Kapitel 2

Fehlererkennende Codes

2.1 Die Hamming-Schranke

Ein Code C über dem Alphabet A heißt *n-Code*, wenn jedes Wort darin die Länge n hat, also $C \subseteq A^n$ gilt. Der *Abstand* (auch: *Hamming-Abstand*) zweier Wörter $u = a_1 \dots a_n$, $w = b_1 \dots b_n$ der gleichen Länge n ist die Zahl

$$d(u, w) := |\{i \in \{1, \dots, n\} : a_i \neq b_i\}|,$$

also die Anzahl Stellen, an denen sich u und w unterscheiden. Dies definiert eine *Metrik* $d : A^n \times A^n \to \mathbb{N}$ auf A^n, das heißt es gilt

(i) $d(u, w) = 0$ genau dann wenn $u = w$ für alle $u, w \in A^n$,

(ii) $d(u, w) = d(w, u)$ für alle $u, w \in A^n$, und

(iii) $d(u, v) + d(v, w) \geq d(u, w)$ für alle $u, v, w \in A^n$.

Die *Kugel vom Radius r um $u \in A^n$* ist definiert durch

$$B_r(u) := \{x \in A^n : d(x, u) \leq r\}.$$

Da es genau $\binom{n}{t}(|A| - 1)^t$ Wörter der Länge n gibt, die sich von einem gegebenen Wort aus A^n an genau t Stellen unterscheiden, folgt

$$|B_r(u)| = \sum_{t=0}^{r} \binom{n}{t} (|A| - 1)^t.$$

Der *Minimalabstand* eines n-Codes ist definiert durch

$$d(C) := \min\{d(u, w) : u \neq w \text{ aus } C\}.$$

Nehmen wir an, daß bei der buchstabenweisen Übertragung eines Wortes u aus C an höchstens r der n Stellen Fehler aufgetreten sind. Der Empfänger erhält dann ein Wort w, das an höchstens r Stellen von u abweicht, also aus $B_r(u)$ stammt. Im Fall $d(C) \geq r+1$ ist u das einzige Wort aus C in $B_r(u)$, der Empfänger kann also aus $w \notin C$ auf eine fehlerhafte Übertragung eines (mit höchstens r Fehlern behafteten) Wortes schließen. Ist sogar $d(C) \geq 2r+1$, so ist u das *einzige* Wort aus C des Abstandes höchstens r von w: Wäre $v \neq w$ ein zweites, so folgte ja $d(v,w) \leq d(v,u) + d(w,u) \leq 2r$, was $d(C) \geq 2r+1$ zuwiderläuft. Der Empfänger kann dann (das mit höchstens r Fehlern behaftete) u rekonstruieren. Infolgedessen nennt man einen n-Code C über A *r-fehlererkennend*, wenn $d(C) \geq r+1$ gilt und *r-fehlerkorrigierend*, wenn $d(C) \geq 2r+1$ gilt.

Satz 1 *Für jeden r-fehlerkorrigierenden n-Code C über A gilt*

$$|C| \leq \frac{|A|^n}{\sum_{t=0}^{r} \binom{n}{t} (|A|-1)^t}.$$

Beweis. Für $v \neq w$ aus C kann es kein $u \in B_r(v) \cap B_r(w)$ geben (sonst wäre wie oben $d(v,w) \leq d(v,u) + d(w,u) \leq 2r$, Widerspruch). Daher sind die $B_r(w)$, $w \in C$, paarweise disjunkte $\sum_{t=0}^{r} \binom{n}{t} (|A|-1)^t$-elementige Teilmengen von A^n, und so folgt $|C| \cdot \sum_{t=0}^{r} \binom{n}{t} (|A|-1)^t \leq |A|^n$, also die Behauptung. □

Die obere Schranke aus Satz 1 heißt *Hamming-Schranke* und wird mit $H_A(n,r)$ bezeichnet. Ein r-fehlerkorrigierender n-Code C über A mit $|C| = H_A(n,r)$ heißt *perfekt*. Aus dem Beweis des Satzes folgt unmittelbar, daß dies dann und nur dann der Fall ist, wenn die $B_r(w)$, $w \in C$, eine Partition von A^n bilden (also eine „Kugelpackung"). Tatsächlich ist letztere Bedingung schon hinreichend dafür, daß C r-fehlerkorrigierend ist.

