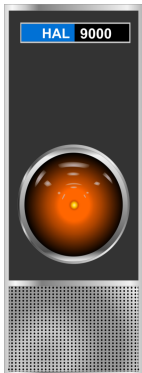
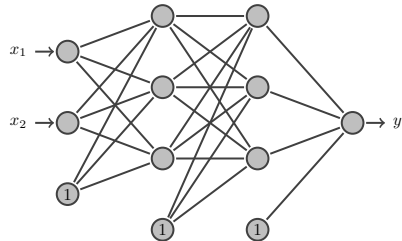


Von Daten und Arten ...

Prof. Dr. Jörg Frochte

Maschinelles Lernen



Strukturierte und unstrukturierte Daten

- Beim maschinellen Lernen haben wir es i. W. mit zwei Arten von Daten zu tun:

Structured Data

Feature Record	f_1	f_2	\dots	f_{n-2}	f_{n-1}	f_n
1						
2						
\vdots						
$m - 1$						
m						

Unstructured Data



Foto von einem Hund

Ich bin der Geist, der stets verneint! Und das mit Recht; denn alles, was entsteht, ist wert, daß es zugrunde geht; Drum besser wär's, daß nichts entstünde. So ist denn alles, was ihr Sünde, Zerstörung, kurz, das Böse nennt, Mein eigentliches Element.

Mephistopheles stellt sich vor (Textauszug aus Goethes Faust)

Supervised learning bzw. überwachtes Lernen

Regression und Klassifikation

Ziel: Lerne eine Funktion $f : X \rightarrow Y$

Regression

Zielmenge ist $Y \subset \mathbb{R}^n$ bzw. eine metrische Skala. Beispiele:

- Ausfallwahrscheinlichkeiten im Anwendungsfeld Predictive Maintenance abschätzen
- Vorhersage von Nachfragen, z. B. im Bereich Car- oder Fahrrad-Sharing

Klassifikation

Zielmenge Y ist diskret und kategorial.

Die Merkmale stammen hingegen aus verschiedenen Skalen. Sie entsprechen z. B. den Längen von unterschiedlichen Blütenblättern.



Wichtig:

Die Werte in $X \times Y$ sind i. A. nicht exakt. Statistische Fehler sind die Regel und systematisch möglich.

Supervised learning bzw. überwachtes Lernen

Labeling ist arbeitsintensiv, Balance ist wichtig

Labeling ist arbeitsintensiv

- Die Herausforderung beim überwachten Lernen sind die Zielwerte.
- Zielwerte zu erstellen ist aufwendig oder man braucht Experten.
- Die Qualität der Daten hängt von den Menschen ab, die diese **labeln**.

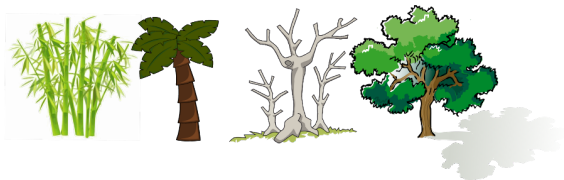


Balance ist wichtig

- Populäre Haustiere (in D): Katzen, Hunde & Kaninchen
- Ein Algorithmus soll mit Hilfe von Internetbildern trainiert werden, diese Tierarten zu unterscheiden.
- Problem: Nehmen wir alle Bilder, die wir finden können, sind die Kaninchen benachteiligt, weil durch die geringe Anzahl ihre Vorhersage nicht so wichtig ist und die wenigen Bilder vielleicht nicht alle Situationen darstellen.



Unüberwachtes Lernen



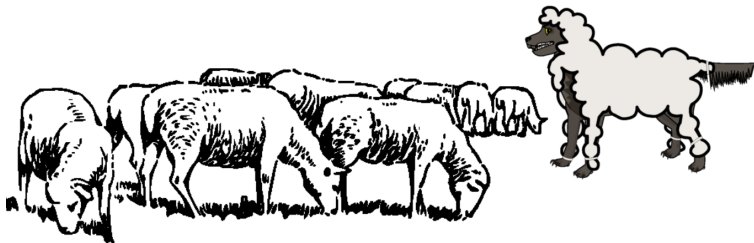
Fragestellung: *Welche 2 oder 3 Abbildungen links bilden eine Gruppe?*

- Ziel ist hier die Bildung von Gruppen bzw. Teilmengen.
- Dies geschieht ohne Labels in den Records. Wir haben quasi nur X vorliegen.
- Hierzu nutzt man die Ähnlichkeit zwischen Elementen $x, y \in X$ als ein wichtiges Kriterium.
- Diese werden über Metriken definiert, wie man sie aus der linearen Algebra kennt.
- Dieses Konzept wird dann eingesetzt, um eine große Ähnlichkeit innerhalb der Gruppe und möglichst große Unterschiede zwischen den Gruppen zu erreichen.

Unüberwachtes Lernen

Outlier Detection

- Ein besonderer Anwendungsfall ist die so genannte Outlier Detection.
- Hier geht es um die Identifizierung von Datensätzen, die nicht mit einem erwarteten Muster oder anderen Elementen in einer Datenbank übereinstimmen.
- Ein Beispiel ist das *Rauschen*, was eben keiner Gruppe zugeordnet wird.



Bestärkendes Lernen bzw. Reinforcement Learning

- Beim bestärkenden Lernen steht nicht die Information über die richtige Lösung zur Verfügung,
- sondern lediglich ein Signal

$$r : s \mapsto \mathbb{R},$$

welches eine Belohnung/Bestrafung für den erreichten Zustand s wiedergibt.

- Auf dieser Basis gilt es, eine optimale Strategie für einen Agenten zu finden.
- Für viele Ansätze ist dabei die Approximation einer Nutzenfunktion nötig, die wiederum überwacht Lernen erfordert.



Nein, der Roboter leidet unter negativem Feedback nicht

