Informatik IV - Tutorium XVI & XVII

Tut Nr. 6 – Informationstheorie

David Münch

Universität Karlsruhe (TH)
Fakultät für Informatik
IBDS Prautzsch

Dank an Yusuf für die Zusammenarbeit.

10. Juni 2009

Universität Karlsruhe (TH)
Forschungsuniversität • gegründet 1825

Inhaltsverzeichnis

Inhaltsverzeichnis

Inhaltsverzeichnis

Inhaltsverzeichnis

Organisatorisches

Email: muenchdavid@gmail.com

https://www.stud.uni-karlsruhe.de/~uhbro/

Tutorium 16: Freitags 8:00 Uhr - Raum -107
Tutorium 17: Freitags 9:45 Uhr - Raum -119

Schein / Übungsblätter

Bearbeitete Hausaufgaben können abgeben werden.
Es gibt vier Scheinklausuren mit je 50 Punkten an folgenden Terminen:

- 19.05.2009 $\surd$ $\varnothing 31$ Punkte
- 09.06.2009 $\surd$ $\varnothing 35$ Punkte
- 02.07.2009
- 23.07.2009

120 Punkte aus den Scheinklausuren sind notwendig, um einen Schein zu erhalten.

Schein / Übungsblätter

Bearbeitete Hausaufgaben können abgeben werden.
Es gibt vier Scheinklausuren mit je 50 Punkten an folgenden Terminen:

- 19.05.2009 $\surd$ $\varnothing 31$ Punkte
- 09.06.2009 $\surd$ $\varnothing 35$ Punkte
- 02.07.2009
- 23.07.2009

120 Punkte aus den Scheinklausuren sind notwendig, um einen Schein zu erhalten.

Ab 140 Punkten gibt es einen Notenbonus von ca. 1/3 Note.

Schein / Übungsblätter

Bearbeitete Hausaufgaben können abgeben werden.
Es gibt vier Scheinklausuren mit je 50 Punkten an folgenden Terminen:

- 19.05.2009 $\surd$ $\varnothing 31$ Punkte
- 09.06.2009 $\surd$ $\varnothing 35$ Punkte
- 02.07.2009
- 23.07.2009

120 Punkte aus den Scheinklausuren sind notwendig, um einen Schein zu erhalten.

Ab 140 Punkten gibt es einen Notenbonus von ca. 1/3 Note.

Abgabe in Zweiergruppe erlaubt und ausdrücklich erwünscht!

Schein / Übungsblätter

Bearbeitete Hausaufgaben können abgeben werden.
Es gibt vier Scheinklausuren mit je 50 Punkten an folgenden Terminen:

- 19.05.2009 $\surd$ $\varnothing 31$ Punkte
- 09.06.2009 $\surd$ $\varnothing 35$ Punkte
- 02.07.2009
- 23.07.2009

120 Punkte aus den Scheinklausuren sind notwendig, um einen Schein zu erhalten.

Ab 140 Punkten gibt es einen Notenbonus von ca. 1/3 Note.

Abgabe in Zweiergruppe erlaubt und ausdrücklich erwünscht!

Bitte folgendes Deckblatt verwenden:
`http://www.stud.uni-karlsruhe.de/~unbdh/deckblatt/`

Literatur

Boehm, Prautzsch: Numerical Methods. AK Peters 1993. ISBN 3-528-06350-5
`http://www.ubka.uni-karlsruhe.de/hylib-bin/suche.cgi?opacdb=UBKA_OPAC&fbt=7319953&nd=3204657`

Ash: Information Theory. Dover 1990. ISBN 0-486-66521-6
`http://www.ubka.uni-karlsruhe.de/hylib-bin/suche.cgi?opacdb=UBKA_OPAC&nd=9866904`

Goos: Vorlesungen über Informatik. Bd. 4, Springer 1998. ISBN 3-540-60650-5
`http://www.ubka.uni-karlsruhe.de/hylib-bin/suche.cgi?opacdb=UBKA_OPAC&fbt=9316367&nd=6568301`

Was wollen wir heute erreichen?

