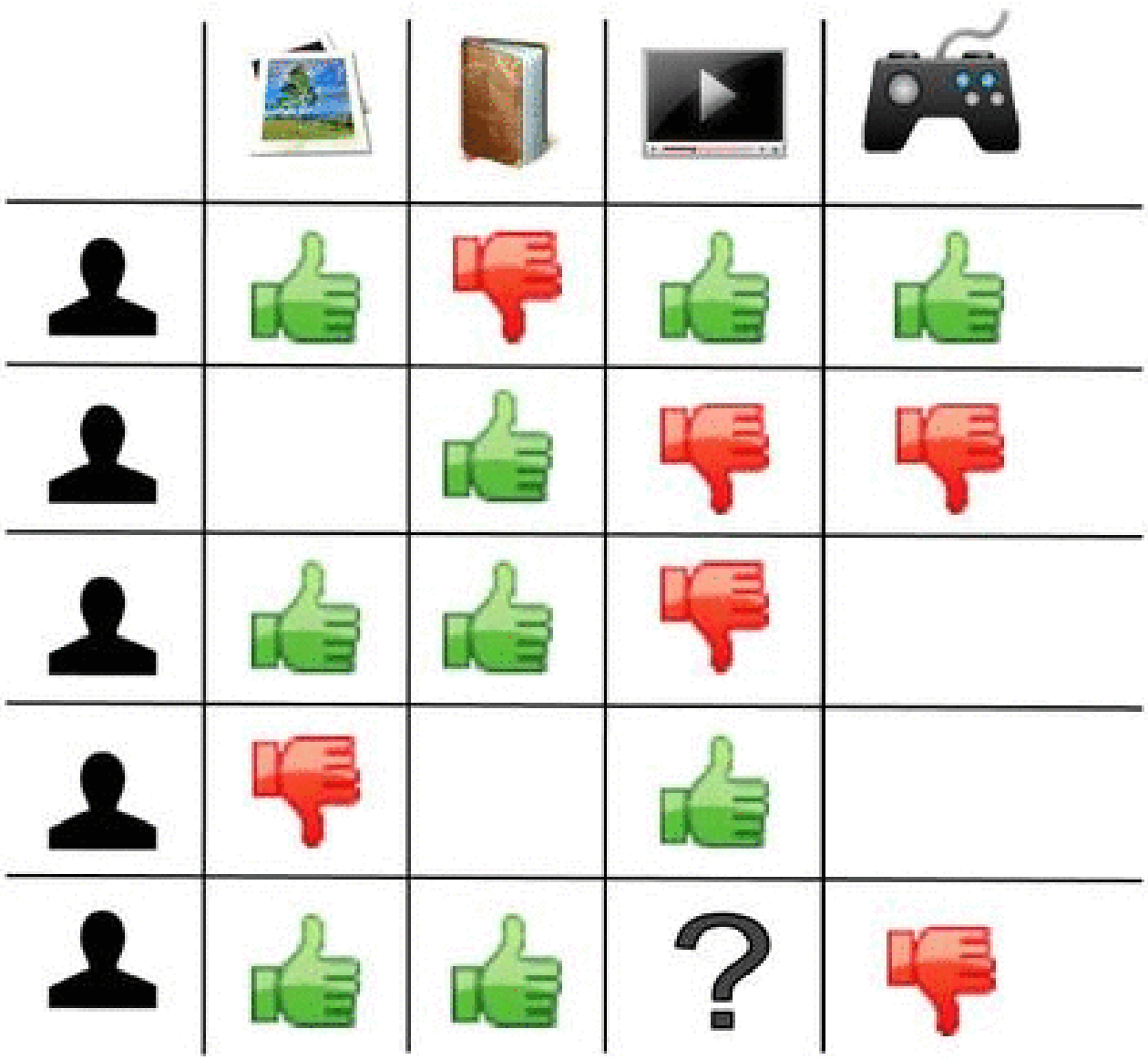


Regression



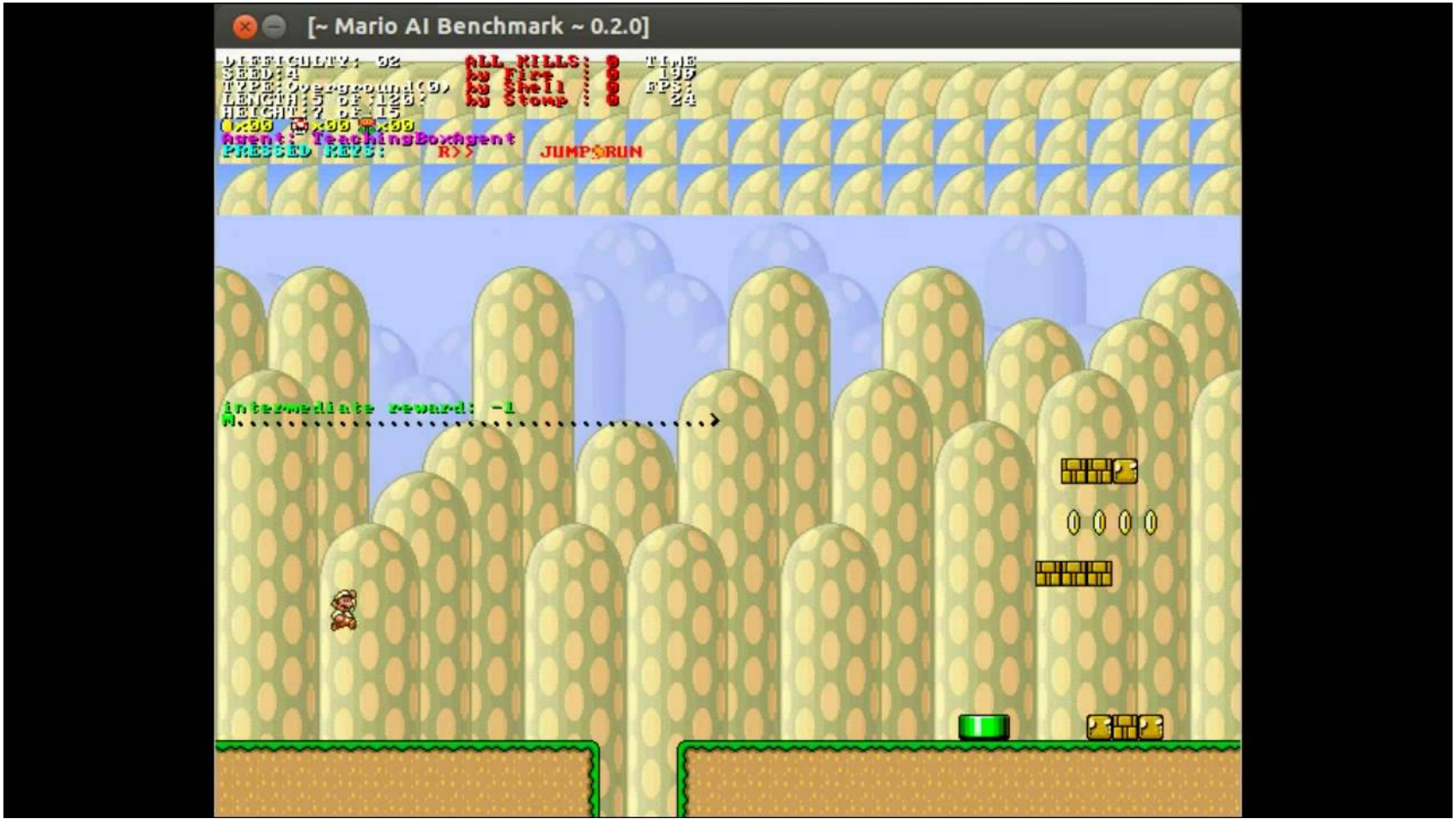
und Clustering (unüberwacht)

Was ist Machine Learning?



Collaborative Filtering

Was ist Machine Learning?



Reinforcement Learning (RL)

