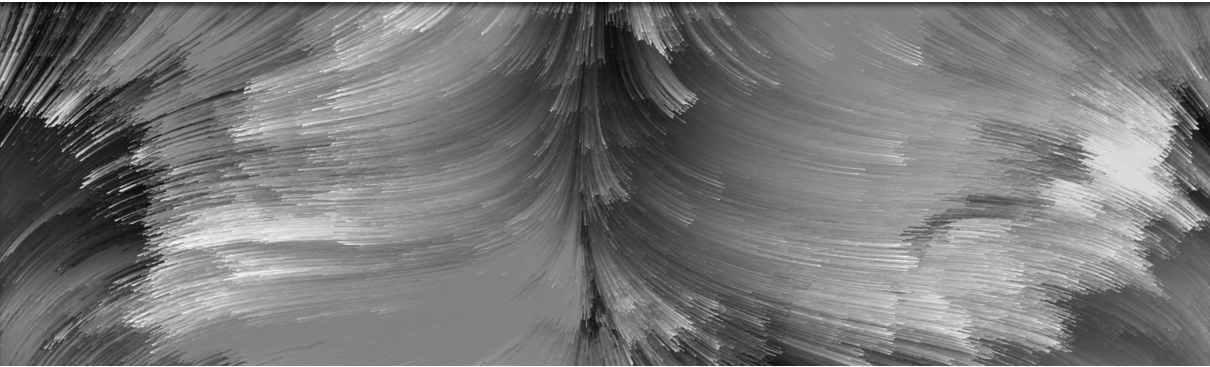
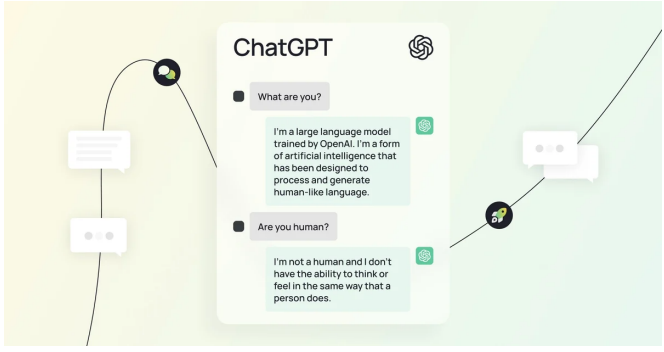


Künstliche Intelligenz (Sommersemester 2024)

Kapitel 02: Machine Learning – Grundlagen

Prof. Dr. Adrian Ulges

Verblüffung...



"Machine intelligence is the last invention that humanity will ever need to make."

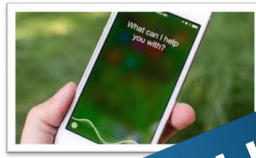
(Nick Bostrom, "Superintelligence")

Maschinelles Lernen (ML): Bereiche aktueller Erfolge image: [4]

autonomous vehicles

dialog systems

speech recognition



predictive maintenance

medical diagnosis

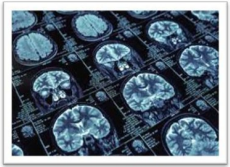


image recognition

machine

ML: Anwendungsfelder

kaggle

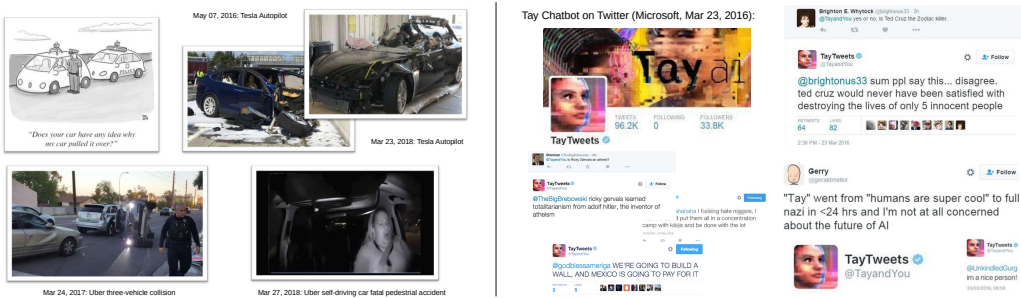
'Klassische' Anwendungsfelder

- **Computer Vision**: Handschrifterkennung, Objekterkennung, ...
- **Nutzermodellierung**: Suchmaschinen, Empfehlungssysteme, Targeting, ...
- **NLP**: Informationsentnahme, Stimmungsanalyse, Spam, ...

Sonstige Anwendungsfelder ...

- Restaurant-Umsatzprognose
  Vorhersage des Jahresumsatzes von zu eröffnenden Restaurants
- Fahrertelematik-Analyse (AXA)
  anhand von GPS-Routen den Fahrer eines Autos identifizieren
- Wal-Erkennung
  Walgesänge in Audio erkennen, Kollisionen mit Schiffsverkehr verhindern
- ...

