Foto: Randy Fath auf Unsplash

Training von Klassifikationsbäumen

… und Overfitting

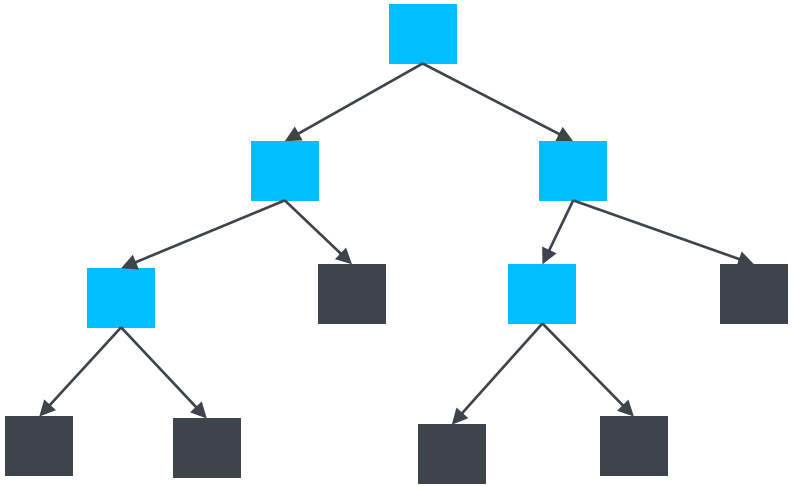
INNOVET
KI B³
Künstliche Intelligenz in die berufliche Bildung bringen

GEFÖRDERT VOM
INNOVET
Bundesministerium für Bildung und Forschung
bibb Bundesinstitut für Berufsbildung

Gefördert als InnoVET-Projekt aus Mitteln des Bundesministeriums für Bildung und Forschung.

Wie baut man einen Entscheidungsbaum?

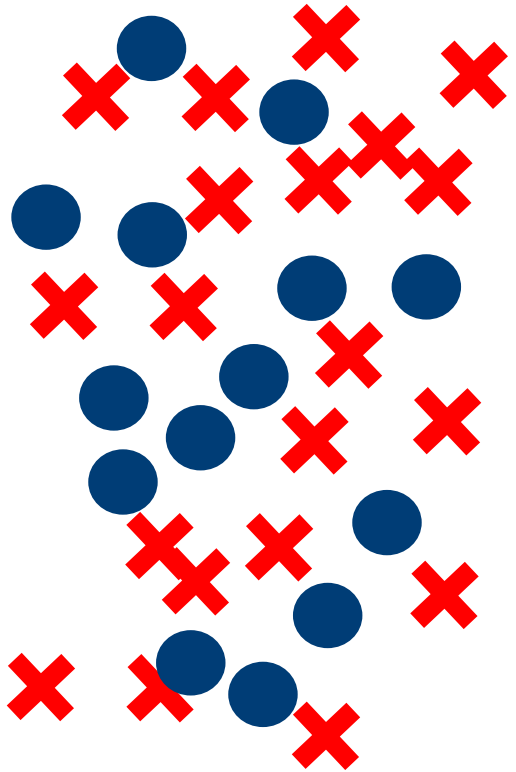
- Zwei Teilprobleme:
  - Wie baut man Zweige?
  - Wie baut man Blätter?
- Zweige = verzweigende Knoten
- Blätter = nicht verzweigende Knoten, Terminalknoten
- Start immer oben, also zuerst: Zweige



Wie baut man Zweige?