ImageNet

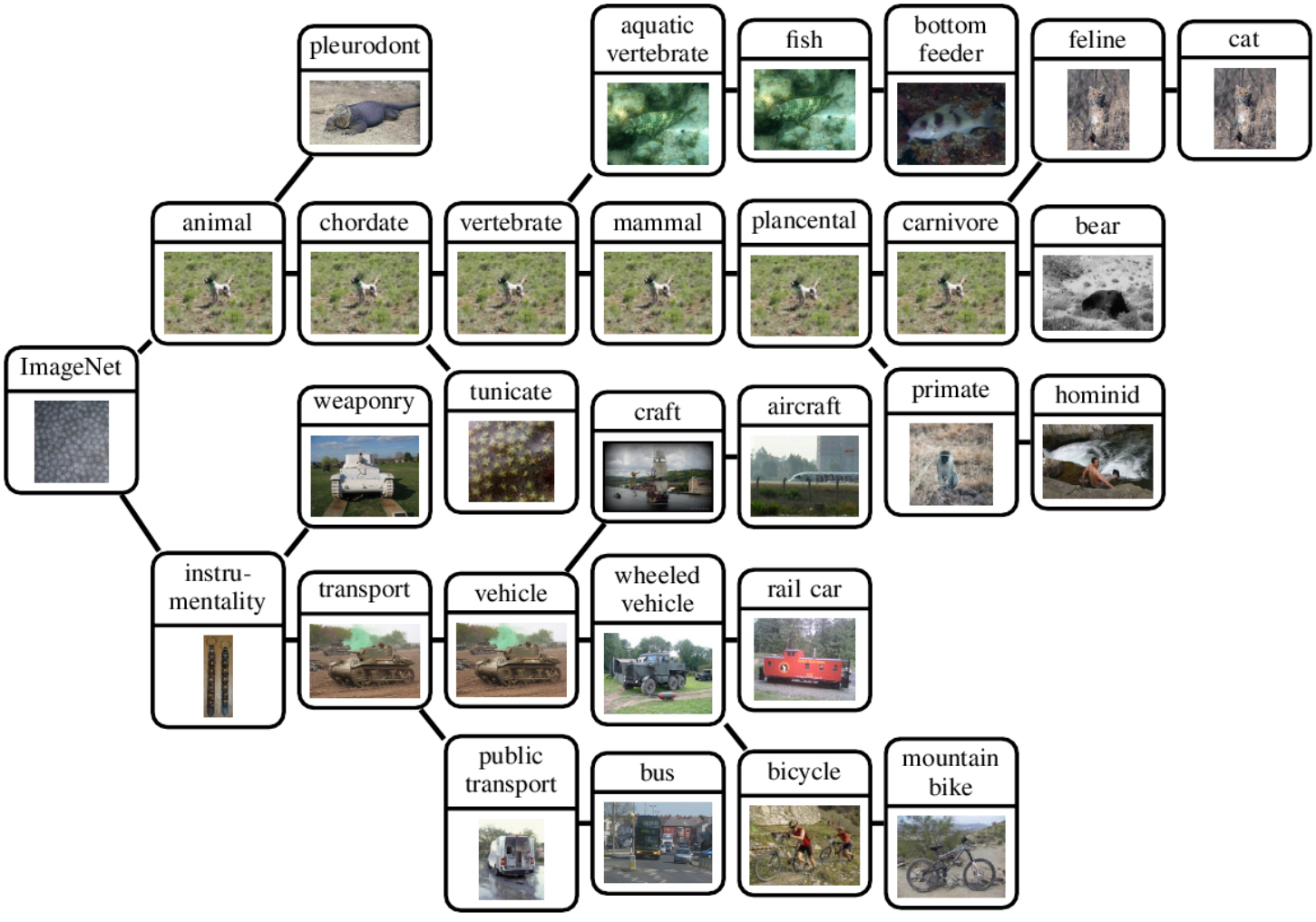


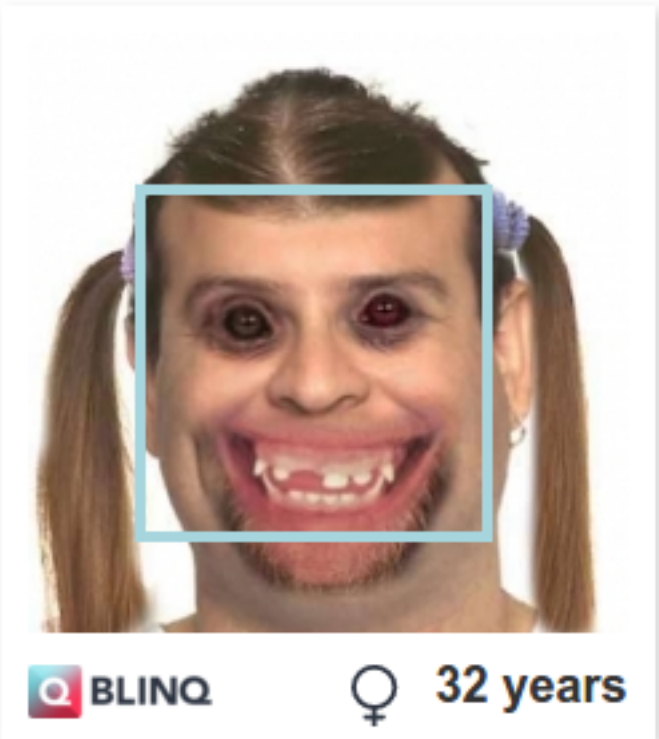
Image by Thomas Deselaers

21 841 Synsets, 14 197 122 Bilder

https://faces.ethz.ch

Let Artificial Intelligence guess your attractiveness and age

#howhot



Hmm.. | Ok | Nice | Hot | Stunning | Godlike

https://howhot.io

Let Artificial Intelligence guess your attractiveness and age

#howhot

BLINQ ♂ 31 years

Hmm.. | Ok | Nice | Hot | Stunning | Godlike

https://howhot.io

Let Artificial Intelligence guess your attractiveness and age

#howhot

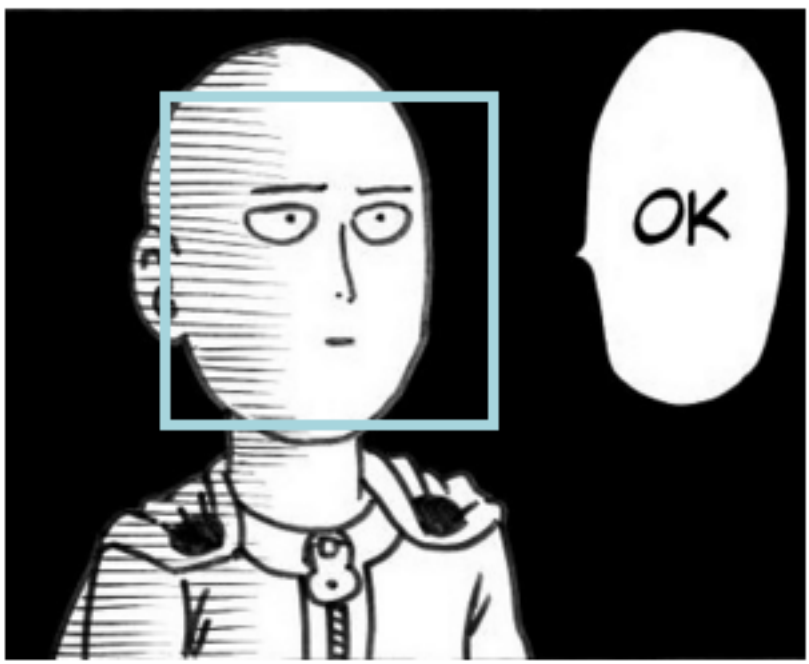
BLINQ

♂ 60 years

Hmm.. | Ok | Nice | Hot | Stunning | Godlike

Let Artificial Intelligence guess your attractiveness and age

#howhot