ML: Misserfolge image: [4]



- ML kommt zunehmend in sicherheitskritischen Anwendungen zum Einsatz: selbstfahrende Fahrzeuge, Gesundheit, Pharmazie, Aktienhandel ...
- Die regulatorischen Auswirkungen sind enorm *(Die Genauigkeit ist begrenzt)*!
- ML-Modelle sollten sicher, fair, transparent, ressourceneffizient, datenschutzgerecht sein.

Maschinelles Lernen: Definition

"The field of study that gives computers the ability to learn **without being explicitly programmed.**"

(Arthur Samuel (1959))

"A computer program is said to **learn** *from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with Experience E."*

(Tom Mitchell (1998))

Eine ML Beispielanwendung images: [1] [2]

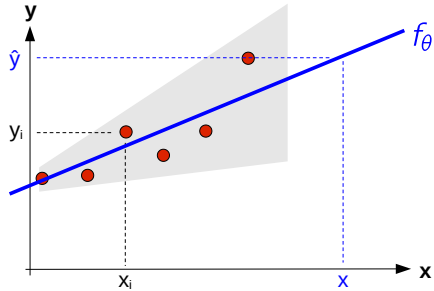
> *"***Jedes If-Statement** *ist eine* **potenzielle Anwendung** *für maschinelles Lernen"*
>
> (Thomas M. Breuel (2004))

- Ein Computersystem soll eine nicht-triviale Entscheidung treffen, z.B. **Spam-Filterung**.
- Warum nicht die Entscheidungslogik hart codieren?

Probleme

- Hoher initialer Verständnisaufwand.
- Schwierig, das bestmögliche Programm zu erreichen.
- Überprüfung der Optimalität ist schwierig.
- Code ist extrem schwierig zu updaten/warten.
- Das Verfolgen von Datendrift ist schwierig, wenn z.B. Spammer ihre Strategien ändern.
- Es gibt keine Möglichkeit, das Feedback der Benutzer zu berücksichtigen.

ML Hello World?



- **Ziel**: Vorhersage des Gewichts einer Person in der Zukunft!
- **Gegeben**: Stichprobe $x_1, ..., x_n$ *(Zeitpunkte)* mit sogenannten "Labels" $y_1, ..., y_n$ *(dem jeweiligen Gewicht der Person)*.
- **Vorgehensweise** (*lineare Regression*):
  - Wir fitten eine Linie f_θ auf die Punkte.
  - Gegeben einen Zeitpunkt x, verwenden wir $\hat{y} := f_\theta(x)$ als Prognose des Gewichts.
- **Ist dies maschinelles Lernen**?

ML Hello World!

- Wir definieren unsere Linie als eine Funktion f mit Parametern $\theta = (a, b)$

$$f_\theta(x) = a \cdot x + b$$

- Wir messen die Qualität einer bestimmten Linie f_θ mit einer Zielfunktion $\mathcal{L}$:

$$\mathcal{L}(\theta) = \sum_{i=1}^{n} \Big(f_\theta(x_i) - y_i \Big)^2$$

- Die beste Linie ist diejenige, die $\mathcal{L}$ minimiert:

$$\theta^* = \arg\min_{\theta \in \mathbb{R}^2} \mathcal{L}(\theta)$$

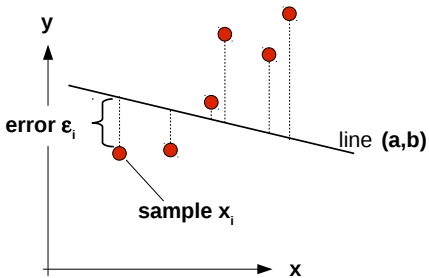
- Wir setzen die partiellen Ableitungen $(\partial\mathcal{L}/\partial a, \partial\mathcal{L}/\partial b)$ gleich null. Es ergibt sich:

$$a^* = \Big(\sum_i y_i x_i - \bar{y} \sum_i x_i \Big) / \Big(\sum_i x_i^2 - \bar{x} \sum_i x_i \Big)$$

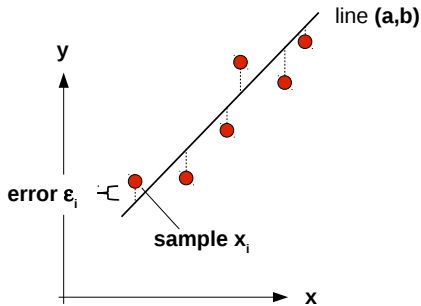
$$b^* = \frac{1}{n} \sum_i y_i - a^* \cdot \frac{1}{n} \sum_i x_i$$

ML Hello World

schlechte Lösung θ: Die Fehler $\epsilon_1^2, ..., \epsilon_n^2$ sind hoch.



gute Lösung θ: Die Fehler $\epsilon_1^2, ..., \epsilon_n^2$ sind niedrig.