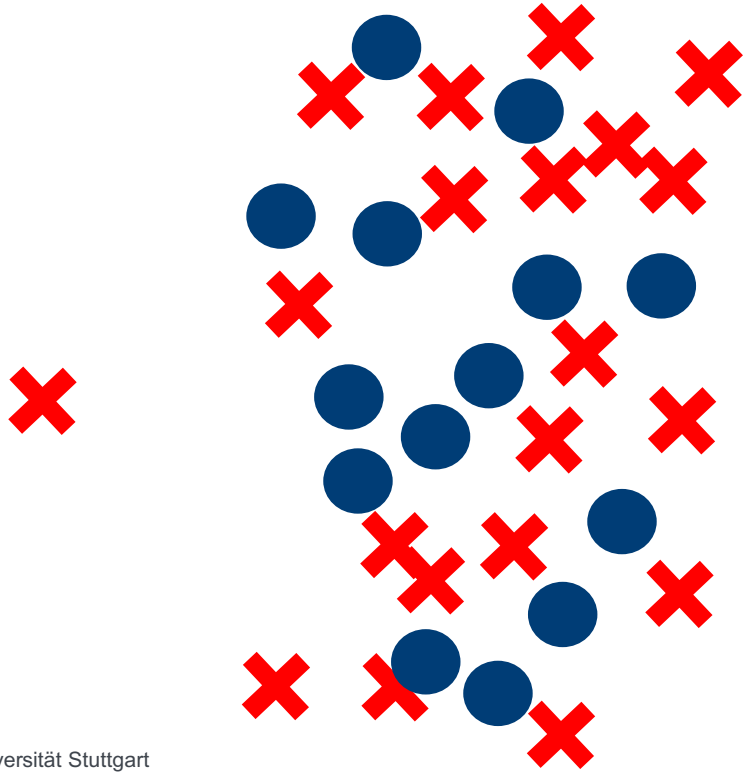
- Gegeben: Datenpunkte
- Gesucht: Kriterium für Aufteilung: ein Feature und ein Schwellwert
  z.B. Temperatur <= 16 Grad
- Gehe alle Features und alle möglichen Schwellwerte durch
  - Wie würden die Daten in zwei Mengen aufgeteilt?
  - Welche Entropien hätten die Verteilungen der Klassen in beiden Mengen jeweils?
  - Welche Entropie ergibt das insgesamt (gewichtet nach Häufigkeit)?

Beispiel aus dem Jupyter-Notebook



- Eis-Beispieldaten: 20 Fälle von “geschlossen“, 14 Fälle von „offen“
- Aufteilung nach Temperatur
- Niedrigster möglicher Schwellwert: -0.5 Grad
- Ergibt Aufteilung in zwei Untermengen:
  - Untermenge <= -0.5: 1 mal geschlossen
  - Untermenge > -0.5: 19 mal geschlossen, 14 mal offen

Beispiel aus dem Jupyter-Notebook

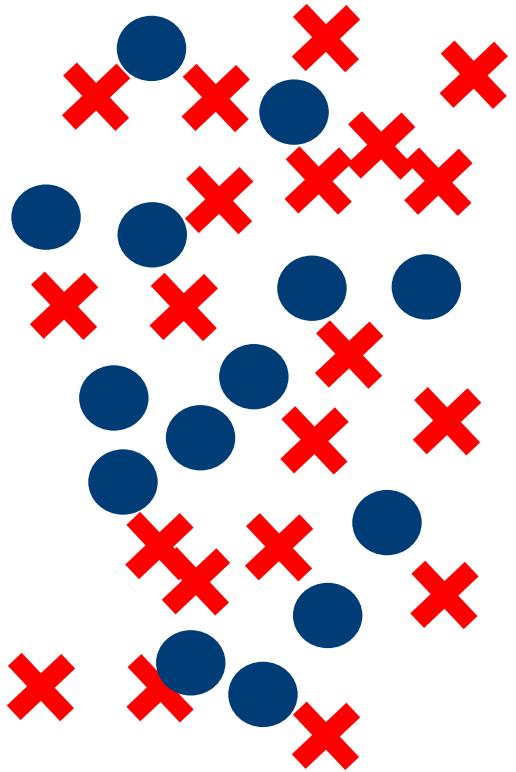


- Entropie für Untermenge 1:
  - 0
- Entropie für Untermenge 2:
  - Verteilung 19:14
  - Insgesamt 33 Datenpunkte
  - $19/33 * - \log_2(19/33) + 14/33 * - \log_2(14/33)$

