

Grundlagen des maschinellen Lernens

github.com/ML-KA/presentations

Martin Thoma

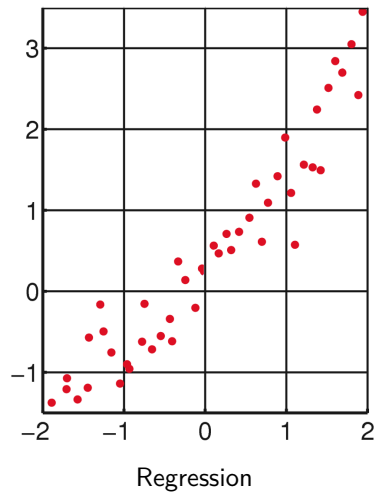
28. Oktober 2015

Was ist Machine Learning?

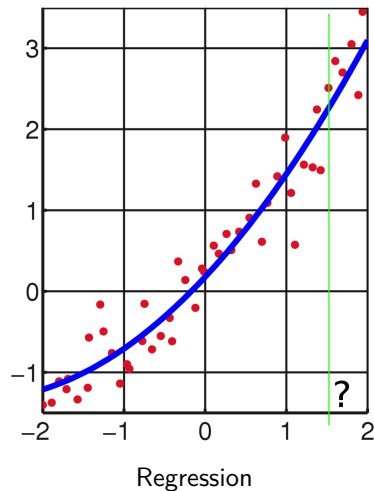
Definition by Tom Mitchell: ML

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 .

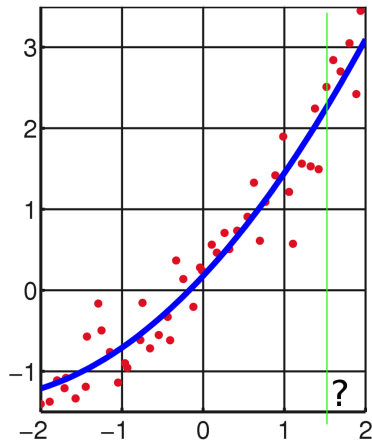
Problemtypen



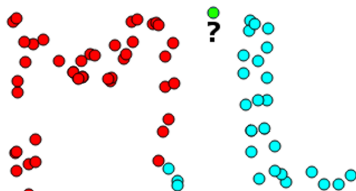
Problemtypen



Problemtypen

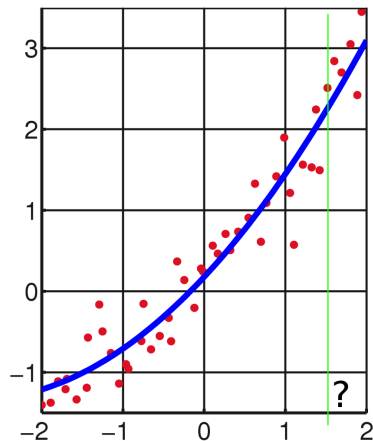


Regression

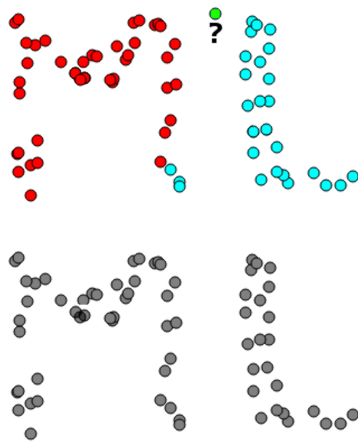


Klassifikation (überwacht)

Problemtypen

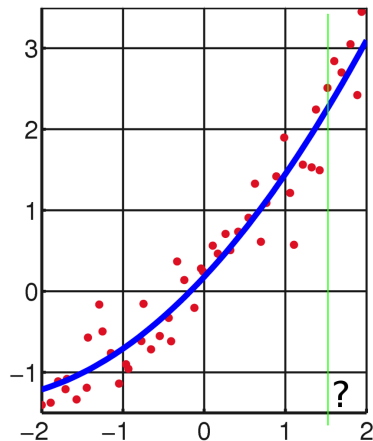


Regression

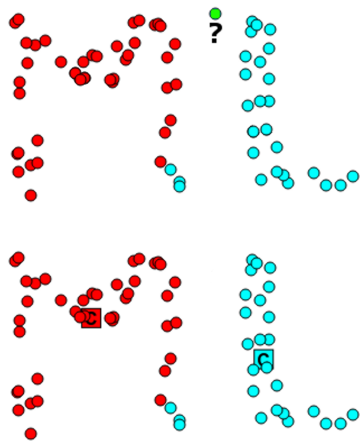


und Clustering (unüberwacht)