Was wollen wir heute erreichen?

- 2. Scheinklausur besprechen

Was wollen wir heute erreichen?

- 2. Scheinklausur besprechen
- Einführen von Grundbegriffen der Informationstheorie

Was wollen wir heute erreichen?

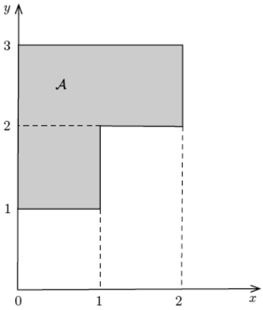
- 2. Scheinklausur besprechen
- Einführen von Grundbegriffen der Informationstheorie
- Überblick über Codes geben

Was wollen wir heute erreichen?

- 2. Scheinklausur besprechen
- Einführen von Grundbegriffen der Informationstheorie
- Überblick über Codes geben
- Kanäle verstehen

Aufgabe 1

Konstruieren Sie ein neuronales Netz mit $-1/1$-Neuronen, zwei Eingabe- und einem Ausgabeneuron, um die in der Abbildung gezeigte geschlossene Figur $\mathcal{A}(\subset \mathbb{R}^2)$ zu klassifizieren. Falls die Eingabe $[x\ y]^t \in \mathcal{A}$, soll die Ausgabe 1 sein und anderenfalls -1. (Zur Erinnerung: Ein Neuron bekommt bei nicht-negativer Erregung den Wert 1 und sonst -1.)



Aufgabe 2

Konstruieren Sie für jede der folgenden boolschen Funktionen ein neuronales Netz, das diese berechnet. Benutzen Sie binäre Neuronen. Der Wert 1 soll dabei TRUE bedeuten; der Wert 0 bzw. -1 den Wahrheitswert FALSE.

(1) Modellieren Sie die Funktion $f(A,B) = (A \Rightarrow B)$ mit $-1/1$-Neuronen.

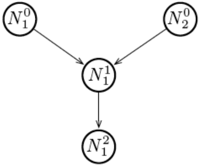
(2) Modellieren Sie die Funktion $f(A,B,C) = (A \text{ AND } \neg B) \text{ OR } C$ mit $-1/1$-Neuronen.

(3) Modellieren Sie die Funktion $f(A,B,C) = A \text{ XOR } B \text{ XOR } C$ mit 0/1-Neuronen, wobei

$$X \text{ XOR } Y = \text{TRUE} \quad :\Leftrightarrow \quad \begin{array}{l} (X = \text{TRUE} \;\; \text{AND} \;\; Y = \text{FALSE}) \quad \text{OR} \\ (Y = \text{TRUE} \;\; \text{AND} \;\; X = \text{FALSE}). \end{array}$$

Aufgabe 3

Wir betrachten ein vorwärtsgerichtetes Perzeptron wie in der folgenden Abbildung mit 2 Schichten, Zustandsmenge $Q = \mathbb{R}$, linearer Erregungsfunktion $h(x) = x$, Gewichten $\omega_{ij}^k = 1$, Schwellwerten $\sigma_k = 0$ und Lernmuster $(\mathbf{e}, \mathbf{a}) = ([1, 2]^t, [2])$.

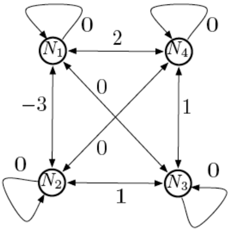


(1) Berechnen Sie die Ausgabe $\mathbf{q}_2$ ausgehend von Eingabe $\mathbf{e}$.

(2) Berechnen Sie den Wert der Fehlerfunktion.

(3) Führen Sie einen Schritt des Lernens mit Rückkopplung und mit Lernrate $\eta = 0{,}1$ durch.

Aufgabe 4



Die Abbildung zeigt ein Hopfield-Netz mit $-1/1$-Neuronen, bei dem alle Schwellwerte Null sind.

