Stochastik für die Informatik, Vorlesung 26

Inhalt

- Einführung in die Informationstheorie, Präfixcodes
- Shannon-Codes, Quellencodierungssatz
- Huffman-Codes

Lernziele

- Mit Präfix-Codes umgehen können
- Shannon- und Huffman-Codes und ihre Bedeutung kennen
- Huffman-Codes konstruieren 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...

Grundbegriffe

- Ein Alphabet ist eine höchstens abzählbare Menge S.
- Ein Buchstabe ist ein Element eines Alphabets.
- Ein binärer Code ist eine *injektive* Abbildung

$$k : S \to \bigcup_{l \geq 1} \{0,1\}^l,$$

also eine Abbildung, die jedem Buchstaben $a \in S$ eine Folge $k(a) = k_1(a), \ldots, k_l(a)$ der Länge $l = \ell(a)$ mit Folgengliedern $k_i(a) \in \{0,1\}$ als Codewort zuordnet.

Präfixcodes

Länge als Maß für die Güte des Codes: Möglichst kurz, aber auch leicht zu entschlüsseln

- Ein Präfixcode ist ein binärer Code, bei dem kein Codewort Anfangsstück eines anderen Codeworts ist. Das bedeutet: Für $a, b \in S$ mit $a \neq b$ gibt es *keine* Folge $f = (f_1, ..., f_m) \in \{0,1\}^m$ mit $k(a)f = k(b)$.

- (Code als Fragestrategie)
- (Baumdarstellung)

Binäre Codebäume

- Binärer Baum mit Wurzel (ganz oben)
- Von jedem Knoten führen höchstens zwei Kanten nach unten auf die nächste Ebene
- Knoten ohne Kante nach unten heißen Blätter oder Endknoten
- Die anderen Knoten heißen innere Knoten
- Falls jeder innere Knoten zwei nach unten führende Kanten hat, handelt es sich um einen vollen Binärbaum.

1-1-Zuordnung von binären Präfixcodes und Binärbäumen:

- Die Kanten der inneren Knoten entsprechen 0 bzw 1
- Buchstaben von S entsprechen den Blättern
- Codewörter entsprechen dem eindeutigen Weg abwärts von der Wurzel
- Länge des Codes = Tiefe des Blattes

Fano-Kraft-Ungleichung

(Satz) Sei k ein Präfixcode für ein Alphabet S. Sei $\ell(a)$ die Länge des Codewortes von $a \in S$. Dann gilt

$$\sum_{a \in S} 2^{-\ell(a)} \leq 1$$

- (Beweis)
- (Gleichheit, strikte Ungleichheit)
- Konsequenz: Zu jeder Funktion $\ell : S \to \mathbb{N}$ welche die Fano-Kraft-Ungleichung erfüllt, kann ein Präfixcode gefunden werden, dessen Wortlängen durch ℓ gegeben sind.

Zufällige Buchstaben

Sei X ein zufälliger Buchstabe, also eine Zufallsvariable mit Wertebereich S. Sei ρ die Häufigkeitsverteilung der Buchstaben von S, also

$$\mathbb{P}(X = a) = \rho(a), \quad a \in S.$$

Die erwartete Wortlänge eines Codes k mit Längenfunktion ℓ ist damit

$$\mathbb{E}[\ell(X)] = \sum_{a \in S} \ell(a)\rho(a).$$

- Sparsames Codieren: Häufige Buchstaben sollen kürzere Codewörter haben. Optimal: Huffman-Code.
- Annähernd optimal: Shannon-Code.
- Schranken: Quellencodierungssatz.
- Beispiel: Gleichverteilte Buchstaben

Shannon-Codes

Ein Shannon-Code ist ein Präfix-Code, bei dem die Wortlängen

$$-\log_2 \rho(a) \leq \ell(a) < -\log_2 \rho(a) + 1$$

erfüllen, Notation: $\ell(a) = \lceil -\log_2 \rho(a) \rceil$.

- Existenz von Shannon-Codes für jede Häufigkeitsverteilung ρ aus der Fano-Kraft-Ungleichung
- Erwartete Wortlänge eines Shannon-Codes:

$$\mathbb{E}[\ell(X)] = \sum_{a \in S} \lceil -\log_2 \rho(a) \rceil \cdot \rho(a)$$

Der Quellencodierungssatz

Für einen zufälligen Buchstaben aus S mit Verteilung ρ definiere

$$H_2[X] := -\sum_{a \in S} \rho(a) \log_2 \rho(a).$$

(Quellencodierungssatz) Für jeden binären Präfixcode gilt

$$\mathbb{E}[\ell(X)] \geq H_2[X].$$

Für binäre Shannon-Codes gilt außerdem $\mathbb{E}[\ell(X)] < H_2[X] + 1$.