Problemtypen



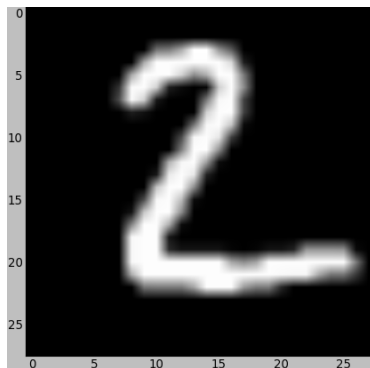
Regression



und Clustering (unüberwacht)

MNIST - Ziffern klassifizieren

- Klassen: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9
- 60 000 Trainingsdaten, 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

ImageNet

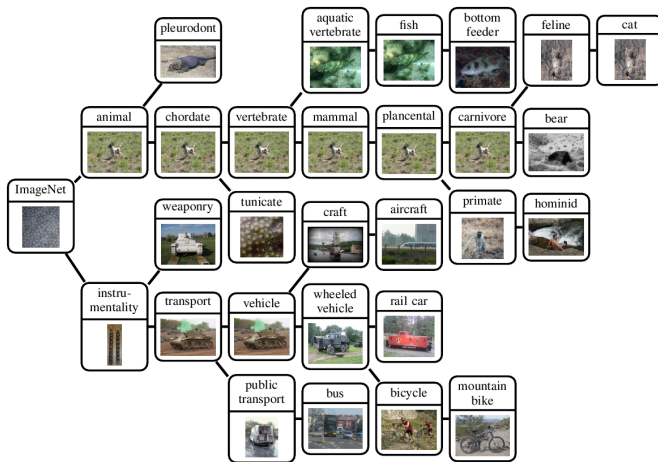


Image by Thomas Deselaers
21 841 Synsets, 14 197 122 Bilder

- In der Klassifikation: Tupel (X, y) , wobei $X \in \mathbb{R}^n$ ein Feature-Vektor, $y \in \{1, \dots, k\}$ das Label und k die Anzahl der Klassen ist.
- Skalenniveaus
 - Nominal: Namen, Geschlecht
 - Ordinal: Konfektionsgrößen
 - Intervall: Anfangszeit einer Veranstaltung
 - Verhältnis: Temperatur in K
 - Absolut: Anzahl Personen
- Zeitreihen, Mengenwertige Daten
- Datenmenge: “There is no data like more data”

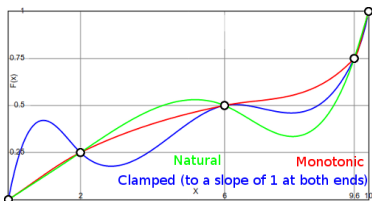
vgl. Vorlesung “Mustererkennung”

- Wie bekomme ich meine Features $X \in \mathbb{R}^n$?
- Bilder: Pixel-Werte
 - Kleiner Skalieren? Rotieren?
 - Farbraum? (z.B. RGB, HSV, HSL, HSI)
 - Normalisieren auf $[0, 1]$
- Verhältnis zweier Größen
- Deep Learning: Auto-Encoder

vgl. Vorlesung “Neuronale Netze”

Generalisierung und Overfitting

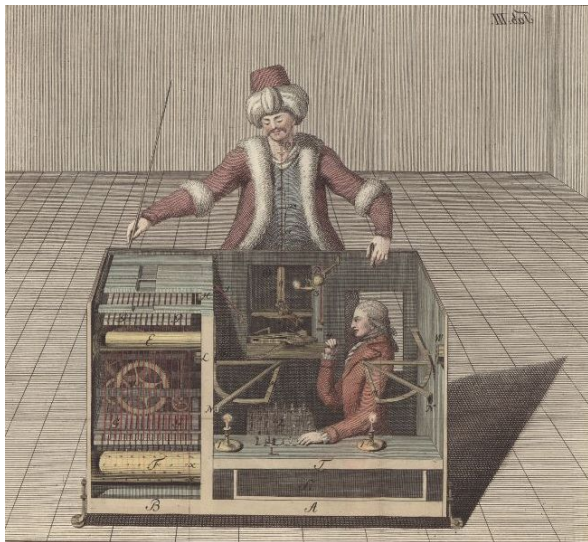
- Generalisierung: Wie gut ist man auf ungesehenen Daten?
- Overfitting: Auswendig lernen