Für einen n-Code C über A definieren wir die *Informationsrate* von C durch

$$r(C) := \frac{\log_{|A|} |C|}{n}.$$

Denken wir uns ein langes Wort w über dem Alphabet A (einen „Klartext"), das wir in Form von Wörtern des n-Codes C über A übertragen wollen. Dazu würden wir w in kleine Teilwörter der Länge ℓ zerlegen und jedem Teilwort $u \in A^\ell$ ein Codewort $f(u) \in C$ zuweisen, welche dann übertragen werden; damit der Empfänger nach der Rekonstruktion von $f(u)$ imstande ist, u zu bestimmen, muß $f(u) \neq f(u')$ für $u \neq u'$ aus A^ℓ gelten, das heißt f injektiv von A^ℓ nach C abbilden. Das geht natürlich nur, wenn $|A^\ell| \leq |C|$ ist, also $\ell \leq \log_{|A|} |C|$ gilt. Zerfällt dann w in k Teilworte der Länge ℓ, hat w also die Länge $\ell \cdot k$, so werden k viele Codewörter, alle der Länge n, übertragen, also insgesamt ein Chiffrat aus $k \cdot n$ Zeichen. Das Verhältnis von Klartextlänge zur Chiffratlänge ist infolgedessen ℓ/n, und wegen

$$\frac{\ell}{n} \leq \frac{\log_{|A|} |C|}{n} = r(C)$$

ist die Informationsrate eine obere Schranke dieses Verhältnisses, die umso genauer ist, je näher $|C|$ oberhalb einer Potenz von $|A|$ liegt.

Betrachten wir den *n-Wiederholungscode über A*, gegeben durch

$$C = \{\underbrace{xx\ldots x}_{n\text{-mal}} : x \in A\}.$$

Dies ist ein n-Code mit $|C| = |A|$, und infolgedessen ist

$$r(C) = \frac{\log_{|A|} |C|}{n} = \frac{1}{n}.$$

Bei der Übertragung ver-n-facht man schlicht jedes Zeichen, das Verhältnis Klartext zu Chiffrat ist also tatsächlich gleich $1/n$. Da der Minimalabstand $d(C)$ dieses Codes gleich n ist, ist zum Beispiel ein 5-Wiederholungscode, etwa über $A = \{0, 1\}$, 2-fehlerkorrigierend. Findet man nun andere 2-fehlerkorrigierenden Codes C' über $\{0, 1\}$ mit mehr und mehr als $|C|$ Wörtern, so wird $\log_{|A|} |C'|$ und damit $r(C)$ größer und größer (jedoch nicht größer als 1). Die Hamming-Schranke $H_{\{0,1\}}(5, 2)$ ist jedoch gleich 2, die Informationsrate eines 2-fehlerkorrigierenden 5-Codes kann daher nicht größer als $1/5$ sein. Läßt man sich aber Spielraum beim n, so kann man zum Beispiel einen 1-fehlerkorrigierenden 8-Code über $A = \{0, 1\}$ mit 16 Wörtern finden (sogar: Minimalabstand 4); ein solcher hat die Informationsrate $\log_2 16/8 = 1/2$.

2.2 Der Satz von Shannon

Sei C ein n-Code über $\mathbb{Z}_2$. Wir denken uns die Übertragung vom Sender zum Empfänger codewortweise, wobei jedes der n Bits mit gleicher Wahrscheinlichkeit p gestört wird. Die Störung für das ganze Wort modellieren wir daher durch einen Vektor $X = (X_1, \ldots, X_n)$ von stochastisch unabhängigen Bernoulli-verteilten Zufallsvariablen X_i mit gleichem Parameter $p \in [0, 1]$. Für ein festes Wort w ist somit das entsprechend gestörte Wort gleich $w + X$ (Rechnung punktweise in $\mathbb{Z}_2$) und ebenfalls eine vektorwertige Zufallsvariable.