(1) Berechnen Sie die Energie des Netzes zu den Zuständen $\mathbf{q}_i, i \in \mathbb{N}$, die sich bei synchroner Schaltung aus dem Zustand $\mathbf{q}_{i-1}$ ergeben, wenn $\mathbf{q}_0 = [1\ 1\ 1\ 1]^t$. (Zur Erinnerung: Nicht erregte Zustände bleiben unverändert.)

(2) Ist der Zustand $\mathbf{q} = [1\ \ -1\ 1\ 1]^t$ stabil bei synchroner Schaltung? Begründen Sie.

Aufgabe 5

Gegeben sei ein Hopfield-Netz mit n Neuronen. Seien alle Schwellwerte Null. Seien $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_p$ $p\ (1 < p \leq n)$ paarweise orthogonale Lernmuster und sei $W = \frac{1}{n} \sum_{i=1}^{p} \mathbf{a}_i \mathbf{a}_i^t - \frac{p}{n} E$ die Gewichtsmatrix.

(1) Sind alle $\mathbf{a}_j$ immer stabil für $Q = \{-1,\ 1\}$?

(2) Sind alle $\mathbf{a}_j$ immer stabil für $Q = \{0,\ 1\}$?

Begründen Sie Ihre Behauptungen durch einen Beweis oder durch ein Gegenbeispiel.

Grundbegriffe der Informationstheorie

Kommunikationssystem

Rauschen

$\downarrow$

Quelle $\rightarrow$ **Codierer** $\rightarrow$ **Kanal** $\rightarrow$ **Decodierer** $\rightarrow$ **Empfänger**

Die Informationstheorie versucht nun die einzelnen Blöcke durch ein mathematisches Modell zu bescheiben.

Zentrale Fragen der Informationstheorie

- Wieviel Information enthält eine Nachricht?

Zentrale Fragen der Informationstheorie

- Wieviel Information enthält eine Nachricht?
- Wieviel von der Nachricht ist überflüssig?

Zentrale Fragen der Informationstheorie

- Wieviel Information enthält eine Nachricht?
- Wieviel von der Nachricht ist überflüssig?
- Kann die Nachricht noch gelesen werden wenn Störungen auftreten?

Beispiel:
Sendet man statt einer 1 n Einsen, dann kann man für $n \longrightarrow \infty$ die Fehlerwahrscheinlichkeit beliebig klein machen auf Kosten der Übertragungsgeschwindigkeit.

Beispiel:
Sendet man statt einer 1 n Einsen, dann kann man für $n \to \infty$ die Fehlerwahrscheinlichkeit beliebig klein machen auf Kosten der Übertragungsgeschwindigkeit.

Fundamentalsatz der Informationstheorie bzw. Shannonsches Kanalcodierungstheorem

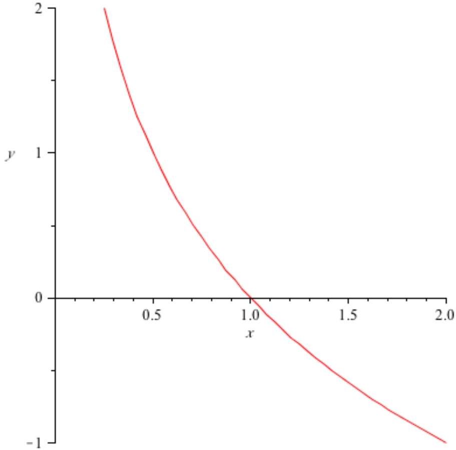
Für einen rauschbehafteten Kanal mit der Kanalkapazität C existiert ein Kodierungsverfahren, so dass für eine Übertragungsrate $R < C$ die Fehlerwahrscheinlichkeit am Empfänger beliebig klein gemacht werden kann. Das heisst, dass es dann theoretisch möglich ist, Information fehlerfrei zu übertragen. Für $R > C$ ist keine fehlerfreie Übertragung möglich.