Hier zur Basis 2 wie in sklearn

- Gewichtete Entropie:
  $1/34 * 0 + 33/34 * 0.98 = 0.954$

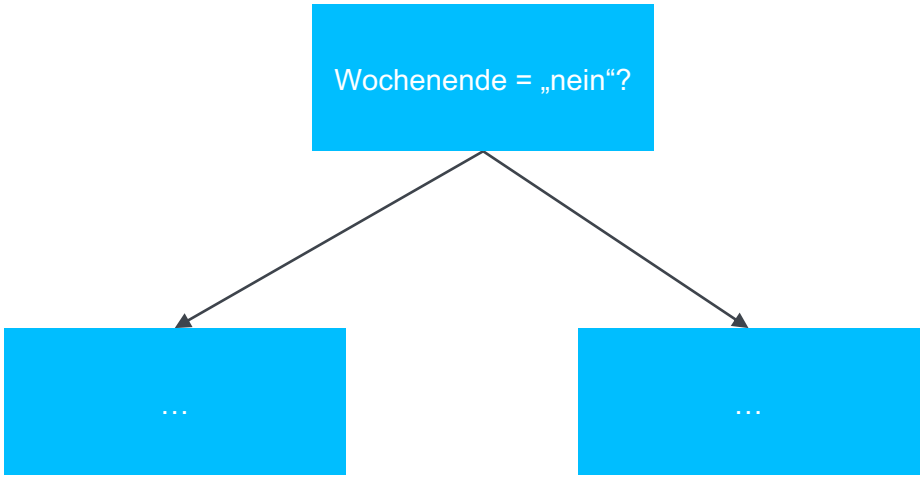
Beispiel aus dem Jupyter-Notebook



- Beste Aufteilung für Temperatur: 16.5 Grad
- Untermenge 1: 19 Fälle geschlossen, Entropie 0
- Untermenge 2: 1 Fall geschlossen, 14 Fälle offen
  - Insgesamt 15 Datenpunkte
  - $1/15 * - \log_2(1/15) + 14/15 * - \log_2(14/15) = 0.353$
- Gewichtete Entropie:
  $19/34 * 0 + 15/34 * 0.353 = 0.156$

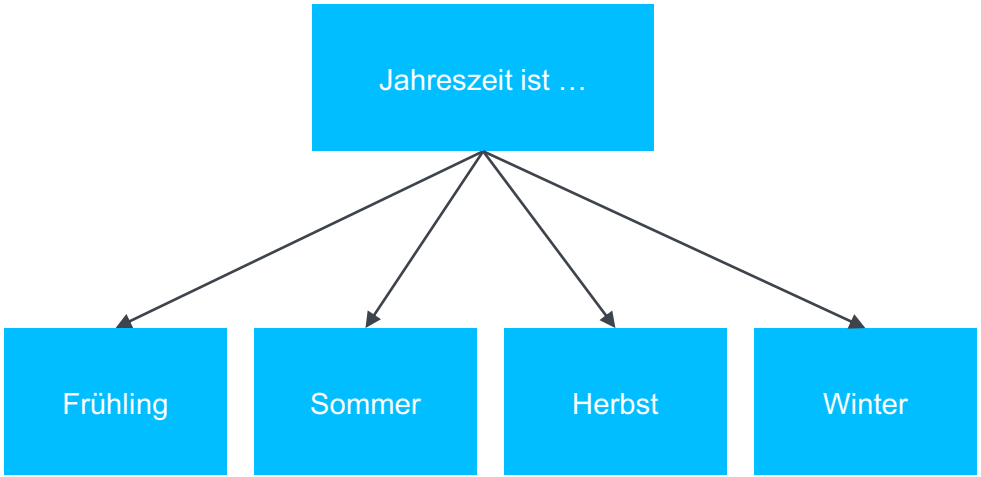
Kategoriale Features

- Was macht man, wenn die Features keine kontinuierlichen Werte aufweisen?
- Zum Beispiel: Wochenende nein/ja
  - Kodierung als 0 und 1
  - Oder Verzweigungen wie Wochenende = “nein“



Kategoriale Features mit mehreren Klassen

- Was macht man, wenn die Features mehrere Klassen haben?
- Zum Beispiel: Jahreszeit
  - Dummy-Kodierung als 0 und 1, alles wie bisher
  - Oder Mehrfachverzweigungen erlauben
- Achtung: sklearn erlaubt nur binäre Verzweigungen und nur kontinuierliche Werte



Wie baut man Zweige?

