Datenkompression

Ausgegeben am 15.05.2014 — Besprechung in der Übung am 22.05.2014

Aufgabe 2.1

Auf einem Alphabet Σ sei die Wahrscheinlichkeitsfunktion $P\colon \Sigma \to [0,1]$ definiert. Für Zeichenketten aus Σ^m, deren Zeichen unabhängig voneinander gezogen werden, erhalten wir die Wahrscheinlichkeitsfunktion $P_m\colon \Sigma^m \to [0,1]$, $(c_1,\ldots,c_m) \mapsto P(c_1)\cdots P(c_m)$.

Beweisen Sie die Additivität der Entropie: $H(\Sigma^m, P_m) = mH(\Sigma, P)$.

Aufgabe 2.2

Berechnen Sie die Entropien der folgenden Wahrscheinlichkeitsfunktionen P und P'. Vergleichen Sie Ihre Ergebnisse mit den mittleren Codewortlängen aus Aufgabe 1.3.

x	a	e	i	o	u
$P(x)$	1/8	1/4	1/8	7/16	1/16
$P'(x)$	5/32	3/16	5/32	3/8	1/8

Aufgabe 2.3

Die Zeichen des Alphabetes $\{\mathtt{a},\mathtt{b},\mathtt{c}\}$ mit der Anordnung $\mathtt{a} < \mathtt{b} < \mathtt{c}$ treten mit den Wahrscheinlichkeiten $P(\mathtt{a}) = 0{,}3$, $P(\mathtt{b}) = 0{,}5$ und $P(\mathtt{c}) = 0{,}2$ auf.

a) Berechnen Sie die arithmetische Kodierung des Wortes `cabb`.

b) Dekodieren Sie ein Wort der Länge 6 aus dem Codewort `0100000000`.

Aufgabe 2.4

Schreiben Sie ein Programm, das für einen Eingabestring die mittlere Codewortlänge bei Shannon-Kodierung, Shannon-Fano-Kodierung bzw. Huffman-Kodierung ausgibt. Ihr Programm sollte die relativen Buchstabenhäufigkeiten in der Eingabe als Wahrscheinlichkeitsverteilung verwenden.

Verwenden Sie hierbei diejenige Variante der Shannon-Fano-Kodierung, bei der Listen von absteigend sortierten Wahrscheinlichkeiten in zwei Teile gespalten werden.

Aufgabe 2.5

Es soll die Konkavität der Entropie gezeigt werden.

a) Die binäre Entropie h ist die Entropie eines Zufallsprozesses mit zwei möglichen Ausgängen:

$$h\colon [0,1] \to \mathbb{R}, \quad p \mapsto -p\log_2(p) - (1-p)\log_2(1-p)\,.$$

Beweisen Sie, dass für $\lambda, p, q \in [0,1]$ gilt:

$$h\big(\lambda p + (1-\lambda)q\big) \geq \lambda h(p) + (1-\lambda)h(q)\,.$$

b) Es seien P, Q zwei Wahrscheinlichkeitsverteilungen auf dem Alphabet Σ und $\lambda \in [0,1]$. Die Wahrscheinlichkeitsverteilung R auf Σ sei definiert durch $R(c) \coloneqq \lambda P(c) + (1-\lambda)Q(c)$ für alle $c \in \Sigma$. Beweisen Sie: $H(\Sigma, R) \geq \lambda H(\Sigma, P) + (1-\lambda)H(\Sigma, Q)$.