Entropie

Definition:

Sei X eine Zufallsvariable mit den unabhängigen Ereignissen $x_1, \ldots, x_m$ und sei $p_i = p(x_i) = P(X = x_i)$ die Wahrscheinlichkeit eines Ereignisses x_i, das mit der Unsicherheit (Information) $h(x_i)$ behaftet ist.
Eine Ereignissfolge $x_i x_j$ tritt mit der Wahrscheinlichkeit $p_{ij} = p(x_i x_j) = p_i \cdot p_j$ auf und hat die Unsicherheit $h(p_{ij}) = h(p_i) + h(p_j)$.
U.a. daraus folgt, dass die Unsicherheitsfunktion (oder Information eines Zeichens) von der Gestalt $h(p) = -\log_b p$ sein muss.

Entropie



Entropie

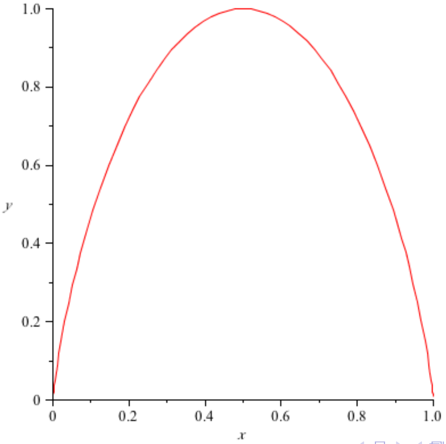
Definition: Entropie

Die Entropie (Informationsgehalt) einer Zufallsvariablen X ist
$H(X) = H(p_1, ..., p_m) = -\sum_{i=1}^{m} p_i \log p_i$

Die Entropie ist die Information pro Zeichen, die wir erwarten.
Oder: Die Entropie ist die durchschnittliche Anzahl von Entscheidungen (bits), die benötigt werden, um ein Zeichen aus einer Zeichenmenge zu identifizieren oder zu isolieren.

Entropie

Beispiel: Entropie von Münzen: $H(p, 1-p)$



Aufgabe

Ein Empfänger beobachtet eine Quelle über dem Alphabet $X = \{A, B, C, D, R\}$, die folgende Zeichen aussendet:

$$S := ABRACADABRA$$

a) Berechne die Wahrscheinlichkeitsverteilung von S.
b) Welche Information hat das Zeichen A, welche das Zeichen C?
c) Wieviel Bits sind nötig um S zu codieren?

Lösung

a) $|S| = 11$
$\Rightarrow p_A = 5/11, p_B = 2/11, p_C = 1/11, p_D = 1/11, p_R = 2/11$

Lösung

a) $|S| = 11$
$\Rightarrow p_A = 5/11, p_B = 2/11, p_C = 1/11, p_D = 1/11, p_R = 2/11$

b) Mit $I(x_i) = h(p_i) = -\log p_i$ folgt:
$I(A) = -\log 5/11 \approx 1,137 bit$
$I(C) = -\log 1/11 \approx 3,459 bit$

Lösung

a) $|S| = 11$
$\Rightarrow p_A = 5/11, p_B = 2/11, p_C = 1/11, p_D = 1/11, p_R = 2/11$

b) Mit $I(x_i) = h(p_i) = -\log p_i$ folgt:
$I(A) = -\log 5/11 \approx 1,137bit$
$I(C) = -\log 1/11 \approx 3,459bit$

c) Berechne Informationsgehalt der kompletten Zeichenkette.
Entropie:
$H(X) = -\sum_{i=1}^{5} p_i \log p_i \approx 2,04bit/Zeichen$
Multiplizieren von $H(X)$ mit der Länge der Zeichenkette:
$H(X) \cdot |S| = 22,444bit$
$\Rightarrow$ minimale Anzahl benötigter Bits beträgt 23.

Gemeinsame Entropie

Seien X und Y zwei Zufallsvariablen mit den Werten $x_1, ..., x_m$ bzw. $y_1, ..., y_n$.
Sei $p_{ij} = p(x_i y_j) = P(X = x_i \text{ und } Y = y_j)$.
Wenn x_i und y_i (wie oben) unabhängig sind,
dann gilt: $p_{ij} = p(x_i) \cdot p(y_j)$
sonst $p_{ij} = p(x_i) \cdot p(y_j | x_i) = p(y_j) \cdot p(x_i | y_j)$.

