Elemente der Kodierungstheorie

A endliches Alphabet, $|A| = q$

betrachte Wörter der (festen) Länge n über A,

d.h. Elemente von $A^n = \underbrace{A \times \dots \times A}_{n \text{ mal}} = \{(a_1, \dots, a_n) \mid a_i \in A\}$

A^n heißt auch

Hamming-Raum $H(n, q)$ bzw. $H(n, A)$

oft: $A = \mathbb{F}_q$ Körper mit q Elementen (d.h. q ist Primzahlpotenz)

$H(n, \mathbb{F}_q) = \mathbb{F}_q^n$ n-dimensionaler $\mathbb{F}_q$-Vektorraum

Insbesondere $q = 2$, $\mathbb{F}_2 = \{0, 1\}$, $H(8, \mathbb{F}_2)$ Bytes

Bspi:

- ASCII-Code (uralter)

$(a_1, \dots, a_7, a_8)$ ← Kontrollziffer

a_8 so, dass $\sum_{i=1}^{8} a_i = 0$ in $\mathbb{F}_2$

Binärstell, eindeutige Codierung eines Zeichens, z.B.

Buchstaben, Satzzeichen, Steuerungszeichen

„parity check"

Code erkennt einen Fehler!

$(1,1,0,0,1,0,1,0)$

- alte ISBN-Nummer

10stellig $(b_1, \dots, b_9, b_{10})$ ← Kontrollzahl

eindeutige Information

$b_1, \dots, b_9 \in \{0,1,\dots,9\}$, $b_{10} \in \{0,1,\dots,9,X\}$

$(1,0,0,0,1,0,1,0)$

$\sum_{j=1}^{10} j \cdot b_j = 0$ in $\mathbb{F}_{11}$

erkennt eine falsche Ziffer und Vertauschungen von 2 Ziffern

- neue ISBN-Code, 13-stellig $(c_1, \dots, c_{13})$ $c_i \in \mathbb{Z}/10\mathbb{Z}$

$1 \cdot c_1 + 3c_2 + 1 \cdot c_3 + 3c_4 + \dots = 0$ in $\mathbb{Z}/10\mathbb{Z}$ beachte $1, 3 \in (\mathbb{Z}/10\mathbb{Z})^*$

erkennt eine falsche Ziffer und manche Vertauschungen

Wie misst man Anzahl der Fehler?

Hier nur falsche „Ziffern", „falsche Bits"

Anzahl Fehler = Anzahl der Stellen mit falscher Information

(insbesondere keine Gewichtung innerhalb des Alphabets)
und keine Gewichtung der Stelle

Hamming-Metrik (oder Hamming-Abstand)

$\bar{v} = (v_1, \dots, v_n)$, $\bar{u} = (u_1, \dots, u_n) \in H(n,q)$

$$d(\bar{v}, \bar{u}) := | \{ i \in \{1, \dots, n\} \mid u_i \neq v_i \} |$$

„distance"

Eigenschaften:

- $d(\bar{v}, \bar{u}) \geq 0$
  $d(\bar{v}, \bar{u}) = 0 \iff \bar{v} = \bar{u}$
- $d(\bar{v}, \bar{u}) = d(\bar{u}, \bar{v})$
- $d(\bar{u}, \bar{w}) \leq d(\bar{u}, \bar{v}) + d(\bar{v}, \bar{w})$

} klar

|| Metrik

denn $u_i \neq w_i \Rightarrow u_i \neq v_i$ oder $v_i \neq w_i$

Angenommen $(A,+)$ ist eine (kommut.) Gruppe

Dann $d(\bar{v},\bar{u}) = d(\bar{v}+\bar{w}, \bar{u}+\bar{w})$ denn $v_i \neq u_i \Leftrightarrow v_i + w_i \neq u_i + w_i$

("Hamming-Metrik ist translations-invariant")

also $d(\bar{v},\bar{u}) = d(\bar{v}-\bar{u}, \bar{0})$

Außerdem: $d(\bar{v},\bar{u}) = d(-\bar{v}, -\bar{u})$ denn $v_i \neq u_i \Leftrightarrow -v_i \neq -u_i$

