

Entscheidungsbaumlernen

Aufgabe

Eine KI soll in einem Online-Shop Jugendlichen Smartphone-Modelle empfehlen, die zu ihren Auswahlkriterien passen. Dafür soll aus den folgenden Trainingsdaten ein Entscheidungsbaum gelernt werden.

Jeder Datensatz beschreibt, welche **Empfehlung** in der Vergangenheit von menschlichen Verkäufern bei einem bestimmten **Budget** des Jugendlichen und bestimmten Vorstellungen über **Akkulaufzeit**, **Kamera** und **Speichergröße** des Handys gegeben wurde.

Budget	Akkulaufzeit	Kamera	Speicher	Empfehlung
niedrig	lang	mittel	groß	Modell A
hoch	mittel	wichtig	groß	Modell B
mittel	lang	mittel	groß	Modell C
niedrig	kurz	egal	klein	Modell A
hoch	lang	wichtig	groß	Modell B
mittel	mittel	mittel	klein	Modell C
niedrig	lang	egal	klein	Modell A
hoch	mittel	mittel	groß	Modell B
mittel	lang	wichtig	groß	Modell C
niedrig	kurz	egal	klein	Modell A
mittel	mittel	wichtig	groß	Modell C
niedrig	lang	mittel	groß	Modell C
hoch	mittel	wichtig	groß	Modell B
niedrig	lang	wichtig	klein	Modell B
hoch	lang	mittel	klein	Modell B
mittel	lang	wichtig	klein	Modell B
niedrig	lang	wichtig	groß	Modell C
hoch	kurz	mittel	groß	Modell C
mittel	kurz	egal	groß	Modell A

a) Berechne die **Gini-Unreinheit** der gesamten Tabelle bezüglich des **Zielfeatures** „Empfehlung“ (Du kannst dazu die Formeln aus Kapitel 6 (KI) der Formelsammlung nutzen).

b) Berechne die *gewichtete* Gini-Unreinheit für das Merkmal „Budget“ (wieder bezüglich des Zielfeatures „Empfehlung“).

c) Angenommen, es ergäben sich die folgenden gewichteten Gini-Unreinheiten für die verschiedenen Features:

- Budget: 0,456
- Akkulaufzeit: 0,542
- Kamera: 0,408
- Speicher: 0,612

Nenne das Merkmal, das als Wurzelknoten des Entscheidungsbaums ausgewählt werden sollte.

d) Die Gini-Werte aus der vorigen Teilaufgabe sind korrekt. Zeichne den daraus resultieren Wurzelknoten des Entscheidungsbaums und seine ausgehenden Äste.

e) Der Algorithmus wird als nächstes für jeden der drei Äste der Wurzelknotens *rekursiv* aufgerufen.

Entscheidungsbaumlernen

- Mit welchen Datensätze arbeitet der Algorithmus weiter, z.B. für den Fall *Kamera=wichtig*? Stelle als Tabelle dar.
 - Welche Features stehen an dieser Stelle des Baums (Ast *Kamera=wichtig*) noch zur Auswahl und welches davon wird wohl als nächster Entscheidungsknoten eingefügt? Argumentiere mithilfe deiner Tabelle aus der vorigen Teilaufgabe.
- f) Welche *Tiefe* kann der Baum maximal haben?
Warum ist es normalerweise nicht sinnvoll, einen Entscheidungsbaum bis zur maximal möglichen Tiefe aufzubauen?
- g) Zeichne den Baum bis zur Tiefe 2 (Wurzel, weitere Ebene mit Entscheidungsknoten, Endknoten).
Du kannst dich entscheiden:
- entweder du berechnest *korrekt*, d.h. mit Hilfe der Gini-Unreinheit, welche Knoten auf Ebene 2 an die Äste der Wurzel angehängt werden
 - oder du stellst (ähnlich wie in Teilaufgabe e) Vermutungen über die weiteren Knoten an.
- Vergleiche deinen Baum mit der Musterlösung.