

Probelehrveranstaltung "Training von Entscheidungsbäumen" - Lecture Notes

Contents

1	Entscheidungsbäume	1
1.1	Einleitung	1
1.2	Terminologie	2
1.3	Allgemeiner Algorithmus zum Training eines Entscheidungsbaums	3
1.4	Time Complexity	4
1.5	Information Gain	4
1.6	Information Theory and Entropy	5

1 Entscheidungsbäume

1.1 Einleitung

- Entscheidungsbaum-Algorithmen können als iterative, von oben nach unten verlaufende Konstruktionsmethode für das Model (den Klassifikator) betrachtet werden. Man kann sich einen Entscheidungsbaum als eine Hierarchie von Entscheidungen vorstellen, die einen Datenraum in Teilräume unterteilen.
- Entscheidungsbäume können jede boolesche (binäre) Funktion darstellen, und der Hypothesenraum, der durchsucht wird, ist der gesamte Raum der booleschen Funktionen. Wir müssen jedoch bedenken, dass eine wesentliche Herausforderung im maschinellen Lernen darin besteht, ob ein Algorithmus in der Lage ist, die "richtige" Funktion oder eine gute Annäherung innerhalb des zu durchsuchenden Teilraums zu lernen oder zu finden.
- Wenn nur binäre (oder boolesche) Merkmale betrachtet werden, gibt es an jedem Knoten 2^m mögliche Aufteilungen, die bewertet werden müssen, vorausgesetzt, dass der Datenraum m Merkmale umfasst.
- Entscheidungsbaum-Algorithmen durchsuchen den Hypothesenraum mit einem greedy-Ansatz über alle möglichen Bäume. Eine vollständige Suche ist aufgrund der exponentiellen Natur des Problems nicht praktikabel. Angenommen, wir haben m boolesche Merkmale, dann gibt es 2^m mögliche

Merkmalskombinationen. Wenn wir davon ausgehen, dass es sich um ein binäres Klassifikationsproblem handelt, gibt es 2^{2^m} mögliche Möglichkeiten, die Daten zu kennzeichnen. Wenn jeder Baum einer eindeutigen Zuordnungsfunktion entspricht, wird deutlich, wie unmöglich es wird, eine vollständige Suche durchzuführen.

1.2 Terminologie

- Wurzel: keine eingehende Kante, repräsentiert den gesamten Datenraum.
- Entscheidungsknoten: eine eingehende Kante, zwei oder mehr ausgehende Kanten.
- Blatt: jedem Blatt ist ein Label zugeordnet, keine ausgehenden Kanten.

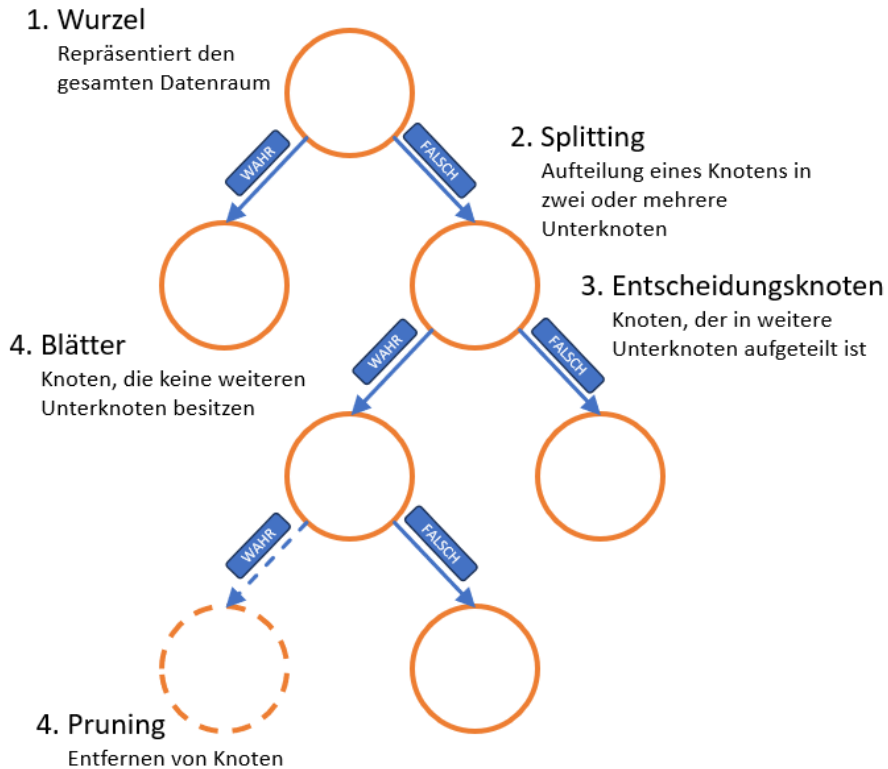


Figure 1: Beispiel eines binären Entscheidungsbaums.

1.3 Allgemeiner Algorithmus zum Training eines Entscheidungsbaums

1.4 Laufzeit des Algorithmus

$\mathcal{O}(m \cdot n^2 \log n)$.

1.5 Informationsgewinn

$$\text{GAIN}(\mathcal{D}, x_j) = H(\mathcal{D}) - \sum_{v \in \text{Values}(x_j)} \frac{|\mathcal{D}_v|}{|\mathcal{D}|} H(\mathcal{D}_v) \quad (5)$$

1.6 Informationstheorie und Entropie

Referenzen

- Sebastian Raschka, Decision Trees - Lecture Notes, 2019