ML: Terminologie

- Wir nennen die Punkte $(x_1, y_1), ..., (x_n, y_n)$ die Trainingsdaten.
- Die „wahren“ Werte y_i werden auch als die Labels, Targets oder Grundwahrheit (engl. “ground truth”) bezeichnet.
- Wir nennen unsere Linie $f_\theta(x) = a \cdot x + b$ das Modell.
- Wir bezeichnen den Prozess der Schätzung der Modellparameter $\theta = (a, b)$ als Training oder Fitting.
- Eine typische Trainingsstrategie besteht darin, eine Zielfunktion *(engl. “objective function” oder “loss”)* $\mathcal{L}$ zu optimieren, oft der Form:

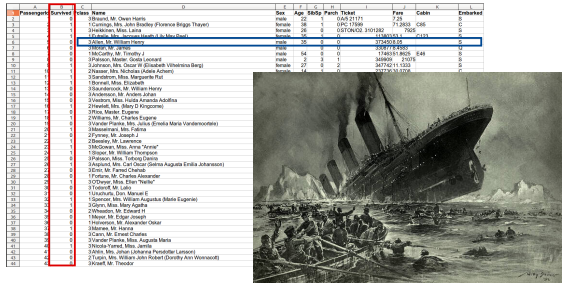
$$\mathcal{L}(\theta) = \frac{1}{n} \sum_i \ell(f_\theta(x_i), y_i)$$

Anmerkungen

- Praktische Modelle haben deutlich mehr Parameter *(GPT-3.5: $\#\theta = 175$ Mrd.)*.
- Im obigen Beispiel haben wir die Lösung manuell abgeleiten können (wir sagen: es gibt eine *analytische* Lösung). In der Praxis ist $\mathcal{L}$ meist schwieriger zu optimieren, und die Optimierung wird per lokaler Suche durchgeführt.

ML ist multi-variat! image: [3]

- ML soll also Prognosen y über Eingabeobjekte x treffen.
- In der Praxis sind x und y keine Skalare, sondern Vektoren **x** und **y**!



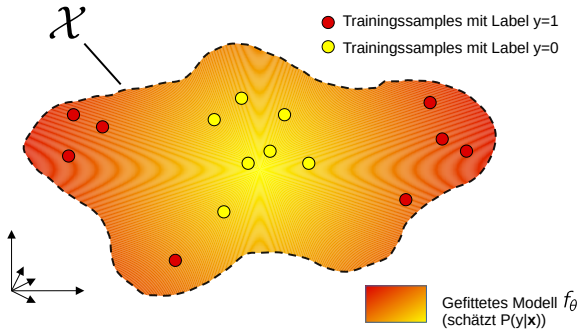
- Unser Modell wird zu einer multivariaten Funktion $f_\theta : \mathcal{X} \to \mathcal{Y}$, wobei $\mathcal{X} \subseteq \mathbb{R}^d$ und $\mathcal{Y} \subseteq \mathbb{R}^{d'}$.
- Wir bezeichnen die Einträge des Merkmalsvektors **x**, z.B. Geschlecht, Alter ..., als Merkmale (engl. "features"), und nennen $\mathcal{X}$ den Merkmalsraum (engl. "feature space").

ML ist multi-variat (cont'd)

Anmerkungen

- Im Allgemeinen können viele Merkmale für das Zielproblem irrelevant sein. Während des Trainings müssen ML-Modelle die relevanten auswählen.
- Ein Merkmal kann auch erst in Kombination mit anderen Merkmalen nützlich sein.

ML: Der Merkmalsraum ("Feature Space")



- Merkmalsvektoren **x** können als Punkte im Merkmalsraum $\mathcal{X}$ interpretiert werden.
- Wir können uns ein Modell $f_\theta : \mathcal{X} \to \mathbb{R}$ als einen Farbverlauf vorstellen, der jedem Punkt **x** einen Wert $f_\theta(\mathbf{x})$ zuweist.
- Im obigen Beispiel schätzt das Modell f_θ die Klassenzugehörigkeit eines Objekts **x**, d.h. $f_\theta(\mathbf{x}) \approx P(Y{=}1 \,|\, X{=}\mathbf{x})$.

References I

[1] Brizzle born and Bread.
https://www.flickr.com/photos/brizzlebornandbred/5292576151/ (retrieved: Oct 2016).

[2] Spam (Monty Python).
https://en.wikipedia.org/wiki/Spam_(Monty_Python) (retrieved: Oct 2016).

[3] 'Untergang der Titanic' Illustration von Willy Stöwer für die Zeitschrift Die Gartenlaube.
https://de.wikipedia.org/wiki/RMS_Titanic (retrieved: Oct 2016).

[4] Damian Borth.
Machine Learning (M.Sc. Course), University St. Gallen, summer term 2022.
(retrieved: Aug 2022).