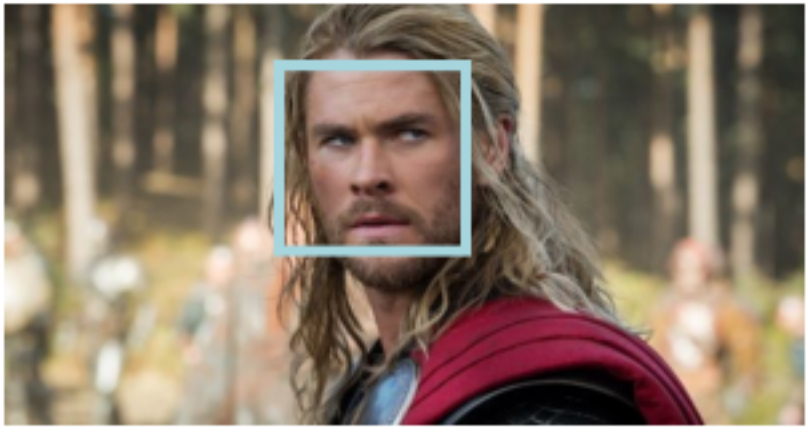


Hmm.. | Ok | Nice | Hot | Stunning | Godlike

https://howhot.io

Let Artificial Intelligence guess your attractiveness and age

#howhot



Hmm.. | Ok | Nice | Hot | Stunning | Godlike

Let Artificial Intelligence guess your attractiveness and age

#howhot

Hmm.. | Ok | Nice | Hot | Stunning | Godlike

https://howhot.io

Let Artificial Intelligence guess your attractiveness and age

#howhot

BLINQ

♂ 61 years

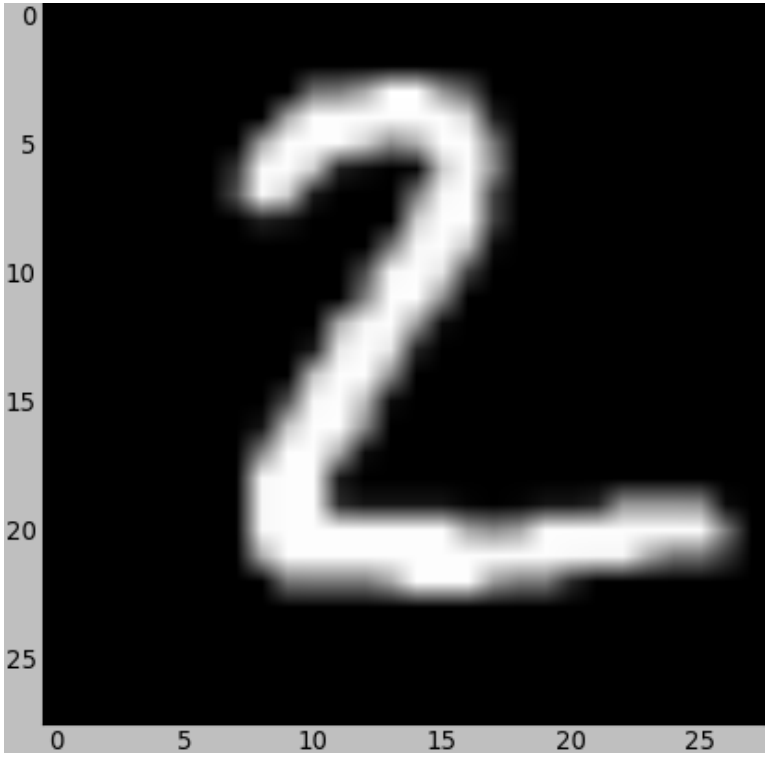
Hmm.. | Ok | Nice | Hot | Stunning | Godlike

Anwendungen

- Klassifikation
  - write-math.com: Symbole erkennen
  - howhot.io: Geschlecht schätzen
  - Bildersuche
- Regression
  - howhot.io, how-old.net: Alter / "Hotness" schätzen
