

---

**Mustererkennung WiSe 12/13**  
**Übung 11**

Lutz Freitag, Sebastian Kürten

---

## 2 Aufgabe 2: ID3-Trees

Wir haben insgesamt 14 Datensätze. Der hervorzusagende Wert ist  $\{yes, no\}$  für das Feld Play. Yes tritt 9 mal auf, No 5 Mal. Die Wahrscheinlichkeiten für Yes sei mit  $p_1$  bezeichnet, die für No mit  $p_2$ . Die Entropie berechnet sie dann wie folgt:

$$p_1 = \frac{9}{14}, p_2 = \frac{5}{14}$$

$$E = -p_1 * \ln(p_1) - p_2 * \ln(p_2) = 0,652$$

Die Frage ist nun, nach welchem der beiden Kriterien Outlook oder Temperature auf der ersten Stufe des Baumes getrennt wird. Dazu betrachten wir für beide Möglichkeiten die durchschnittliche Entropie auf der jeweils nächsten Stufe.

### 2.1 Fall 1: Outlook

Outlook	Yes	No	Summe
Sunny	2	3	5
Overcast	4	0	4
Rain	3	2	5

$$E_{Sunny} = -\frac{2}{5} * \ln(\frac{2}{5}) - \frac{3}{5} * \ln(\frac{3}{5}) = 0,673$$

$$E_{Overcast} = -1 * \ln(1) - 0 * \ln(0) = 0$$

$$E_{Rain} = -\frac{3}{5} * \ln(\frac{3}{5}) - \frac{2}{5} * \ln(\frac{2}{5}) = 0,673$$

Damit ergibt sich im Mittel:

$$E_{Outlook}^1 = t_{Sunny} * E_{Sunny} + t_{Overcast} * E_{Overcast} + t_{Rain} * E_{Rain}$$

wobei

$$t_{Sunny} = \frac{5}{14}, t_{Overcast} = \frac{4}{14}, t_{Rain} = \frac{5}{14}$$

also:

$$E_{Outlook}^1 = 0,481$$

## 2.2 Fall 2: Temperature

Temperature	Yes	No	Summe
Hot	2	2	4
Mild	4	2	6
Cool	3	1	4

$$E_{Hot} = -\frac{1}{2} * \ln(\frac{1}{2}) - \frac{1}{2} * \ln(\frac{1}{2}) = 0,693$$

$$E_{Mild} = -\frac{4}{6} * \ln(\frac{4}{6}) - \frac{2}{6} * \ln(\frac{2}{6}) = 0,637$$

$$E_{Cool} = -\frac{3}{4} * \ln(\frac{3}{4}) - \frac{1}{4} * \ln(\frac{1}{4}) = 0,562$$

Damit ergibt sich im Mittel:

$$E_{Temperature}^1 = t_{Hot} * E_{Hot} + t_{Mild} * E_{Mild} + t_{Cool} * E_{Cool}$$

wobei

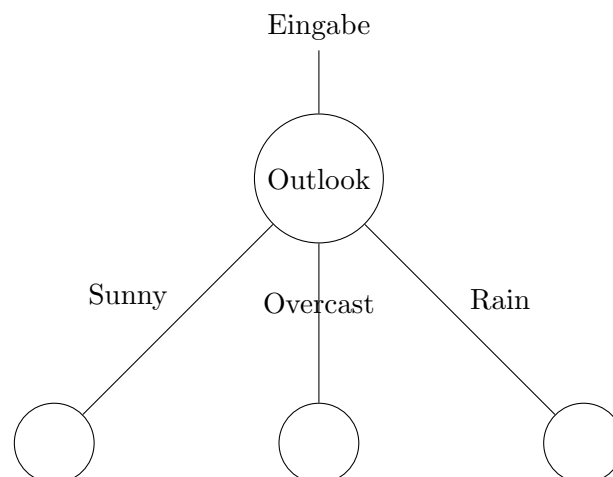
$$t_{Hot} = \frac{4}{14}, t_{Mild} = \frac{6}{14}, t_{Cool} = \frac{4}{14}$$

also:

$$E_{Temperature}^1 = 0,632$$

## 2.3 Auswahl der Trennkriteriums

Die mittlere Entropie für das Trennkriterium Outlook ist geringer als die des Kriteriums Temperature. Daher wird auf der ersten Stufe des Baums nach Outlook getrennt.