Angenommen A ist K-Vektorraum (K endlicher Körper)

$k \neq 0 \Rightarrow d(\bar{v},\bar{u}) = d(k\bar{v}, k\cdot\bar{u})$ denn $v_i \neq u_i \Leftrightarrow k\cdot v_i \neq k\cdot u_i$

Bem: $(A,+)$ kommutative Gruppe, p Primzahl (z.B. $p=2$)

Falls $p \cdot a = \underbrace{a + \dots + a}_{p\text{-mal}} = 0$ für alle $a \in A$, so ist A $\mathbb{F}_p$-Vektorraum.

noch mehr Definitionen ...

- Ein Code ist eine Teilmenge von $H(n,q)$ bzw $H(n,A)$

  „Code der Länge n über A"

  „q-ärer Code der Länge n"

  (binär, ternär, ...)

Minimalabstand des Codes ist $\min\{ d(\bar{u},\bar{v}) \mid \bar{u},\bar{v} \in C,\ \bar{u} \neq \bar{v} \}$

- Ein linearer Code ist ein Untervektorraum von $H(n, \mathbb{F}_q) = \mathbb{F}_q^n$

  „$[n,k]$-Code", falls $k = \dim C$

Gewicht von $\bar{v} \in C$ ist $d(\bar{v},\bar{0}) = wt(\bar{v}) = w(\bar{v})$ („weight")

Minimalgewicht von C ist $\min\{ wt(\bar{v}) \mid \bar{v} \in C,\ \bar{v} \neq 0 \}$

„Minimalabstand von C, da $d(\bar{v},\bar{u}) = d(\bar{v}-\bar{u}, 0)$

„$[n,k,d]$-Code": linearer Code der Länge n, der Dimension k und des Minimalgewichts d

Bem: $C \subseteq \underset{\text{Teilraum}}{\mathbb{F}_p^n}$ p Primzahl

Falls C unter Addition abgeschlossen (d.h. $\bar{c}_1, \bar{c}_2 \in C \Rightarrow \bar{c}_1 + \bar{c}_2 \in C$)
dann ist C bereits ein Untervektorraum.
(denn $\ell \cdot \bar{c} = \underbrace{\bar{c} + \dots + \bar{c}}_{\ell \text{ mal}}$, $-\bar{c} = (p-1) \cdot \bar{c}$)

- Code C erkennt n Fehler, falls
$d(\bar{u}, \bar{v}) > n$ für $\bar{u}, \bar{v} \in C$, $\bar{u} \neq \bar{v}$

Bem: Falls Minimalabstand von $C \geq m$ ist, dann erkennt der Code $m-1$ Fehler

- Code C korrigiert n Fehler, falls für jedes $\bar{v} \in H(n, A)$
höchstens ein $\bar{u} \in C$ existiert mit $d(\bar{v}, \bar{u}) \leq n$

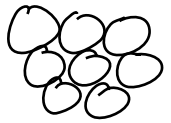
Bem Falls Minimalabstand von $C \geq m$ ist,
dann korrigiert der Code $\lfloor \frac{m-1}{2} \rfloor$ Fehler

Ball vom Radius r um $\bar{c} \in C$ $= \{\bar{v} \in H(n, A) \mid d(\bar{c}, \bar{v}) \leq r\} = B_r(\bar{c})$

C erkennt r Fehler $\Leftrightarrow$ kein $\bar{c}' \in B_r(\bar{c})$ für $\bar{c}, \bar{c}' \in C$, $\bar{c} \neq \bar{c}'$

C korrigiert r Fehler $\Leftrightarrow$ die $B_r(\bar{c})$ für $\bar{c} \in C$ sind paarweise disjunkt
(schneiden sich nicht)

$\rightsquigarrow$ Kugelpackungsproblem



Zwei Beispiele:

Wie korrigiert man 1 Fehler?

eigentliche Information: binäre Wörter der Länge 4

"naive" Methode: 3fache Wiederholung, d.h. $\bar{x} \in H(4, \mathbb{F}_2)$
wird kodiert als $\bar{x} \hat{} \bar{x} \hat{} \bar{x} \in H(12, \mathbb{F}_2)$

$C = \{(c_1, \ldots, c_{12}) \mid c_i \in \mathbb{F}_2,\ c_1 = c_5 = c_9,\ c_2 = c_6 = c_{10}, \text{etc.}\}$

linearer $[12, 4, 3]$-Code

Effizient?