  - Regenmenge / Verkaufszahlen schätzen
- Collaborative Filtering: Netflix-Empfehlungen
- RL: Spiele automatisch spielen (Video 1, 2), Selbstfahrende Autos, Roboterarme
- Weiteres: Deep dream

MNIST - Ziffern klassifizieren

- Klassen: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9
- 60 000 Trainigsdaten, 10 000 Testdaten auf yann.lecun.com/exdb/mnist
- Algorithmen zur Klassifizierung: **SVMs** (Support Vector Machines), **CNNs** (Convolutional Neural Networks), k Nearest Neighbors (siehe tinyurl.com/knn-interact)



Datensatz der Klasse “2”; 28 px $\times$ 28 px

Wie löst man das?

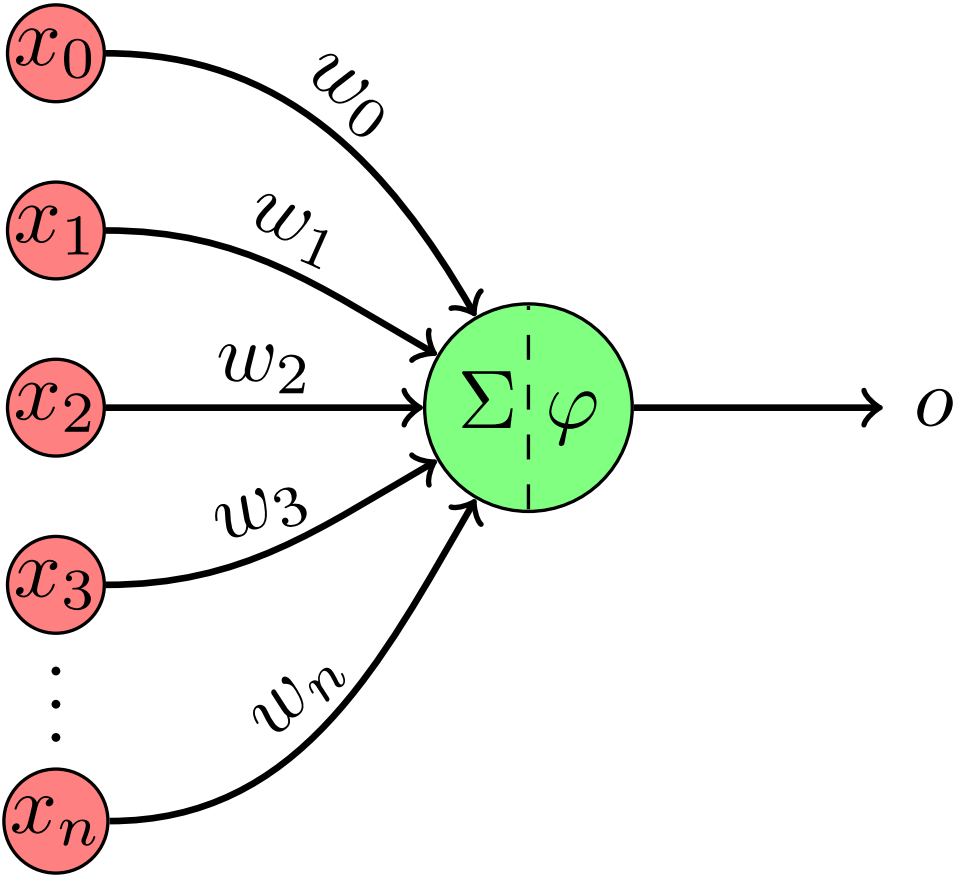
- **Situation**: Daten im $\mathbb{R}^{28\times 28}$, Lösungen in $\{\,0, 1, 2, 3, 4, 5, 6, 7, 8, 9\,\}$