Der Empfänger erhält eine konkrete Ausprägung x eines n-Wortes und kennt natürlich den Code C. Durch die Störung kann es geschehen, daß x nicht aus C kommt; der Empfänger sucht dann nach dem „ähnlichsten" Wort aus C, also nach einem Codewort mit kleinstem Hamming-Abstand von x. Wird dieser Abstand von mehreren Wörtern aus C realisiert, so soll die daraus entstehende Ratlosigkeit ebenfalls angezeigt werden. Formal definieren wir so:

Sei $w = w_1, \ldots, w_m \in \mathbb{Z}_2^n$ eine Folge von Wörtern. Sei

$$\begin{aligned} m_w &: \mathbb{Z}_2^n \to \{w_1, \ldots, w_m, ⍰\} \\ m_w(x) &:= \begin{cases} w_i & \text{falls } d_{\mathbb{Z}_2^n}(w_i, x) < d_{\mathbb{Z}_2^n}(w_j, x) \text{ für alle } j \in \mathbb{N}_m \setminus \{i\}, \\ ⍰ & \text{sonst.} \end{cases} \end{aligned}$$

Dabei ist $\mathbb{N}_m := \{1, \ldots, m\}$ und ⍰ ein willkürlich gewähltes Fehlerwort oder -symbol, das jedenfalls nicht in der Folge w vorkommen darf. Die Abbildung m_w ist wohldefiniert (das bedeutet hier, daß i durch die definierende Bedingung eindeutig bestimmt ist) und heißt *Maximum–Likelihood–Decodierung*. Die Wahrscheinlichkeit dafür, daß ein Wort w_i unter obiger Störung X falsch übertragen wird, ist also

$$P(m_w(w_i + X) \neq w_i),$$

und der über alle Wörter w_i gemittelte Wert, also die Zahl

$$p_w := \frac{1}{m} \sum_{i=1}^{m} P(m_w(w_i + X) \neq w_i)$$

heißt *Fehlerwahrscheinlichkeit* der Folge w unter der *Störung* X bei *Maximum–Likelihood–Decodierung* m_w. Man beachte, daß p_w nicht von der Anordnung der w_i innerhalb der Folge abhängt.

Die Angabe der Codewörter als *Folge* ist nur der Ökomie im Beweis des Satzes von Shannon geschuldet. Dort werden wir „zufällige Codewortfolgen" betrachten, deren Verteilung umgänglicher ist als die von „zufälligen Codewortmengen". Niemand wird schließlich ein Codewort ambivalent verwenden wollen (im Falle $w_i = w_j$ für $j \neq i$ ist ja ohnehin $m_w(w_i) = m_w(w_j) =$ ⍰).

Sind entsprechend $w_1, \ldots, w_m$ die m paarweise verschiedenen Wörter eines n-Codes C, so heißt $p_C := p_{w_1,\ldots,w_m}$ *Fehlerwahrscheinlichkeit* von C unter der *Störung* X bei *Maximum–Likelihood–Decodierung* $m_C := m_{w_1,\ldots,w_m}$.

Sei nun zum Beispiel C ein perfekt r-fehlerkorrigierender n-Code über $\mathbb{Z}_2$. Dann gibt es zu jedem $x \in \mathbb{Z}_2^n$ genau ein $w \in C$ mit $d_C(x, w) \leq r$. Für ein beliebiges $w \in C$ ist daher $m_C(w + x) = w$ genau dann, wenn $d(x, 0) \leq r$ gilt, und folglich

$$\begin{aligned} p_C &= \frac{1}{|C|} \sum_{w \in C} P(m_C(w+X) \neq w) \\ &= \frac{1}{|C|} \sum_{w \in C} P(d(X,0) > r) \\ &= P(\underbrace{d(X,0)}_{\text{binomialverteilt!}} > r) \\ &= \sum_{i=r+1}^{n} \binom{n}{i} p^i (1-p)^{n-i}. \end{aligned}$$

Die Stochastik kennt Approximationsverfahren für solche Summen.