- Gegeben: Datenpunkte
- Gesucht: Kriterium für Aufteilung: ein Feature und ein Schwellwert
- Gehe alle Features und alle möglichen Schwellwerte durch
  - Wie würden die Daten in zwei Mengen aufgeteilt?
  - Welche Entropien hätten die Verteilungen der Klassen in beiden Mengen jeweils?
  - Welche Entropie ergibt das insgesamt für den aktuellen Knoten (gewichtet nach Häufigkeit)?
- Wähle das Feature und den Schwellwert mit der niedrigsten (also „besten“) Gesamt-Entropie
- Füge einen Knoten ein sowie einen linken Zweig und einen rechten Zweig
- Teile die Daten gemäß Feature und Schwellwert auf diese beiden Zweige auf

Wie baut man ein Blatt?

- Füge einen Knoten ein
- Weise ihm die Klasse zu, die in den Daten am häufigsten vorkommt

Wie baut man einen Baum?

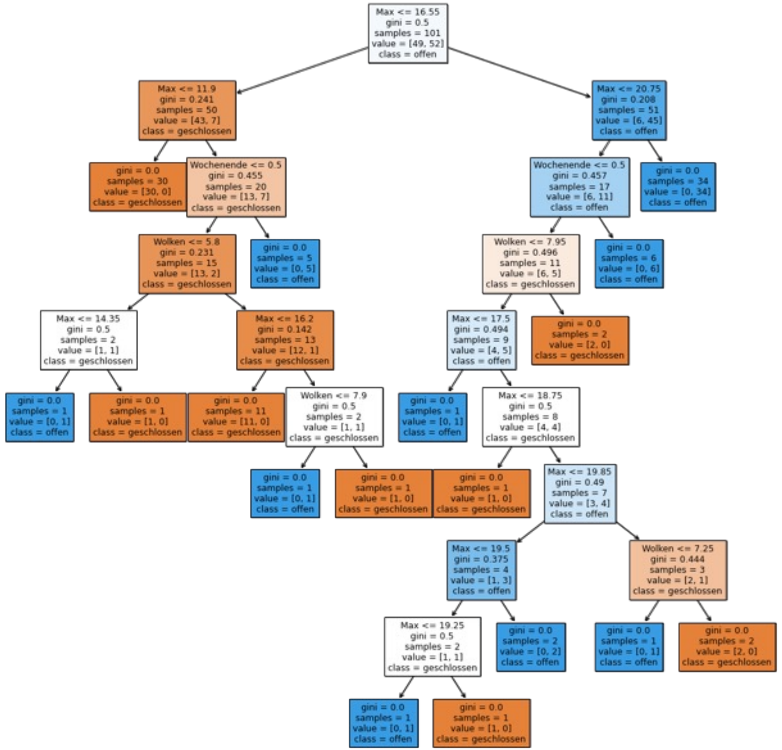
- Falls alle Datenpunkte dieselbe Klasse haben, erstelle ein Blatt
- Andernfalls:
  - Erstelle einen Zweig
  - Baue einen Baum für den linken Zweig
  - Baue einen Baum für den rechten Zweig

Eigentlich bedeutet Training: immer und immer wieder üben, um bessere Ergebnisse zu erzielen…

Das Erstellen eines Entscheidungsbaumes wird oft als „Training“ bezeichnet.

Drag the words: Wie baut man einen Entscheidungsbaum?