Definition: Gemeinsame Entropie

Die Gemeinsame Entropie von X und Y ist
$H(X, Y) = H(p_{11}, .., p_{mn}) = -\sum_{i,j} p_{ij} \log p_{ij}$

Gemeinsame Entropie

Seien X und Y zwei Zufallsvariablen mit den Werten $x_1, ..., x_m$ bzw. $y_1, ..., y_n$.
Sei $p_{ij} = p(x_i y_j) = P(X = x_i$ und $Y = y_j)$.
Wenn x_i und y_i (wie oben) unabhängig sind,
dann gilt: $p_{ij} = p(x_i) \cdot p(y_j)$
sonst $p_{ij} = p(x_i) \cdot p(y_j|x_i) = p(y_j) \cdot p(x_i|y_j)$.

Definition: Gemeinsame Entropie

Die Gemeinsame Entropie von X und Y ist
$H(X, Y) = H(p_{11}, .., p_{mn}) = -\sum_{i,j} p_{ij} \log p_{ij}$

Satz

$H(X_1, ..., X_n) \leq H(X_1) + ... + H(X_n)$, mit Gleichheit $\Leftrightarrow X_i$ unabhängig.

Bedingte Entropie

Definition: Bedingte Entropie

$H(X|Y) = -\sum_{i,j} p_{ij} \cdot \log p(x_i|y_j)$

Bedingte Entropie

Definition: Bedingte Entropie

$H(X|Y) = -\sum_{i,j} p_{ij} \cdot \log p(x_i|y_j)$

Satz

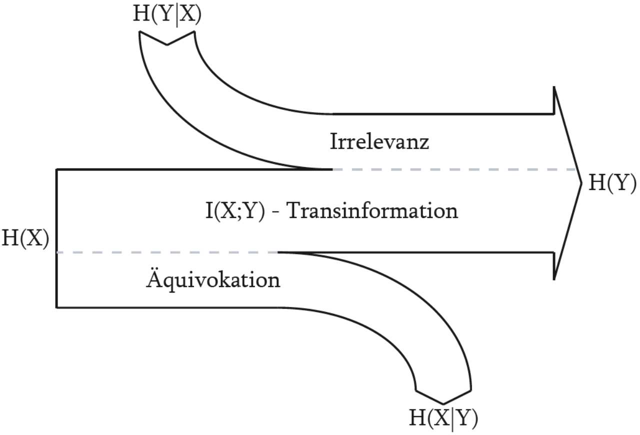
$H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$

Information

Definition: Information

Information ist die beseitigte Unsicherheit. D.h. die durch Kenntnis von Y über X erzielbare Information ist:
$I(X|Y) = H(X) - H(X|Y) = I(Y|X) \geq 0$

Übertragungskanal



Zusammenfassung

Transinformation: $I(X|Y) = H(X) - H(X|Y) = I(Y|X)$

Zusammenfassung

Transinformation: $I(X|Y) = H(X) - H(X|Y) = I(Y|X)$
Äquivokation: $H(X|Y) = H(X) - I(X|Y)$

Zusammenfassung

Transinformation: $I(X|Y) = H(X) - H(X|Y) = I(Y|X)$
Äquivokation: $H(X|Y) = H(X) - I(X|Y)$
Irrelevanz: $H(Y|X) = H(Y) - I(X|Y)$

Zusammenfassung

Transinformation: $I(X|Y) = H(X) - H(X|Y) = I(Y|X)$
Äquivokation: $H(X|Y) = H(X) - I(X|Y)$
Irrelevanz: $H(Y|X) = H(Y) - I(X|Y)$
Totalinformation: $H(Y|X) + I(X|Y) + H(X|Y)$

Kanaleigenschaften

- deterministisch: $H(Y|X) = 0$

Kanaleigenschaften

- deterministisch: $H(Y|X) = 0$
- verlustfrei: $H(X|Y) = 0$

Kanaleigenschaften

- deterministisch: $H(Y|X) = 0$
- verlustfrei: $H(X|Y) = 0$
- störungsfrei: $H(X|Y) = H(Y|X) = 0$