Wir können jetzt den Satz von Shannon formulieren. Alle im Rest dieses Abschnitts auftretenden Logarithmen sind zur Basis 2, wir schreiben stets $\log x$ anstelle von $\log_2 x$.

Satz 2 *(Shannon 1948)*

Seien $p \in (0, \frac{1}{2})$ und $0 < r < 1 + p \log p + (1-p)\log(1-p)$. Sei $\varepsilon > 0$.

Dann gibt es zu jedem genügend großen n einen n-Code über $\mathbb{Z}_2$ mit Informationsrate $\frac{\lfloor nr \rfloor}{n}$ und Fehlerwahrscheinlichkeit kleiner ε.

Beweis. Sei Y eine reellwertige Zufallsvariable mit Erwartungswert μ und Varianz σ^2. Dann gilt für jedes $k > 0$

$$P(|Y - \mu| \geq k\sigma) \quad < \quad \frac{1}{k^2}. \tag{1}$$

(Ungleichung von Tschebyscheff, wird in der Stochastik gezeigt.)

Sei $X = (X_1, \ldots, X_n)$ ein Vektor von stochastisch unabhängigen Bernoulliverteilten Zufallsvariablen X_i mit gleichem Parameter $p \in [0, 1]$. Da $Y := d(X, 0)$ eine binomialverteilte Zufallsvariable mit Erwartungswert $\mu := np$ und Varianz $\sigma^2 := np(1-np)$ ist, erhalten wir speziell für $k = (\varepsilon/2)^{-1/2}$ mit (1)

$$P(d(X, 0) \geq \mu + k\sigma) \quad \leq \quad \varepsilon/2.$$

Wegen $p < 1/2$ gilt für genügend großes n

$$d := \lfloor \mu + k\sigma \rfloor \quad < \quad \frac{n}{2}.$$

Entsprechend gilt für die Kugel $B_d(x)$ um $x \in \mathbb{Z}_2^d$ mit Radius d

$$|B_d(x)| \quad = \quad \sum_{i=0}^{d} \binom{n}{i} < \frac{n}{2}\binom{n}{d} \leq \frac{n}{2}\frac{n^n}{d^d(n-d)^{n-d}}, \tag{2}$$

letzteres wegen $n^n = (d + (n-d))^n \geq \binom{n}{d} d^d (n-d)^{n-d}$ (aus dem Binomialsatz).

Für $x, y \in \mathbb{Z}_2^n$ sei

$$f(x, y) := \begin{cases} 0 & \text{für } d(x, y) > d \text{ und} \\ 1 & \text{für } d(x, y) \leq d. \end{cases}$$

Für eine Folge $w_1, \ldots, w_m \in \mathbb{Z}_2^n$ von Wörtern und $i \in \mathbb{N}_m$ sei $g_i : \mathbb{Z}_2^n \to \mathbb{Z}_{\geq 0}$ definiert durch

$$g_i(y) := (1 - f(y, w_i)) + \sum_{j \in \mathbb{N}_m - \{i\}} f(y, w_j).$$

Ist w_i das einzige Wort der Folge mit $d(y, w_i) \leq d$, so ist $g_i(y) = 0$, andernfalls ist $g_i(y) \geq 1$. Man kann sich $g_i(y)$ als eine Summe von „Strafpunkten" vorstellen, die umso höher ist, je schlechter y das Wort w_i repräsentieren kann. Liegt

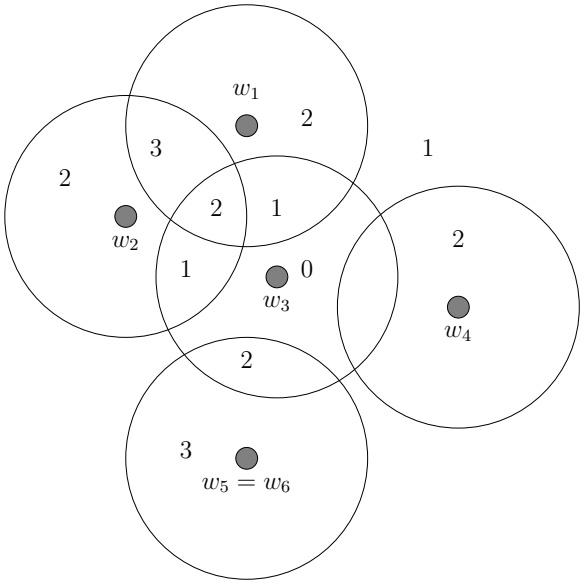

Abbildung 2.1: Zum Beweis des Satzes von Shannon.

