



Agenda

- 1** Einführung
- 2** BI Strategie & Management
- 3** Data Warehouses & OLAP
- 4** ETL-Prozesse und Tools
- 5** Kommunikation und Visualisierung
- 6** Dashboards & Self-Service BI
- 7** Vorgehensmodelle BA
- 8** Klassifikation I
- 9** **Klassifikation II**
- 10** Wirkungsprognosen
- 11** Modellbeurteilung & ML-Pipelines
- 12** Zeitreihenanalyse und –vorhersage
- 13** Nichtüberwachte Verfahren
- 14** Wrap-up und Q&A



Heutige Agenda



Inhalte: Anwendungsfälle und Methoden zur Klassifikation



Lernziele:

- a) Grundlagen Entscheidungsbäume
- b) Grundlagen neuronale Netze



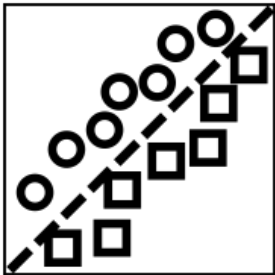
2-Token Leuphana Language Model (LLM)



Algorithmen – Versuch einer Übersicht

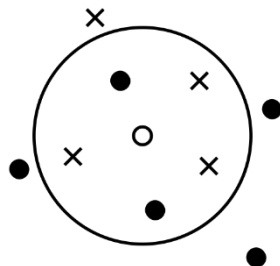
Linear

Annahme: „lineare“
Hyperebene trennt
Gruppen



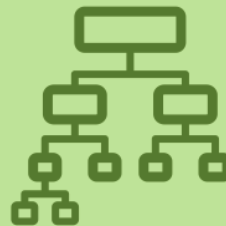
Instance-Based

Klassifikation erfolgt
aufgrund der Klassen
der Nachbarn (Lazy
Learning)



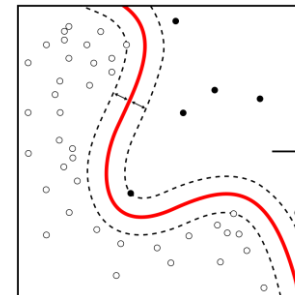
Tree-Based

Klassifikation anhand
eines Entscheidungs-
baums



Kernel-Based

Verwendung nicht-
linearer
Transformationen



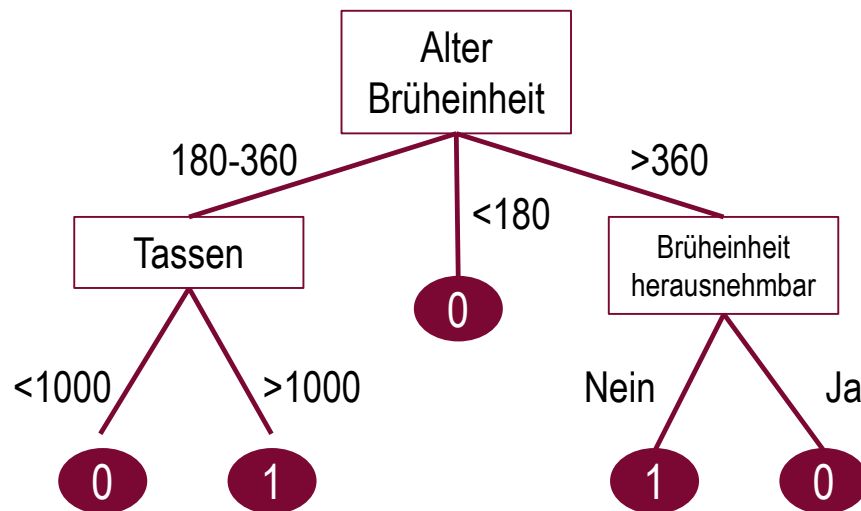
Neurale Netze

Verbindung einfacher
Funktionen zu
komplexen Netzen





Entscheidungsbäume



Algorithmus:

- Entscheidungsbäume modellieren
Entscheidungen in einer baumähnlichen Struktur. Jeder Knoten repräsentiert ein "Kriterium", während die Kanten die möglichen Antworten repräsentieren.
- Verschiedene Ansätze für die Auswahl des jeweiligen Splitkriteriums möglich

Bewertung:

- Sehr flexibel, da EB bis auf einzelne Datenpunkte unterteilen können
- Hohe Flexibilität birgt Gefahr des Overfittings
- EB sind häufig Grundlage für weitergehende Ansätze (s. Ensemble Verfahren)



Entscheidungsbäume – ein klassisches Beispiel

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Tom Mitchell (1998)