Kanaleigenschaften

- deterministisch: $H(Y|X) = 0$
- verlustfrei: $H(X|Y) = 0$
- störungsfrei: $H(X|Y) = H(Y|X) = 0$
- nutzlos: $I(X|Y) = 0$

Kanaleigenschaften

- deterministisch: $H(Y|X) = 0$
- verlustfrei: $H(X|Y) = 0$
- störungsfrei: $H(X|Y) = H(Y|X) = 0$
- nutzlos: $I(X|Y) = 0$

Die Zufallsvariablen X und Y sind genau dann stochastisch unabhängig, wenn der Kanal nutzlos ist.
$H(X,Y) = H(X) + H(Y)$

Kanalkapazität

Definition: Kanalkapazität

$C = max_{P(X)}\{I(X|Y)\}$

C ist die höchste Informationsmenge, die unter allen möglichen Quellenverteilungen über den Kanal übertragen werden kann.

Aufgabe

Gegeben sei folgender binärer, asymetrischer Kanal über den Alphabeten $\mathcal{X} = \mathcal{Y} = \{0, 1\}$:

$$\begin{array}{cc} P(0|0) = 1 - \beta & P(1|0) = \beta \\ P(0|1) = 0 & P(1|1) = 1 \end{array}$$

a) Für welches β hat der Kanal maximale Kapazität?

Aufgabe

Gegeben sei folgender binärer, asymetrischer Kanal über den Alphabeten $\mathcal{X} = \mathcal{Y} = \{0, 1\}$:

$$\begin{array}{cc} P(0|0) = 1 - \beta & P(1|0) = \beta \\ P(0|1) = 0 & P(1|1) = 1 \end{array}$$

a) Für welches β hat der Kanal maximale Kapazität?

b) Für welches $H(X)$ wird bei gegebener Wahrscheinlichkeitsverteilung β der Wert von $H(Y)$ maximal?

Aufgabe

Gegeben sei folgender binärer, asymetrischer Kanal über den Alphabeten $\mathcal{X} = \mathcal{Y} = \{0, 1\}$:

$$\begin{array}{cc} P(0|0) = 1 - \beta & P(1|0) = \beta \\ P(0|1) = 0 & P(1|1) = 1 \end{array}$$

a) Für welches β hat der Kanal maximale Kapazität?

b) Für welches $H(X)$ wird bei gegebener Wahrscheinlichkeitsverteilung β der Wert von $H(Y)$ maximal?

c) Berechne Irrelevanz, Äquivokation und Transinformation für $\beta = 0.9$ und Gleichverteilung auf X.

Redundanz

Redundanz ist überflüssige Information. D.h. sie sieht aus wie Information, ist aber keine.

Definition: Absolute Redundanz

$R_{abs} = \tilde{H} - H$ mit

Realinformation (Entropie)
$H = \mathbb{E}(h(p_i)) = -\sum_{i=1}^{n} p_i \cdot h(p_i) = -\sum_{i=1}^{n} p_i \cdot \log p_i$

Nominalinformation
$\tilde{H} = \mathbb{E}(\text{Länge}(code_i)) = \sum_{i=1}^{n} p_i \cdot \text{Länge}(code_i)$

Redundanz

Redundanz ist überflüssige Information. D.h. sie sieht aus wie Information, ist aber keine.

Definition: Absolute Redundanz

$R_{abs} = \tilde{H} - H$ mit

Realinformation (Entropie)
$H = \mathbb{E}(h(p_i)) = -\sum_{i=1}^{n} p_i \cdot h(p_i) = -\sum_{i=1}^{n} p_i \cdot \log p_i$

Nominalinformation
$\tilde{H} = \mathbb{E}(\text{Länge}(code_i)) = \sum_{i=1}^{n} p_i \cdot \text{Länge}(code_i)$