- Kürzer als $H_2[X]$ kann also kein binärer Präfix-Code im Mittel werden
- Shannon-Codes sind bezüglich der Länge fast optimal (Abweichung höchstens ein Bit)
- (Beweis)
- Die Schranken könen nicht verbessert werden.
- Für einen Shannon-Code mit $\rho(a) = 2^{-\ell(a)}$ ist $\mathbb{E}[\ell(X)] = H_2[X]$.

Der Quellencodierungssatz

Für einen zufälligen Buchstaben aus S mit Verteilung ρ definiere

$$H_2[X] := -\sum_{a \in S} \rho(a) \log_2 \rho(a).$$

(Quellencodierungssatz) Für jeden binären Präfixcode gilt

$$\mathbb{E}[\ell(X)] \geq H_2[X].$$

Für binäre Shannon-Codes gilt außerdem $\mathbb{E}[\ell(X)] < H_2[X] + 1$.

- Kürzer als $H_2[X]$ kann also kein binärer Präfix-Code im Mittel werden
- Shannon-Codes sind bezüglich der Länge fast optimal (Abweichung höchstens ein Bit)
- (Beweis)
- Die Schranken könen nicht verbessert werden.
- Für einen Shannon-Code mit $\rho(a) = 2^{-\ell(a)}$ ist $\mathbb{E}[\ell(X)] = H_2[X]$.

Huffman-Codes

(Def.) Ein optimaler Code ist ein Präfixcode mit kleinstmöglicher erwarteter Wortlänge $\mathbb{E}[\ell(X)]$.

- Nach dem Quellencodierungssatz gilt für einen optimalen Code immer $H_2[X] \leq \mathbb{E}[\ell(X)] < H_2[X] + 1$.

Beobachtungen:

- Für optimale binäre Präfixcodes ist der Codebaum voll.
- Für optimale binäre Präfixcodes folgt aus $\rho(a) < \rho(b)$ stets $\ell(a) \geq \ell(b)$.
- Für zwei Buchstaben $u, v \in S$ mit kleinster Wahrscheinlichkeit (also $\rho(u) \leq \rho(v) \leq \rho(a) \forall a \notin \{u, v\}$) gilt $\ell(v) = \ell(v) \geq \ell(a)$.

Huffman-Codes: Konstruktion

- Konstruktion von den Blättern zur Wurzel
- Für zwei Buchstaben $u, v \in S$ mit kleinster Wahrscheinlichkeit kann man nach der vorigen Folie annehmen, dass ihre Codewörter $k(u)$ und $k(v)$ sich nur an der letzten Stelle unterscheiden. Damit sitzen Sie an derselben Gabel im entsprechenden Code-Baum.
- Somit können sie zu einem neuen Wort $\langle uv \rangle$ verschmolzen werden, und vom Alphabet S zum kleineren Alphabet $S' \setminus \{u, v\} \cup \{\langle uv \rangle\}$ übergehen, wobei $\langle uv \rangle$ die Wahrscheinlichkeit $\rho(\langle uv \rangle) = \rho(u) + \rho(v)$ erhält. Im Baum wird entsprechend die Gabel für u und v durch einen Knoten für $\langle uv \rangle$ ersetzt.
- Ist der ursprüngliche Baum optimal für S, so ist der so konstruierte neue Baum optimal für S'.

(Beispiel)

Huffman-Codes

- Zur Konstruktion von Huffman-Codes muss die Verteilung ρ bekannt sein
- Es gibt Methoden, die unter geeigneten Voraussetzungen asymptotisch optimal sind, welche unabhängig von der Verteilung sind.
- Rekursive Berechnung der erwarteten Codewortlänge:

$$E(\rho(a_1), ..., \rho(a_n)) = \rho(a_1)+\rho(a_2)+E(\rho(a_1)+\rho(a_2), \rho(a_3), ..., \rho(a_n)),$$

wobei $E(\rho(a_1), ..., \rho(a_n))$ die erwartete Codewortlänge für die Buchstabenverteilung ρ ist.

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: Gleichverteilt, geometrisch)