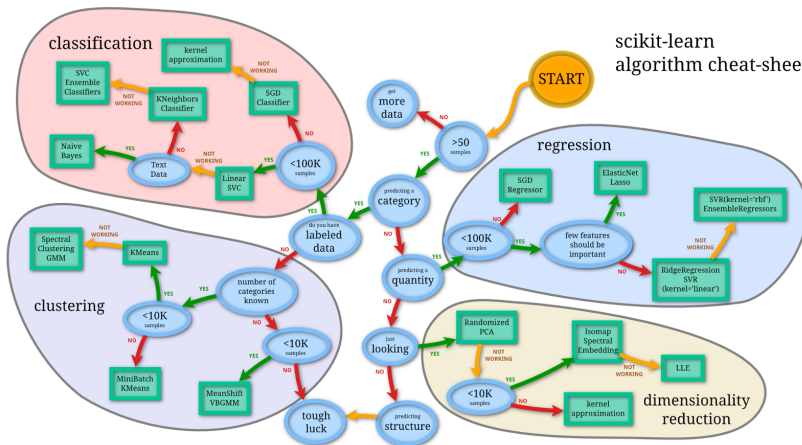
5 Datenpunkte, 3 perfekte Modelle

Trainingsfehler			
		☹	☺
Test- fehler	☹	Verfahren ändern; bessere Daten	Overfitting
	☺	Programmierfehler	Perfekt

Amazon Mechanical Turk



scikit-learn algorithm cheat-sheet



- Neuronale Netze trainieren
- Mit GPU, falls CUDA installiert ist
- github.com/Lasagne
- Lasagne for Python Newbies

Anwendungen: Fotos

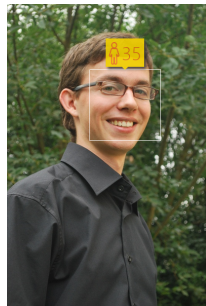
- how-old.net
- “Gelöste” Aufgaben:
 - Gesichter in Bild finden,
z.B. mit Sliding Window
 - Geschlecht klassifizieren: ♂, ♀
 - Regression beim Alter



Wie alt bin ich auf diesem Bild?

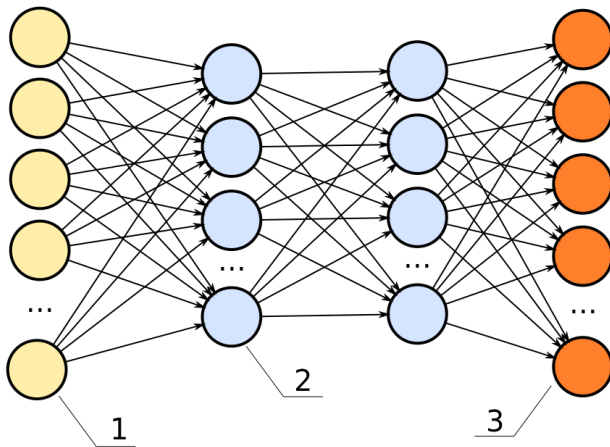
Anwendungen: Fotos

- how-old.net
- “Gelöste” Aufgaben:
 - Gesichter in Bild finden,
z.B. mit Sliding Window
 - Geschlecht
klassifizieren: ♂, ♀
 - Regression beim Alter



20 Jahre alt

Thanks for Your Attention!



- Prof. Dr. Bayerer: Mustererkennung
- Prof. Dr. Waibel:
 - Kognitive Systeme
 - Neuronale Netze (Dr. Kilgour)
- Prof. Dr. Zöllner:
 - Machine Learning 1
 - Machine Learning 2
 - Praktika und Seminare
- Online
 - [Machine Learning: Supervised Learning](#) (Udacity)
 - [Intro to Artificial Intelligence](#) (Udacity)

- Introduction to Machine Learning. 1h 29min
- Machine Learning A Cappella - Overfitting Thriller!. 5min

- [sklearn cheat sheet](#) by Andreas Mueller
- [sklearn cheat sheet](#) by Thomas Deselaers