w_i „weit entfernt" von y, gibt es schon mal einen Strafpunkt, und für jedes w_j, $j \neq i$, das „nahe" bei y liegt, gibt es einen weiteren.

Abbildung 2.1 zeigt ein mögliches Venn–Diagramm von insgesamt 6 als Kreisscheiben veranschaulichten Kugeln $B_d(w_i)$. Die Zahlen innerhalb der Gebiete geben den Wert der Funktion g_3 dort an.

Wir modifizieren die Maximum-Likelihood-Decodierung m_w für eine Folge $w = w_1, \ldots, w_m \in \mathbb{Z}_2^n$ von Wörtern: Sei

$$\begin{array}{rcl} m'_w & : & \mathbb{Z}_2^n \to \{w_1, \ldots, w_m, ?\} \\ m'_w(x) & := & \begin{cases} w_i & \text{falls } w_i \in B_d(x) \text{ und } w_j \notin B_d(x) \text{ für alle } j \in \mathbb{N}_m \setminus \{i\}, \\ ? & \text{sonst.} \end{cases} \end{array}$$

Auch m'_w ist wohldefiniert (d ist fest). Da aus $m_w(w_i + x) \neq w_i$ auch $m'_w(w_i +$

$x) \neq w_i$ folgt, können wir wie folgt abschätzen.

$$\begin{aligned} P(m_w(w_i{+}X) \neq w_i) &\leq P(m'_w(w_i{+}X) \neq w_i) \\ &\leq \sum_{y\in\mathbb{Z}_2^n} P(w_i{+}X = y)\cdot g_i(y) \\ &= \underbrace{\sum_{y\in\mathbb{Z}_2^n} P(w_i{+}X = y)\cdot(1-f(y,w_i))}_{=P(w_i{+}X\notin B_d(w_i))\leq\varepsilon/2} \\ &\quad + \sum_{y\in\mathbb{Z}_2^n}\sum_{j\in\mathbb{N}_m\setminus\{i\}} P(w_i{+}X = y)\cdot f(y,w_j), \end{aligned}$$

also

$$p_w \leq \varepsilon/2 + \frac{1}{m}\sum_{i=1}^{m}\sum_{y\in\mathbb{Z}_2^n}\sum_{j\in\mathbb{N}_m-\{i\}} P(w_i{+}X = y)\cdot f(y,w_j).$$

Wir variieren jetzt die Folge w. Sei

$$p' := \min\{p_w :\ w = w_1,\ldots,w_m \text{ ist Folge von Wörtern aus } \mathbb{Z}_2^n\}.$$

Seien $W_1,\ldots,W_m$ unabhängige Zufallsvariablen mit der durch $P(W_i = w) = 2^{-n}$ für jedes $w\in\mathbb{Z}_2^n$ beschriebenen Gleichverteilung und $W = W_1,\ldots,W_m$. Wir schätzen die erwartete Fehlerwahrscheinlichkeit und damit p' durch ε nach oben ab.

$$\begin{aligned} p' &\leq E(p_W) \\ &\leq \frac{\varepsilon}{2} + \frac{1}{m}\sum_{i=1}^{m}\sum_{y\in\mathbb{Z}_2^n}\sum_{j\in\mathbb{N}_m-\{i\}} E(P(W_i + X = y|W))\underbrace{E(f(y,W_j))}_{=|B_d(y)|/2^n} \\ &= \frac{\varepsilon}{2} + \frac{1}{m}|B_d(0)|/2^n\sum_{i=1}^{m}\sum_{j\in\mathbb{N}_m-\{i\}}\underbrace{\sum_{y\in\mathbb{Z}_2^n} E(P(W_i + X = y|W))}_{=1} \\ &= \frac{\varepsilon}{2} + (m-1)|B_d(0)|/2^n \\ &\overset{(2)}{\leq} \frac{\varepsilon}{2} + m\cdot\frac{n}{2}\cdot\frac{n^n}{d^d(n-d)^{n-d}2^n}. \end{aligned}$$