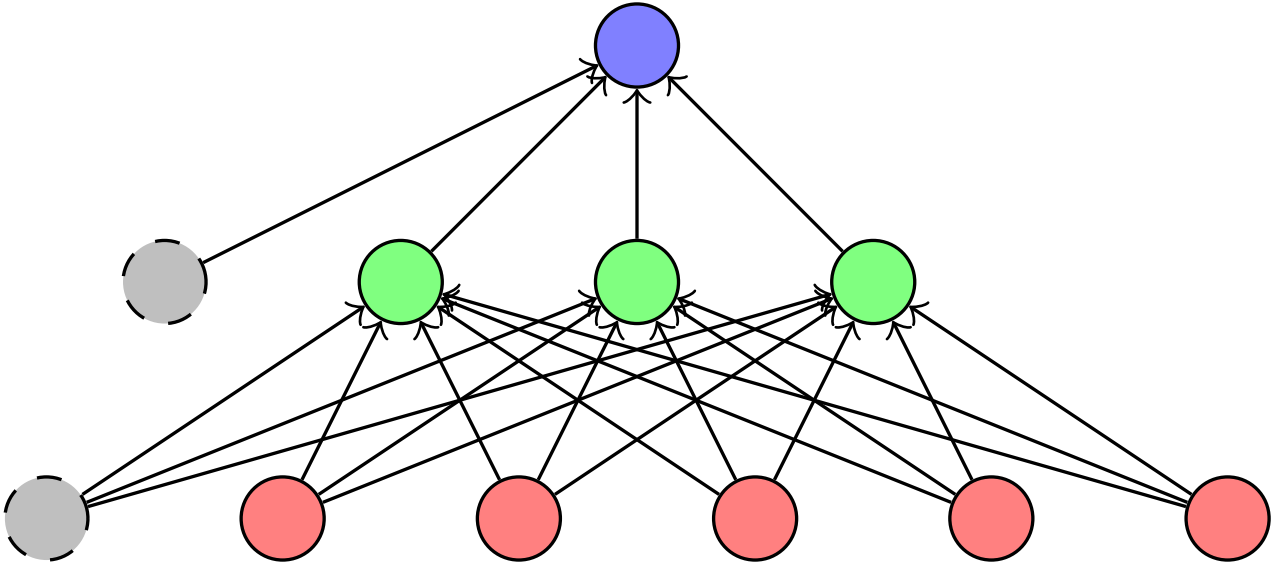
Wie löst man das?

- **Situation**: Daten im $\mathbb{R}^{28\times 28}$, Lösungen in $\{\,0, 1, 2, 3, 4, 5, 6, 7, 8, 9\,\}$
- ... oder $[0, 1]^{10}$, wenn wir eine Wahrscheinlichkeitsverteilung wollen

Wie löst man das?

- **Situation**: Daten im $\mathbb{R}^{28\times 28}$, Lösungen in $\{\,0, 1, 2, 3, 4, 5, 6, 7, 8, 9\,\}$
- ... oder $[0, 1]^{10}$, wenn wir eine Wahrscheinlichkeitsverteilung wollen
- **Lösung**: Neuronale Netze!