## 2.4 Zweite Ebene

Für die zweite Ebene ergeben sich folgende Tabellen: siehe Abbildung 4.

Damit können wir dann den Entscheidungsbaum aufstellen: siehe Abbildung 5

Temperature	Play
Hot	No
Hot	No
Mild	No
Cool	Yes
Mild	Yes

Abbildung 1: Sunny

Temperature	Play
Hot	Yes
Cool	Yes
Mild	Yes
Hot	Yes

Abbildung 2: Overcast

Temperature	Play
Mild	Yes
Cool	Yes
Cool	No
Mild	Yes
Mild	No

Abbildung 3: Rain

Abbildung 4: Daten auf der zweite Ebene

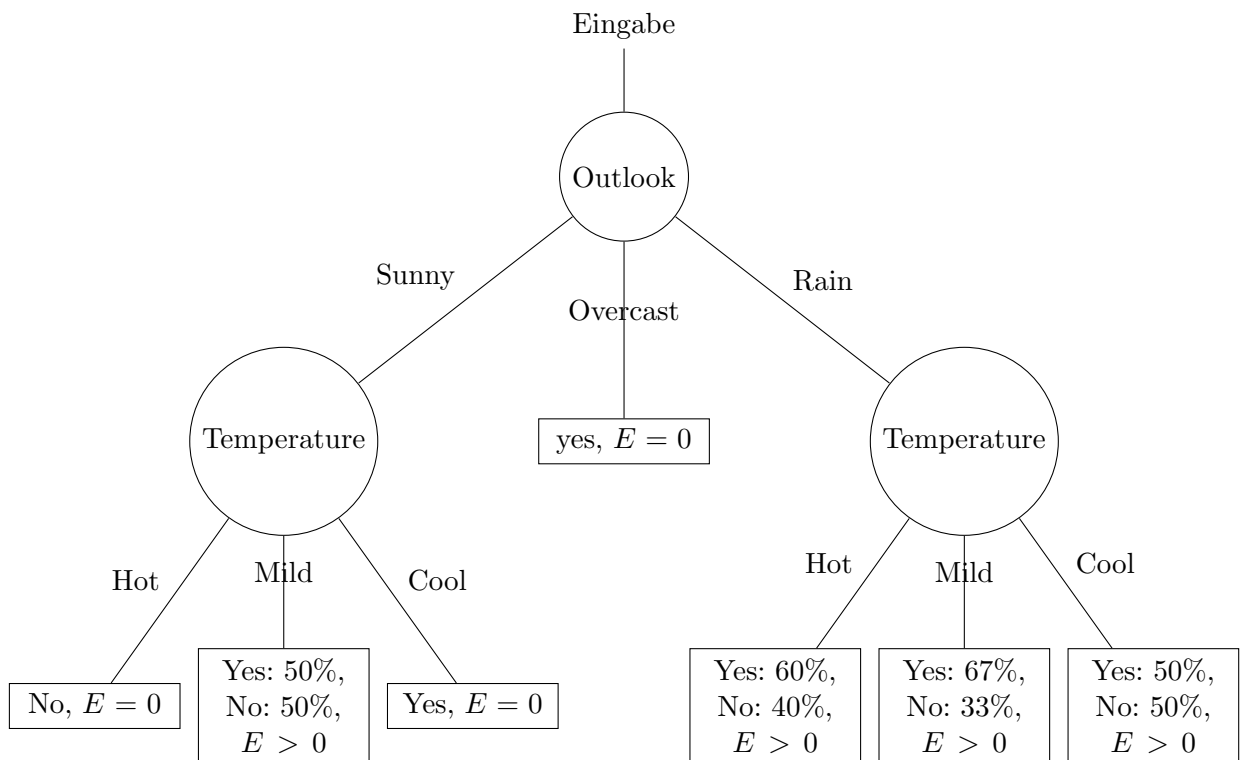


Abbildung 5: fertiger Entscheidungsbaum