Nehmen wir $p' > \varepsilon/2$ an. Subtraktion von $\varepsilon/2$, Logarithmieren und Division durch n liefert

$$\begin{aligned} \frac{\log(p'-\varepsilon/2)}{n} &\leq \frac{1}{n}\log(m\cdot\frac{n}{2}\cdot\frac{n^n}{d^d(n-d)^{n-d}2^n}) \\ &= \frac{\log m}{n} + \frac{\log(n/2)}{n} - \frac{1}{n}\log(2^n\cdot\frac{d^d}{n^d}\cdot\frac{(n-d)^{n-d}}{n^{n-d}}) \\ &= \frac{\log m}{n} + O(n^{-1/2}) - (1 + \frac{d}{n}\log\frac{d}{n} + \frac{n-d}{n}\log\frac{n-d}{n}). \end{aligned}$$

Unter Verwendung von

$$\begin{aligned}\frac{d}{n}\log\frac{d}{n} &= \frac{\lfloor np+k\cdot\sigma\rfloor}{n}\log\frac{\lfloor np+k\cdot\sigma\rfloor}{n}\\ &= p\log p + O(n^{-1/2}) \text{ und, analog,}\\ \frac{n-d}{n}\log\frac{n-d}{n} &= (1-p)\log(1-p)+O(n^{-1/2})\end{aligned}$$

erhalten wir mit $m := 2^{\lfloor rn\rfloor}$ und $\delta := r-(1+p\log p+(1-p)\log(1-p)) > 0$

$$\begin{aligned}\frac{\log(p'-\varepsilon/2)}{n} &\leq \frac{\lfloor rn\rfloor}{n} - r - \delta + O(n^{-1/2})\\ &\leq -\delta + O(n^{-1/2});\end{aligned}$$

wegen $\frac{\lfloor rn\rfloor}{n} \leq r$ ist für genügend große n In der Stochastik zeigt man:

die rechte Seite kleiner als $-\delta/2$, und wir erhalten

$$p' \leq \varepsilon/2 + 2^{-n\delta/2}$$

für diese n. Für genügend großes n ist also $p' \leq \varepsilon$.

Also gibt es eine Folge w mit $p_w \leq \varepsilon$; aus w konstruieren wir abschließend einen entsprechenden m-elementigen Code:

Zu $i \in \mathbb{N}_m$ wählen wir $j(i) := \min\{j \in \mathbb{N}_m : w_i = w_j\}$ (der erste Index j, an der w_i vorkommt) und setzen $J := \{j(i) : i \in \mathbb{N}_m\}$. Wir können die Menge $D := \{w_1, \ldots, w_m\}$ zu einer m-elementigen Teilmenge C von $\mathbb{Z}_2^n$ ergänzen. Es gilt dann $|C \setminus D| = |\mathbb{N}_m \setminus J|$, und wir können C als Folge $v = v_1, \ldots, v_m$ mit $v_i = w_i$ für $i \in J$ darstellen. Für $i \in J$ gilt dann trivialerweise $P(m_v(v_i + X) \neq v_i) = P(m_w(w_i + X) \neq w_i)$, für $i \in \mathbb{N}_m \setminus J$ gilt sogar $m_w(w_i + x) = ?$ für alle $x \in \mathbb{Z}_2^n$ und folglich $P(m_v(v_i + X) \neq v_i) \leq 1 = P(m_w(w_i + X) \neq w_i)$.

Also ist $p_C = p_v \leq p_w \leq \varepsilon$, und folglich ist C ein n-Code mit den gewünschten Eigenschaften. □