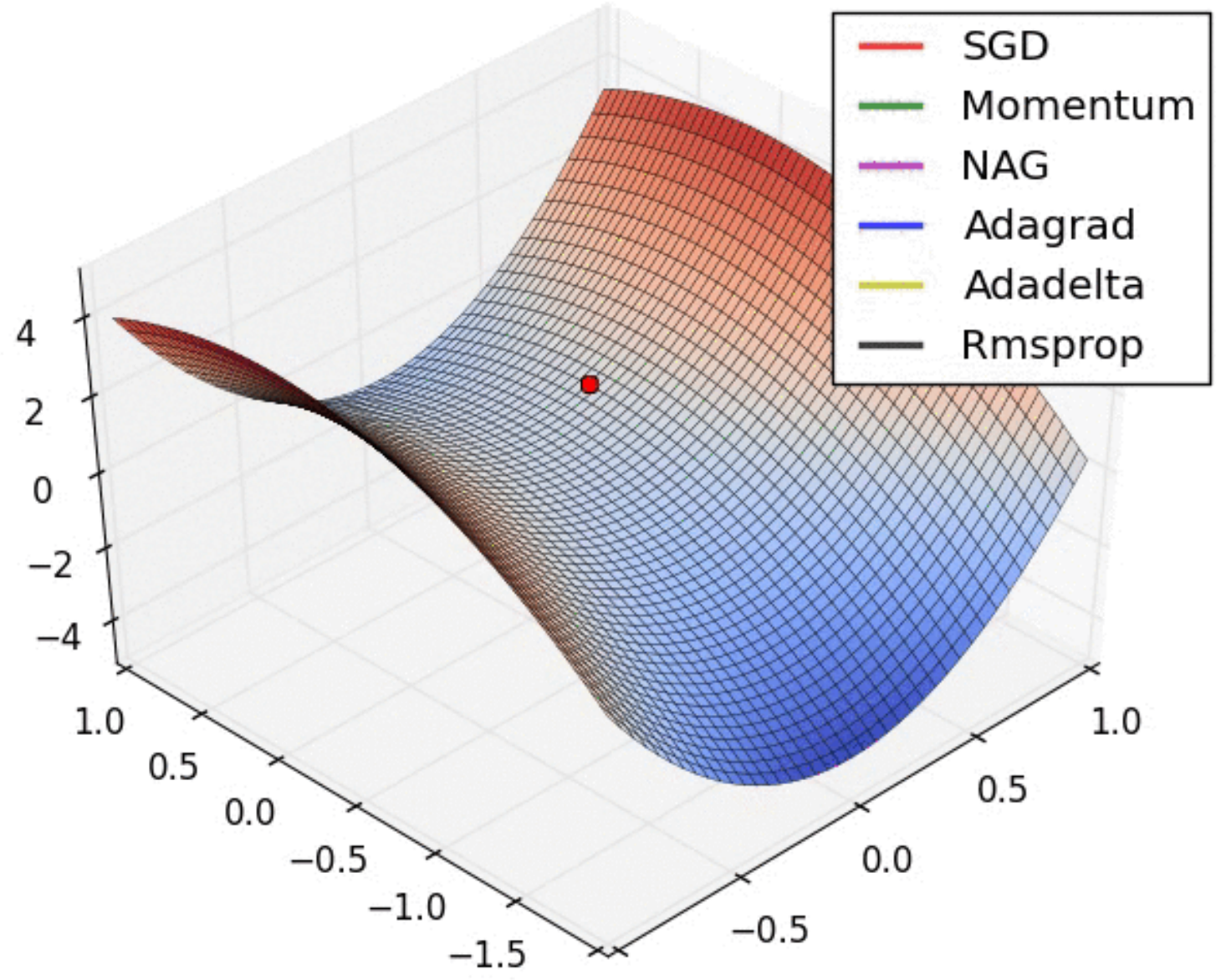

x_0
x_1
x_2
x_3
x_n
w_0
w_1
w_2
w_3
w_n
Σ
φ
o



Wie finden wir die Gewichte?