Definition: Relative Redundanz

$R_{rel} = 1 - \frac{H}{\tilde{H}}$

Definition: Codierung

Sei $\mathbb{A} = \{a_1, .., a_m\}$ ein Alphabet und A eine Zufallsvariable über $\mathbb{A}$ mit den Wahrscheinlichkeiten $p_i = P(A = a_i)$. Eine Codierung von A über einem Codealphabet $\mathbb{X} = \{x_1, .., x_d\}$ ist eine Abbildung: $C : \mathbb{A} \to \mathbb{X}^+$. Diese wird für ganze Worte erweitert: $C^* : \mathbb{A}^* \to \mathbb{X}^*$, $a_{i_1}...a_{i_k} = c_{i_1}...c_{i_k}$

Definition: Codierung

Sei $\mathbb{A} = \{a_1, .., a_m\}$ ein Alphabet und A eine Zufallsvariable über $\mathbb{A}$ mit den Wahrscheinlichkeiten $p_i = P(A = a_i)$. Eine Codierung von A über einem Codealphabet $\mathbb{X} = \{x_1, .., x_d\}$ ist eine Abbildung: $C : \mathbb{A} \to \mathbb{X}^+$. Diese wird für ganze Worte erweitert: $C^* : \mathbb{A}^* \to \mathbb{X}^*$, $a_{i_1} ... a_{i_k} = c_{i_1} ... c_{i_k}$

Definition: Codelänge

Die mittlere Codelänge ist $L(C) = \sum_{i=1}^{m} p_i \cdot l_i$

Codeeigenschaften

- regulär, wenn C injektiv

Codeeigenschaften

- regulär, wenn C injektiv
- dekodierbar, wenn C^* injektiv

Codeeigenschaften

- regulär, wenn C injektiv
- dekodierbar, wenn C^* injektiv
- Präixcode, wenn kein c_i Präfix eines andern c_j ist.

Konstruktion optimaler Codes

Satz

Für jeden Präfix- und dekodierbaren Code gilt die Kraft-Ungleichung:

$$\sum_{i=1}^{m} d^{-l_i} \leq 1$$

Wenn diese Gleichung erfüllt ist, gibt es einen Präfix- bzw. dekodierbaren Code mit diesen Längen l_i

Kodierungstheorem

Theorem

Die Länge eines dekodierbaren Codes C ist mindestens so gross wie die Entropie der Zufallsvariablen A:
$L(C) \geq H_d(p)$
mit Gleichheit $\Leftrightarrow d^{-l_i} = p_i$ für alle i.

Kodierungstheorem

Theorem

Die Länge eines dekodierbaren Codes C ist mindestens so gross wie die Entropie der Zufallsvariablen A:
$L(C) \geq H_d(p)$
mit Gleichheit $\Leftrightarrow d^{-l_i} = p_i$ für alle i.

Ein optimaler Präfixcode L^* hat die Länge:
$H_d(p) \leq L^* < L(C) < H_d(p) + 1$

Quellen

Pajor - Informatik 4 Tutorium SS2007

Prautzsch - Skript Informatik 4 SS2008

Ash: Information Theory. Dover 1990. ISBN 0-486-66521-6
http://www.ubka.uni-karlsruhe.de/hylib-bin/suche.cgi?opacdb=UBKA_OPAC&nd=9866904

Reflexion

Was haben wir heute gelernt?

Reflexion

Was haben wir heute gelernt?

- 2. Scheinklausur besprochen

Reflexion

Was haben wir heute gelernt?

- 2. Scheinklausur besprochen
- Grundbegriffe der Informationstheorie kennen gelernt

Reflexion

Was haben wir heute gelernt?

- 2. Scheinklausur besprochen
- Grundbegriffe der Informationstheorie kennen gelernt
- Verschiedene Codes betrachtet

Reflexion

Was haben wir heute gelernt?

- 2. Scheinklausur besprochen
- Grundbegriffe der Informationstheorie kennen gelernt
- Verschiedene Codes betrachtet
- Diskrete Kanäle behandelt

Noch Fragen?

Vorschau

Vorschau

- Kanäle

Vorschau

- Kanäle
- Fehlerkorrigierende Codes

Bis zum nächsten Mal