Lernalgorithms – Erstellung von Entscheidungsbäumen



Entropie als Maß für den Klassifikationsfehler

- Leitfrage: Wie lässt sich der Fehler / die Reinheit eines Blattes des EBs quantitativ fassen? Missklassifikationsrate?
- Aus der Physik (und später der Informationstheorie) wird häufig die sog. Entropie verwendet

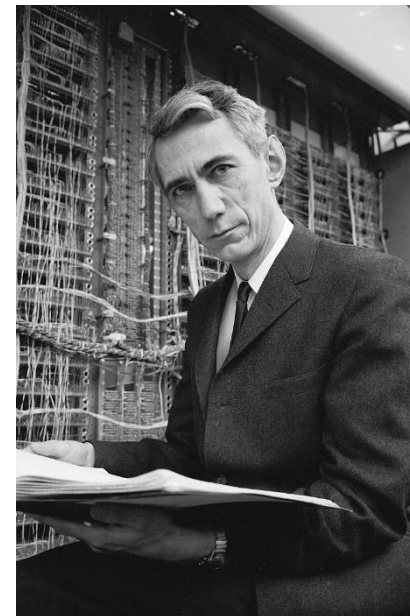
$$H(X) = - \sum_x p(x) \log p(x)$$

- Wie berechnet man die Entropie im Falle der Klassifikation?

$$H(S) = -p_+ \log_2 p_+ - p_- \log_2 p_-$$

wobei $p_{+/-}$ die Verhältnisse zw. +/- und allen Fällen sind

Claude Shannon





Information und Entropie

$$H_1 = 4.76 \text{ bits/char}$$

1. *Zero-order approximation.* (The symbols are independent and equiprobable.)

XFOML RXKHRJFFJUJ ZLPWCFWKCYJ
FFJEYVKCQSGXYD QPAAMKBZAACIBZLHJQD

$$H_2 = 4.03 \text{ bits/char}$$

2. *First-order approximation.* (The symbols are independent. Frequency of letters matches English text.)

OCRO HLI RGWR NMIELWIS EU LL NBNESEBYA TH EEI
ALHENHTTPA OOBTTVA NAH BRL

•
•
•

$$H_2 = 2.8 \text{ bits/char}$$

5. *Fourth-order approximation.* (The frequency of quadruplets of letters matches English text. Each letter depends on the previous three letters. This sentence is from Lucky's book, *Silicon Dreams* [183].)

THE GENERATED JOB PROVIDUAL BETTER TRAND THE
DISPLAYED CODE, ABOVERY UPONDULTS WELL THE
CODERST IN THESTICAL IT DO HOCK BOTHE MERG.

Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell system technical journal*, 30(1), 50-64.

