Entropie

Carsten Gips (FH Bielefeld)

Wie Attribute wählen?

Relevanz => Informationsgehalt

- Shannon/Weaver (1949): **Entropie**
  - Maß für die Unsicherheit einer Zufallsvariablen
  - Anzahl der Bits zur Darstellung der Ergebnisse eines Zufallsexperiments

Beispiele

- Münze, die immer auf dem Rand landet: keine Unsicherheit, 0 Bit
- Faire Münze: Kopf oder Zahl: Entropie 1 Bit
- Fairer 4-seitiger Würfel: 4 mögliche Ausgänge: Entropie 2 Bit
- Münze, die zu 99% auf einer Seite landet: Entropie nahe Null

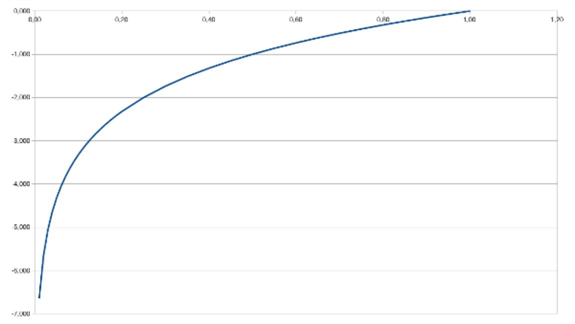
=> Anzahl der Ja/Nein-Fragen, um zur gleichen Information zu kommen

Definition der Entropie $H(V)$ für Zufallsvariable V

- Zufallsvariable V => mögliche Werte v_k
- Wahrscheinlichkeit für v_k sei $p_k = P(v_k)$

$$\boxed{H(V) \quad = \quad -\sum_k p_k \log_2 p_k}$$

Hinweis: $\log_2 x = \frac{\log_{10} x}{\log_{10} 2} = \frac{\log x}{\log 2}$



Beispiele Entropie: faire Münze

$$\text{Entropie: } H(V) \;=\; -\sum_k p_k \log_2 p_k$$

- $v_1 = \text{Kopf}, v_2 = \text{Zahl}$
- $p_1 = 0.5, p_2 = 0.5$
- $H(\text{Fair}) = -(0.5 \log_2 0.5 + 0.5 \log_2 0.5) = 1 \text{ Bit}$

$\log_2 0.5 = -1$

Beispiele Entropie: unfaire Münze

Entropie: $H(V) \;=\; -\sum_k p_k \log_2 p_k$

Beispiele Entropie: unfaire Münze

$$\boxed{\text{Entropie: } H(V) \;=\; -\textstyle\sum_k p_k \log_2 p_k}$$

- $v_1 = \text{Kopf}, v_2 = \text{Zahl}$
- $p_1 = 0.99, p_2 = 0.01$
- $H(\text{UnFair}) = -(0.99 \log_2 0.99 + 0.01 \log_2 0.01)$
  $H(\text{UnFair}) \approx 0.08$ Bit

$\log_2 0.01 \approx -6.64$

$\log_2 0.99 \approx -0,014$

Beispiele Entropie: 4-seitiger Würfel

Entropie: $H(V) \;=\; -\sum_k p_k \log_2 p_k$

Beispiele Entropie: 4-seitiger Würfel

$$\boxed{\text{Entropie: } H(V) \;=\; -\sum_k p_k \log_2 p_k}$$

- $v_1 = 1, v_2 = 2, v_3 = 3, v_4 = 4$
- $p_1 = p_2 = p_3 = p_4 = 0.25$
- $H(\text{Wuerfel}) = -4 \cdot (0.25 \log_2 0.25) = 2$ Bit

$\log_2 0.25 = -2$

Entropie der Trainingsmenge: Häufigkeit der Klassen zählen

Nr.	x_1	x_2	x_3	k
1	0	0	0	A
2	1	0	2	A
3	0	1	1	A
4	1	1	0	B
5	0	1	1	B
6	0	1	0	A

- Anzahl Klasse A: 4
- Anzahl Klasse B: 2
- Gesamtzahl Beispiele: 6

Wahrscheinlichkeit für A: $p_A = 4/6 = 0.667$

Wahrscheinlichkeit für B: $p_B = 2/6 = 0.333$

Entropie der Trainingsmenge: Häufigkeit der Klassen zählen

Nr.	x_1	x_2	x_3	k
1	0	0	0	A
2	1	0	2	A
3	0	1	1	A
4	1	1	0	B
5	0	1	1	B
6	0	1	0	A

- Anzahl Klasse A: 4
- Anzahl Klasse B: 2
- Gesamtzahl Beispiele: 6

Wahrscheinlichkeit für A: $p_A = 4/6 = 0.667$

Wahrscheinlichkeit für B: $p_B = 2/6 = 0.333$

$$\begin{aligned} H(S) &= -\sum_k p_k \log_2 p_k \\ &= -(4/6 \cdot \log_2 4/6 + 2/6 \cdot \log_2 2/6) \\ &= -(-0.39 - 0.53) = 0.92\,\mathrm{Bit} \end{aligned}$$

Mittlere Entropie nach Betrachtung von Attribut A

$$R(S, A) = \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} H(S_v)$$

Entropie der Trainingsmenge nach Attributwahl

Nr.	x_1	x_2	x_3	k
1	0	0	0	A
2	1	0	2	A
3	0	1	1	A
4	1	1	0	B
5	0	1	1	B
6	0	1	0	A

- Sei Attribut x_1 ausgewählt
- x_1 partitioniert die Trainingsmenge
  - $x_1 = 0$ liefert $S_0 = \{1, 3, 5, 6\}$
  - $x_1 = 1$ liefert $S_1 = \{2, 4\}$
  - Häufigkeit für $x_1 = 0$: $4/6$
  - Häufigkeit für $x_1 = 1$: $2/6$
  - Gesamtzahl Beispiele: 6

Entropie der Trainingsmenge nach Attributwahl

Nr.	x_1	x_2	x_3	k
1	0	0	0	A
2	1	0	2	A
3	0	1	1	A
4	1	1	0	B
5	0	1	1	B
6	0	1	0	A

- Sei Attribut x_1 ausgewählt
- x_1 partitioniert die Trainingsmenge
  - $x_1 = 0$ liefert $S_0 = \{1, 3, 5, 6\}$
  - $x_1 = 1$ liefert $S_1 = \{2, 4\}$
  - Häufigkeit für $x_1 = 0$: $4/6$
  - Häufigkeit für $x_1 = 1$: $2/6$
  - Gesamtzahl Beispiele: 6

$$
\begin{aligned}
R(S, A) &= \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} H(S_v) \\
&= 4/6 \cdot H(\{1, 3, 5, 6\}) + 2/6 \cdot H(\{2, 4\}) \\
&= 4/6 \cdot (-3/4 \cdot \log_2 3/4 - 1/4 \cdot \log_2 1/4) + \\
&\quad\; 2/6 \cdot (-1/2 \cdot \log_2 1/2 - 1/2 \cdot \log_2 1/2) \\
&= 0.54 + 0.33 = 0.87\,\text{Bit}
\end{aligned}
$$

Wrap-Up

- Begriff und Berechnung der Entropie: Maß für die Unsicherheit
- Begriff und Berechnung des Informationsgewinns
  - Entropie für eine Trainingsmenge
  - Mittlere Entropie nach Wahl eines Attributs

LICENSE