Overfitting



- 101 Datenpunkte für Eisverkäufe, 33 Knoten
- Verdächtig: nicht alle Knoten erscheinen sinnvoll
- Eisverkäufe unterliegen auch dem Zufall

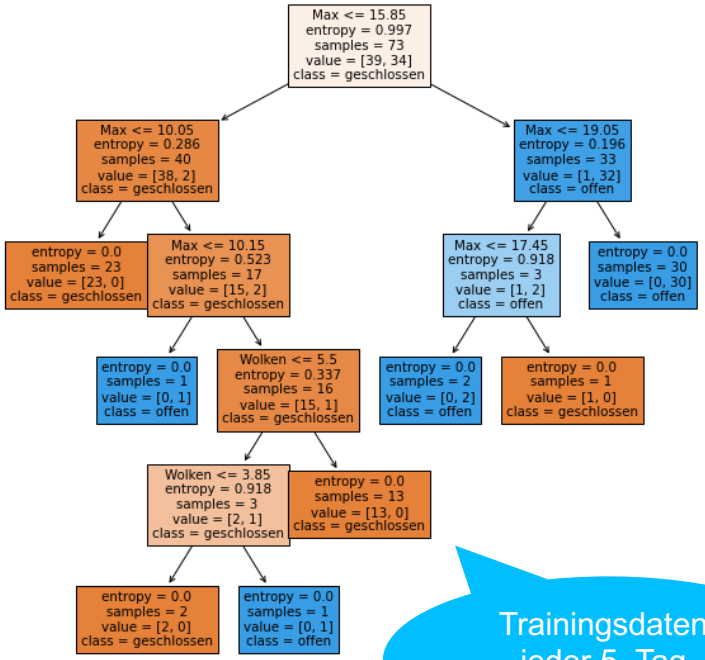
Overfitting

- Dieser Baum versucht trotzdem, die Daten (zu) perfekt zu modellieren

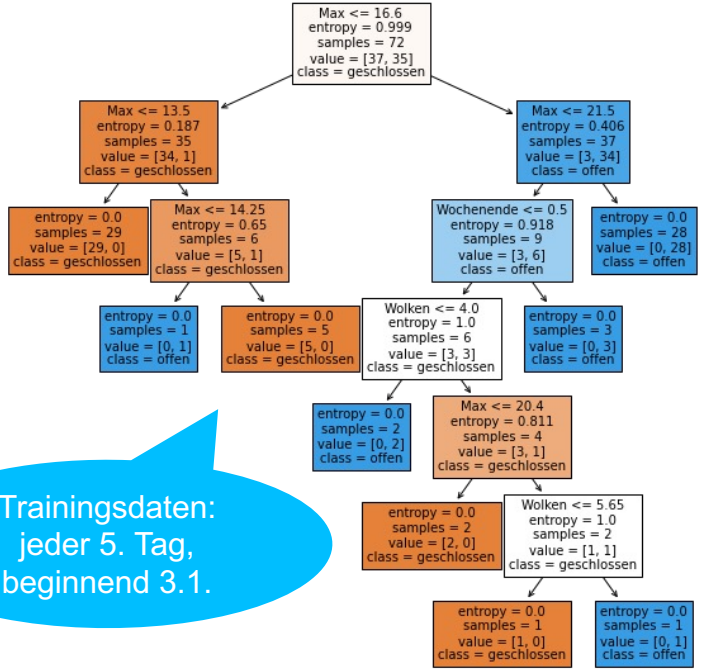
Schlechte Generalisierung

- Es ist unwahrscheinlich, dass dieser Baum auch für Daten von anderen Tagen perfekt ist

Beispiel für schlechte Generalisierung



Trainingsdaten:
jeder 5. Tag,
beginnend 1.1.



Trainingsdaten:
jeder 5. Tag,
beginnend 3.1.

Overfitting bezeichnet die zu starke Anpassung eines KI-Modells auf die Trainingsdaten.

Overfitting ist nicht wünschenswert, weil solche Modelle nicht gut generalisieren: sie erzielen in der Regel auf neuen Datenpunkten deutlich schlechtere Ergebnisse.

Overfitting verhindern

- Erkennen von Overfitting: siehe spätere Kursabschnitte
- Verhindern von Overfitting bei Entscheidungsbäumen
  - Blätter auch dann einfügen, wenn nicht ganz perfekt

⇒ Neue Kriterien zum Verzweigen bzw. zum Erstellen von Blättern!