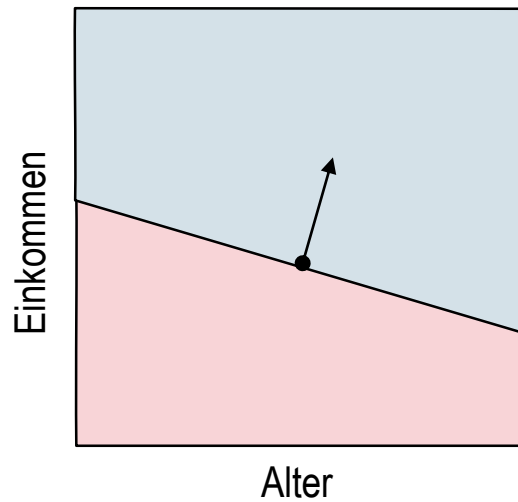


Selektion des Splitkriteriums anhand der Entropie



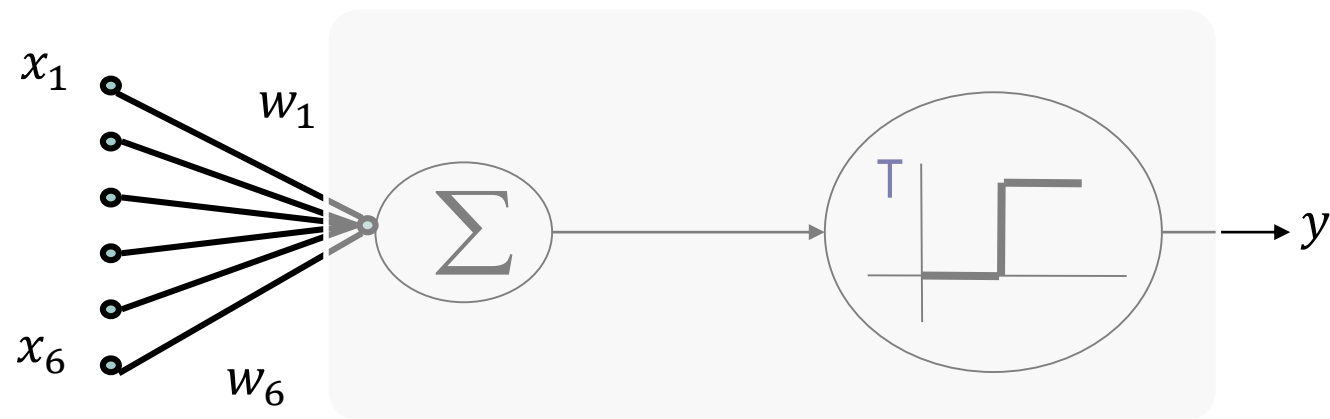
Kreditentscheidung mit dem Perzeptron

$$h(\mathbf{x}) = \text{sign}(w^T \mathbf{x})$$





Aufbau eines Neurons

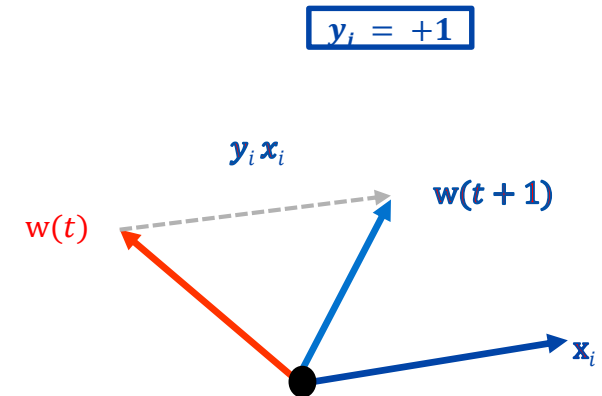


Dan Ross: CS 446: Machine Learning



Perzeptronalgorithmus

Unter der Annahme, dass die Daten linear trennbar sind:



Start with arbitrary $\mathbf{w}(1) = \mathbf{0}$

Do until all training data is correctly classified

Pick any misclassified example (x_i, y_i)

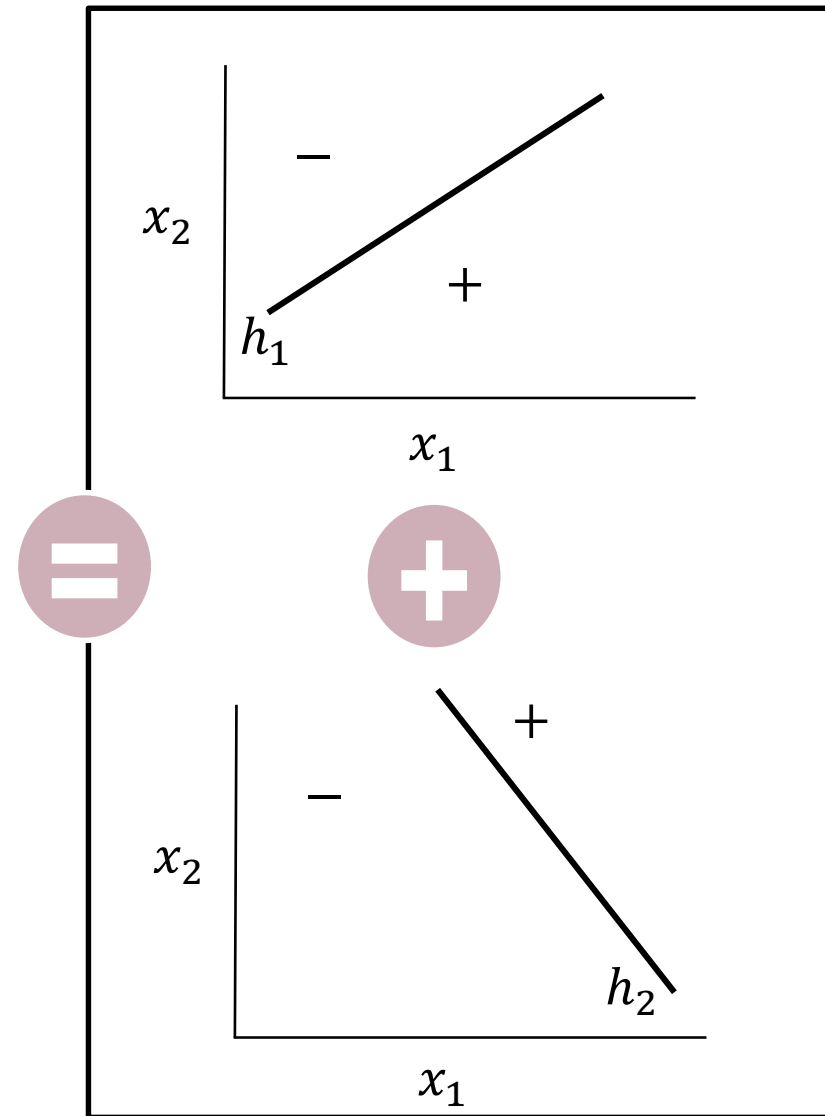
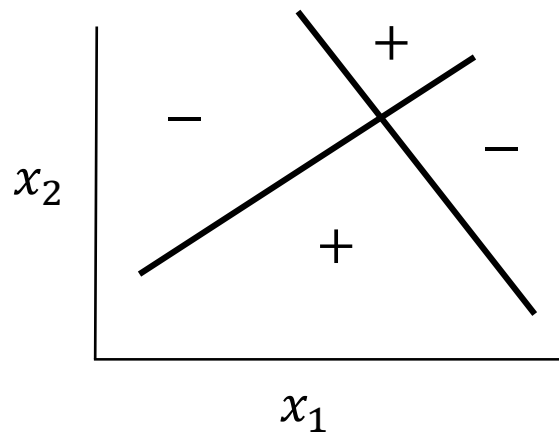
Update the weight:

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \eta y_i x_i$$

PLA implementiert ein schrittweises Lernen, ein zentraler Gedanke im Maschinellen Lernen



Grundlegende Idee von Neuronalen Netzen

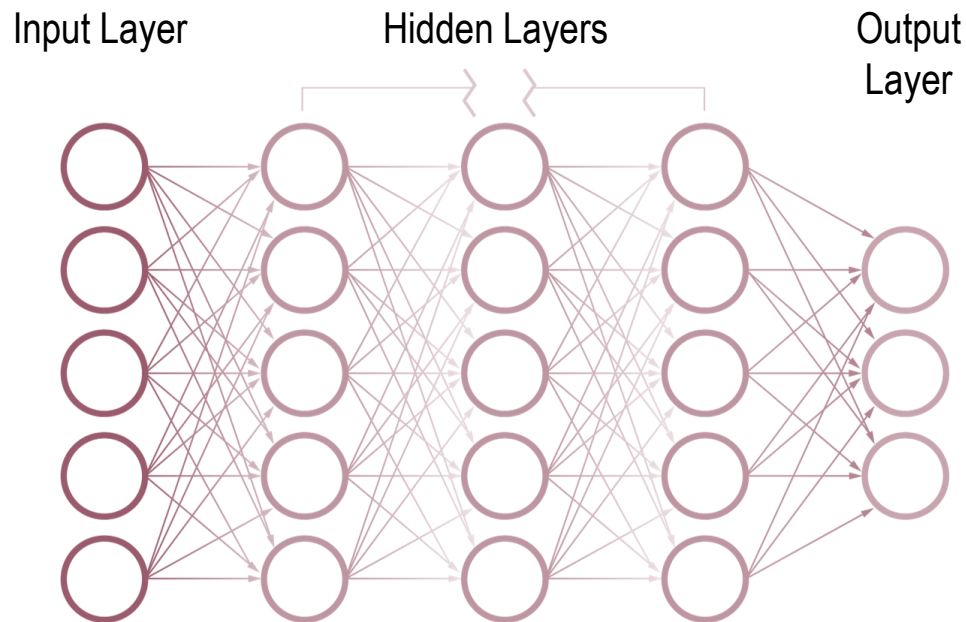




Neuronale Netze als Multi-Layer-Perzeptron



Neuronale Netze



Algorithmus:

- Kombination vieler Neuronen in mehreren Ebenen
- Nicht-lineare Aktivierungsfunktionen
- Parameterschätzungen erfolgen durch Lösung eines Optimierungsproblems (Minimierung der Fehler)

Bewertung:

- State-of-the-art Performance
- Ergebnisse sind schlecht interpretierbar
- Lange Trainingszeiten und hohe Rechenleistungen erforderlich



Zusammenfassung: Relevante Aspekte bei der Modellwahl

1 Datenmenge: Sowohl die Anzahl Beobachtungen (N) als auch Features (d)

2 Angenommene Komplexität der Zusammenhänge (linear vs. nicht-linear)

3 Verfügbare Rechenleistung zum Training und zur Vorhersage

4 Anforderung an die Interpretierbarkeit der Ergebnisse

5 Output: binäre vs multikategoriale Klassifikation



Aufgaben und Quellen

Aufgaben

- Lesen Sie das Han, Kamber und Pei (2011) Kapitel 1.4.3 zur Übersicht, Kapitel 8.2.1-8.2.2 zu Decision Trees
- Lesen Sie Provost & Fawcett (2013) Kapitel 4.3 und 4.4

Quellen

- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining Concepts and Techniques*. Waltman, MA, USA: Morgan Kaufmann Publishers.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business*. O'Reilly Media, Inc.