Raumgröße $|H(12,2)| = 2^{12} = 4096$

Anzahl Codewörter $2^4 = 16$

$|B_1(\bar{c})| = 13$

$4096 - 16 \cdot 13 = 3.888$

„verschwendeter Platz"

erkennt 2 Fehler
korrigiert 1 Fehler

1,0,0,0,1,0,0,0,1,0,0,0
0100 ...
0010 ...
0001 ...

0	0 0 0 0 0 0 0
1 -	0 0 0 1 1 1 1
2 -	0 0 1 0 1 1 0
3	0 0 1 1 0 0 1
4 -	0 1 0 0 1 0 1
5	0 1 0 1 0 1 0
6	0 1 1 0 0 1 1
7	0 1 1 1 1 0 0
8 -	1 0 0 0 0 1 1
9	1 0 0 1 1 0 0
10	1 0 1 0 1 0 1
11	1 0 1 1 0 1 0
12	1 1 0 0 1 1 0
13	1 1 0 1 0 0 1
14	1 1 1 0 0 0 0
15	1 1 1 1 1 1 1

linearer $[7,4,3]$- Code

Raumgröße $|H(7,2)| = 2^7 = 128$

Anzahl Codewörter $2^4 = 16$

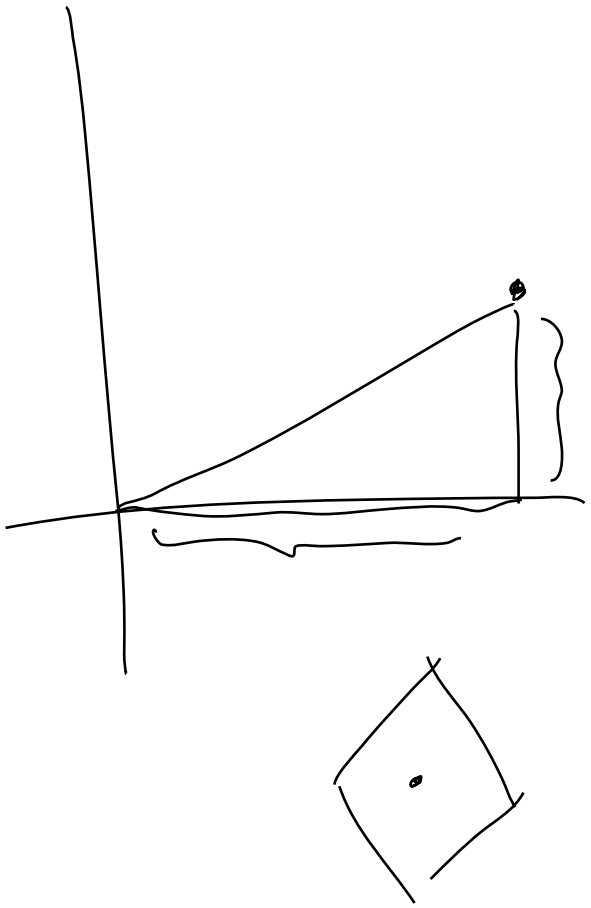
erkennt 2 Fehler

korrigiert 1 Fehler

$|B_1(c)| = 8$

$128 - 16 \cdot 8 = 0$

„perfekter Code"



(v_1, v_2) $\vec{v} \in \mathbb{R}^2$

$$\|\vec{v}\| = \sqrt{v_1^2 + v_2^2}$$

$$\|\vec{v}\|_\infty = |v_1| + |v_2|$$