Wie baut man einen Baum mit weniger Overfitting?

- Falls alle Datenpunkte dieselbe Klasse haben, erstelle ein Blatt
- Andernfalls:
  - Erstelle einen Zweig
  - Baue einen Baum für den linken Zweig
  - Baue einen Baum für den rechten Zweig

Falls der Baum schon sehr tief ist, erstelle ein Blatt

Falls nicht mehr viele Datenpunkte vorhanden sind, erstelle ein Blatt

Wenn es noch nicht zu viele Blätter gibt, erstelle ein Blatt

… aber nur, wenn für beide Zweige genügend Datenpunkte vorhanden sind

Beim Bau von Entscheidungsbäumen gibt es eine Reihe von Parametern, die Overfitting reduzieren können.

Maßnahmen gegen Overfitting, die Zweite

- Fertig erstellte Bäume können nachträglich vereinfacht werden
- Dies bezeichnet man als Pruning
- Engl. „Pruning“ = Zurechtstutzen/Beschneiden von Bäumen
- Dabei werden z.B. Unterbäume durch Blätter ersetzt

Das nachträgliche Vereinfachen von Entscheidungsbäumen bezeichnet man als Pruning.

Pruning ist eine weitere Möglichkeit, Overfitting zu reduzieren.

Die Verhinderung zu komplexer Bäume schon beim Bauen durch Parameter, die die Anzahl der Verzweigungen oder der Blätter begrenzen, bezeichnet man als Pre-Pruning.

Nochmal: wie baut man Zweige?

- Gehe alle Features und alle möglichen Schwellwerte durch
  - Wie würden die Daten in zwei Mengen aufgeteilt?
  - Welche Entropien hätten die Verteilungen der Klassen in beiden Mengen jeweils?
  - Welche Entropie ergibt das insgesamt für den aktuellen Knoten (gewichtet nach Häufigkeit)?

Wähle das Feature und den Schwellwert mit der niedrigsten (also „besten“) Gesamt-Entropie

- Füge einen Knoten ein sowie einen linken Zweig und einen rechten Zweig.
- Teile die Daten gemäß Feature und Schwellwert auf diese beiden Zweige auf

Greedy-Algorithmus:
das in diesem Moment
(„lokal“) beste

Problem mit der Greedy-Strategie

- Garantiert nicht den insgesamt (global) optimalen Baum
- Möglicherweise ergäben sich durch Wahl eines anderen Features/Schwellwerts später andere Verzweigungen
- Evtl. also ein besserer oder auch weniger komplexer Baum

Im Maschinellen Lernen werden häufig „Greedy“-Algorithmen angewandt.

Diese entscheiden sich für die lokal optimale Wahl, garantieren aber nicht, dass das Resultat auch insgesamt (global) optimal ist.

Drag the words: Pruning

INNOVET

KI B³

Künstliche Intelligenz in die berufliche Bildung bringen

Dr. Antje Schweitzer

Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung

Universität Stuttgart
Institut für Maschinelle Sprachverarbeitung
Institut für Software Engineering

IHK Industrie- und Handelskammer Reutlingen
Reutlingen | Tübingen | Zollernalb

IHK Region Stuttgart

IHK Industrie- und Handelskammer Karlsruhe

LMU Ludwig-Maximilians-Universität München

GEFÖRDERT VOM

INNOVET

Bundesministerium für Bildung und Forschung

bibb Bundesinstitut für Berufsbildung

Gefördert als InnoVET-Projekt aus Mitteln des Bundesministeriums für Bildung und Forschung.

Lizenzbestimmungen

Quellenverzeichnis

Titelfoto: Randy Fath (https://unsplash.com/@randyfath), „Amish barn-raising near my home“, auf Unsplash (https://unsplash.com/photos/ymf4_9Y9S_A), ist lizenziert unter Unsplash-Lizenz (https://unsplash.com/license). Bildausschnitt verändert.