Stochastik für die Informatik, Vorlesung 27

Inhalt

- Entropie, relative Entropie
- Bedeutung der Entropie für die Informationstheorie, Entropieschranken

Lernziele

- Den Begriff der Entropie kennen, Entropien berechnen können
- Die Bedeutung der Entropie in der Informationstheorie kennen
- Entropieschranken mittels der relativen Entropie berechnen können

Vorkenntnisse: Zufallsvariablen, Verteilungen, Erwartungswerte; Logarithmen

Kapitel X: Informationstheorie

Literatur: G. Kersting, A. Wakolbinger: Elementare Stochastik; Kapitel VI: Ideen aus der Informationstheorie.

Verfügbar in der Bibliothek, auch als E-Book.

Inhalt: Stochastische Aspekte der Informationsübermittlung: Codierung, Redundanz, Entropie...

Entropie

(Def.) Sei X eine diskrete Zufallsvariable mit Werten in S und Verteilung $\rho(a) = \mathbb{P}(X = a)$. dann ist die Entropie von X definiert als

$$H[X] := -\sum_{a \in S} \rho(a) \cdot \log \rho(a).$$

- $0 \cdot \log 0 = 0$.
- Die Basis des Logarithmus ist hier nicht spezifiziert. Binäre (Shannon-)Codes: Basis 2.
- $H[X] \geq 0$

Interpretation: Entropie als Maß für den Informationsgehalt bzw. der Ungewissheit von X. Sie entspricht (ungefähr) der mittleren Zahl von Ja-Nein Fragen, welche benötigt werden, um den unbekannten Wert von X zu erfragen.

- (Bernoulli-Verteilung)
- (Gleichverteilung)

Relative Entropie

(Def.) Seien ρ und π zwei diskrete Wahrscheinlichkeitsverteilungen. Dann ist die relative Entropie von ρ bezüglich π definiert als

$$D(\rho \| \pi) := \sum_{a \in S} \rho(a) \log \frac{\rho(a)}{\pi(a)}.$$

- Summanden mit $\rho(a) = 0$ werden 0 gesetzt
- Die relative Entropie heißt auch Kullback-Leibler-Information
- Interpretation: Unterschied der erwarteten Codewortlänge, wenn die Verteilung ρ statt π ist (Shannon-Code, Quellencodierungssatz), da

$$D(\rho \| \pi) = -\sum_{a \in S} \rho(a) \log \pi(a) - (-\sum_{a \in S} \rho(a) \log \rho(a))$$

ist.

Relative Entropie und Entropieschranken

(Satz) Für die relative Entropie gilt $D(\rho \| \pi) \geq 0$, und es gilt

$$D(\rho \| \pi) = 0 \quad \Leftrightarrow \rho = \pi.$$

- (Beweis)
- Folgerung:

$$H[X] = -\sum_{a \in S} \rho(a) \log \rho(a) \leq -\sum_{a \in S} \rho(a) \log \pi(a)$$

- (Beispiele: Gleichverteilung, geometrische Verteilung)

Gemeinsame Entropie

Die gemeinsame Entropie von n Zufallsvariablen $X_1, ..., X_n$ ist definiert als die Entropie der Zufallsvariablen $X = (X_1, ..., X_n)$, d.h.

$$H[X_1, ..., X_n] = -\sum_{a_1,...,a_n} \mathbb{P}(X_1 = a_1, ...X_n = a_n) \cdot \log \mathbb{P}(X_1 = a_1, ...X_n = a_n)$$

(Satz) Es gilt: $H[X, Y] \leq H[X] + H[Y]$. Gleichheit gilt genau dann, wenn X und Y unabhängig sind.

- (Beweis)
- (Beispiel: Codieren von Wörtern)

Bedingte Entropie

Die bedingte Entropie von Y gegeben $X = a$ ist definiert als

$$H[Y|X = a] = -\sum_b \mathbb{P}(Y = b|X = a) \cdot \log \mathbb{P}(Y = b|X = a),$$

und die bedingte Entropie von Y gegeben X also

$$H[Y|X] = \sum_a H[Y|X = a] \cdot \mathbb{P}(X = a).$$

- Interpretation: Mittlere Ungewissheit über den Wert von Y, die besteht, wenn man den Wert von X schon kennt.
- Es gilt $H[X, Y] = H[X] + H[Y|X]$ (Interpretation)
- Weiter gilt $H[Y|X] \leq H[Y]$

Wechselseitige Information

Die wechselseitige Information von X und Y ist gegeben durch

$$I(X\|Y) := H[Y] - H[Y|X].$$

- Informationsgewinn über Y durch Beobachtung von X
- $I(X\|Y) \geq 0$
- $I(X\|Y) = H[X] + H[Y] - H[X, Y]$, also insbesondere symmetrisch in X und Y.

Beispiel: Stationäre Quellen

Eine stationäre Quelle ist eine unendliche Folge $X_1, X_2, \ldots$ von zufälligen Buchstaben, für die die Verteillungen von $X_1, \ldots, X_n$ und $X_{m+1}, \ldots, X_{m+n}$ für alle $m, n \in \mathbb{N}$ übereinstimmen.

Die Entropierate der stationären Quelle ist definiert als

$$h_Q := \lim_{n \to \infty} \frac{H_2[X_1, \ldots, X_n]}{n}.$$

- Dieser Grenzwert existiert.
- Es gilt $0 \leq h_Q \leq H_2[X_1]$.
- Interpretation mittels Quellencodierungssatz, Redundanz