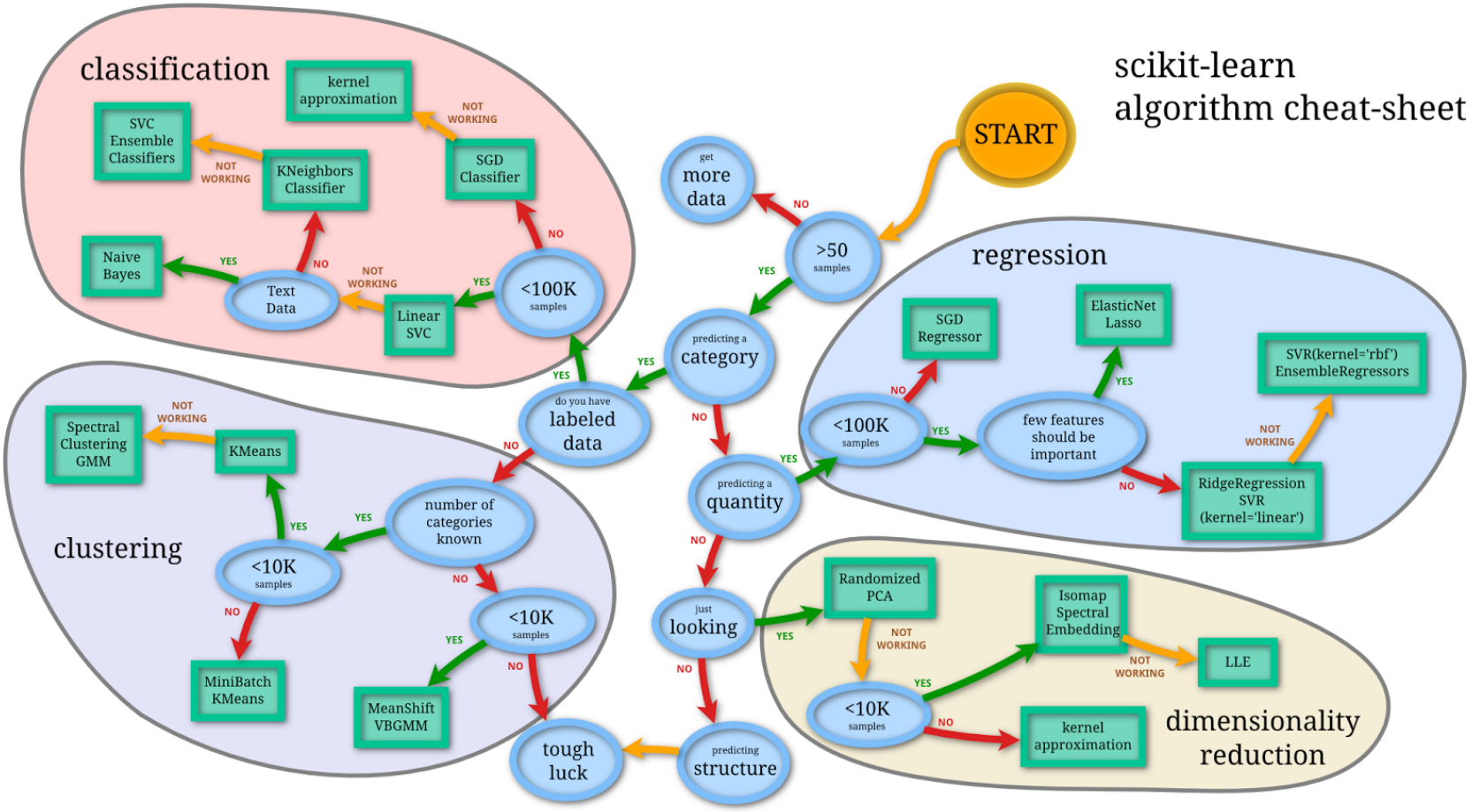
Wie finden wir die Gewichte?
Gradientenabstieg
(Mehrdimensionale Ableitung)

Gradientenabstieg



Quelle: http://imgur.com/a/Hqolp

sklearn


scikit-learn
algorithm cheat-sheet
START
>50 samples
get more data
predicting a category
predicting a quantity
do you have labeled data
just looking
predicting structure
tough luck
classification
kernel approximation
SVC Ensemble Classifiers
KNeighbors Classifier
SGD Classifier
Naive Bayes
Text Data
Linear SVC
<100K samples
clustering
Spectral Clustering GMM
KMeans
number of categories known
<10K samples
MiniBatch KMeans
MeanShift VBGMM
regression
SGD Regressor
ElasticNet Lasso
SVR(kernel='rbf') EnsembleRegressors
few features should be important
RidgeRegression SVR (kernel='linear')
dimensionality reduction
Randomized PCA
Isomap Spectral Embedding
LLE
kernel approximation
YES
NO
NOT WORKING

Weitere Tools

- TensorFlow (Tutorials)
- TensorBox basiert auf TensorFlow, Lokalisierung (Computer Vision)
- Keras: Sehr einfach zu bedienen, abstrahiert von TensorFlow
- Datenvisualisierung
  - Pandas
  - Seaborn