Entropie, Informationsgehalt und Informationsgewinn

Die Informationstheorie wurde von **Claude Shannon** entwickelt, um die Übertragung und Verarbeitung von Informationen zu analysieren und zu optimieren. Die Begriffe **Entropie**, **Informationsgehalt** und **Informationsgewinn** sind zentrale Konzepte in diesem Bereich. Sie helfen, Unsicherheit zu quantifizieren und Muster in Daten zu verstehen.

1. Informationsgehalt $I(x)$

Der **Informationsgehalt** beschreibt, wie überraschend ein einzelnes Ereignis ist. Je unwahrscheinlicher ein Ereignis, desto mehr Information liefert es, wenn es eintritt.

Formel:

$$I(x) = -\log_2 P(x)$$

- $P(x)$: Wahrscheinlichkeit des Ereignisses x.
- $I(x)$: Informationsgehalt des Ereignisses.

Beispiel:

- Ereignis A: "Es regnet morgen." ($P(A) = 0.9$)
  $I(A) = -\log_2(0.9) = 0.15$ → Niedriger Informationsgehalt (es ist fast sicher).
- Ereignis B: "Ein Meteorit schlägt ein." ($P(B) = 0.01$)
  $I(B) = -\log_2(0.01) = 6.64$ → Hoher Informationsgehalt (sehr unwahrscheinlich und überraschend).

Der Informationsgehalt misst die Überraschung eines einzelnen Ereignisses.

---

2. Entropie $H(X)$

Die **Entropie** misst die **durchschnittliche Unsicherheit** oder den **durchschnittlichen Informationsgehalt** eines Systems. Sie betrachtet alle möglichen Ereignisse und deren Wahrscheinlichkeiten.

Formel:

$$H(X) = -\sum_i P(x_i) \cdot \log_2 P(x_i)$$

- $P(x_i)$: Wahrscheinlichkeit des i-ten möglichen Ereignisses.
- $H(X)$: Entropie des Systems X.

Eigenschaften:

- **Hohe Entropie**: Alle Ereignisse sind etwa gleich wahrscheinlich, die Unsicherheit ist groß.
  Beispiel: Ein fairer Münzwurf ($P(Kopf) = 0.5$, $P(Zahl) = 0.5$) hat maximale Entropie.
- **Niedrige Entropie**: Ein Ereignis dominiert, die Unsicherheit ist gering.
  Beispiel: Eine fast sichere Münze ($P(Kopf) = 0.99$, $P(Zahl) = 0.01$).

Beispiel:

1. Fairer Würfel ($P(1) = P(2) = ... = P(6) = \frac{1}{6}$):
   $H(X) = -\sum_{i=1}^{6} \frac{1}{6} \log_2 \frac{1}{6} = 2.58$.
2. Würfel, bei dem 1 sehr wahrscheinlich ist ($P(1) = 0.9$, $P(2) = ... = P(6) = 0.02$):
   $H(X) \approx 0.5$.

Entropie beschreibt die durchschnittliche Unsicherheit in einem System.

---

3. Informationsgewinn

Der **Informationsgewinn** misst die Reduktion der Entropie durch zusätzliche Informationen oder durch Aufteilung von Daten. Er wird häufig im maschinellen Lernen, insbesondere bei Entscheidungsbäumen, genutzt.

Formel:

$$\text{Informationsgewinn} = H(S) - \sum_i \frac{|S_i|}{|S|} H(S_i)$$

- $H(S)$: Entropie des gesamten Datensatzes vor der Aufteilung.
- $H(S_i)$: Entropie der Teilmengen nach der Aufteilung.
- $\frac{|S_i|}{|S|}$: Gewichtung der Teilmengen relativ zur Gesamtmenge.

Beispiel in Entscheidungsbäumen:

Ein Datensatz hat 100 Instanzen, davon 50 in Klasse A und 50 in Klasse B.

1. Entropie vor der Aufteilung:
   $H(S) = -\left(\frac{50}{100} \log_2 \frac{50}{100} + \frac{50}{100} \log_2 \frac{50}{100}\right) = 1.$
2. Nach einer Aufteilung, bei der die Klassen getrennt werden, wird die Entropie reduziert. Der Informationsgewinn ist die Differenz.

Der Informationsgewinn zeigt, wie stark die Unsicherheit reduziert wird.