1 Error correcting code (Hamming Code)

Übertagung von Daten und Datenverlust und wie man diese minimiert.

- Datenveränderung bei Datenübertragung oder -speicherung
- Paritätsbits und überprüfung von Parität(Gleichheit)
- Dreimal schreiben $abc \to aaabbbccc$

Definition 1. *Der Hamming Code funktioniert wie folgend: Statt der Folge $v = abcd$ schickt man die Folge $w = abcdefg$, wobei $e \coloneqq a+b+c$, $f \coloneqq a+b+d$ und $g \coloneqq a+c+d$.*

Definition 2. *Alphabet S ist ein endliche Menge. Bsp. $S = \{0,1\}$ oder $S = \{a,b,c\dots,z\}$, wir sehen diese als $\mathbb{Z}/n\mathbb{Z}$ an. Elemente aus S nennen wir Buchstaben oder auch Bits wenn $n = 2$. Ein Wort ist eine beliebig lange endliche Folge an Buchstaben. Für die Menge aller Wörter der Länge n Folge wir $S^n = \{w = a_1a_2\dots a_n | a_1, a_2\dots, a_n \in S\}$. Ist ein Vektorraum mit std Vektorraumaddition und Skalarmultiplikation.*

Definition 3. *Ein Code über S der Länge n ist ein Teilmenge $C \subseteq S^n$*

Definition 4. *Wir Definieren Folgende Terminologie:*

- *Die Hamming Distanz von 2 Wörtern $u, v \in S^n$ ist*

$$d(u,v) := \#\{i | u_i \neq v_i, i = 1,2,\dots,n\} \tag{1}$$

  Das Bedeutet, dass man v erhält wenn man d(u,v) "Fehler" macht.

- *Ein Code Korrigiert t Fehler wenn für jedes $u \in S^n$ existiert maximal ein $v \in C$ mit $d(u,v) \leq t$.*
- *Die minimum Distanz von einem Code ist definiert als $d(C) \coloneqq min\{d(u,v)|u,v \in C, u \neq v\}$*

Theorem 5. *Ein Code C korrigiert t Fehler genau dann wenn $d(C) \geq 2t+1$.*

Beweis mit Kugeln um w

Codieren und Decodieren 6. *Um diesen Code auch wirklich nutzen zu können (es sei denn dass du Einsen und Nullen verschicken willst) müssen wir eine injektive Abbildung $c : \Sigma^k \to C$ finden. Wobei Σ ein Alphabet ist und k die Länge der Wörter ist.*

Um eine gegeben Nachricht $v \in \Sigma^k$, erstellen wir das Wort $w = c(v) \in C$ und Senden dieses. Nun erhalten wir das Wort $w' \in S^n$, mit diesem suchen wir ein Wort mit $w'' \in C$, so dass $d(w', w'')$ möglichst klein ist. Nun berechnen wir $v' = c^{-1}(w'') \in \Sigma^k$ Wenn maximal t Fehler entstanden sind und D t Fehler korrigiert, gilt: $w'' = w$ und $v' = v$.

Definition 7 (Liniarer Code)**.** *Linearer Code ist eine besondere Art von Code und der Hamming Code ist einer davon. In diesem Fall gilt, das alphabet S ist ein endliches Feld (Am Wichtigsten $S = \mathbb{F}_2$), dadurch ist S^n ein Vektorraum. Jede Untervektorraum von S^n ist ein Linearer Code.*

Beweis ist Trivial

Observation 8. *Für alle Linearen Codes gilt:*

$$d(C) = min\{d(0,w) | w \in C, w \neq 0\} \tag{2}$$

Da wir in der Linearen Algebra arbeiten müssen wir C nicht als Wortliste Geben sonder es gibt einfachere Möglichkeiten, welche uns auch beim Codieren und Decodieren Helfen

Theorem 9 (Eine Basis)**.** *Wir Können eine Generations Matrix G bauen mit welcher wir C bauen. Die Generationsmatrix ist eine $n \times k$ Matrix mit $n := dim(C)$, wobei die Reihen eine Basis von C sind.*

Die Generations Matrix ist sehr gut zum Codieren da wenn wir den Vektor v haben, senden wir den Vektor $w := G^T v \in C$ Man kann eine Generationsmatrix erhalten in der Form $G = (Id_n|A)$, wenn man eine geeignete Basis wählt.

Observation 10. *Der Vektor w stimmt mit v in den ersten k Stellen überein. Das bedeutet das $k-n$ extra Symbole angehängt werden. Diese werden Paritätsbits genannt.*

Beispiel 11. *Der Hamming Code hat die Generationsmatrix:*

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{pmatrix}$$

Theorem 12 (Ein lineares Gleichungssystem)**.** *Ebenfalls kann man ein UVR mit einem Linearen Gleichungssystem definieren:*

$$C = \{w | w \in S^n, Pw = 0\} \tag{3}$$

Wobei P die Paritätsüberprüfungsmatrix vom Code C ist.

Wen vom Code C die Generationsmatrix $G = (Id_n|A)$ ist, ist es einfach zu überprüfen das $P := (-A^T|Id_{k-n})$ eine Paritätsüberprüfungsmatrix von C ist.

Beispiel 13. *Eine Paritätsüberprüfungsmatrix vom Hemming Code ist:*

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

Satz 14 (Generalisierter Hammingcode)**.** *Der Generalisierte Hamming Code kann folgenderweise Bewiesen werden:*

Indem man eine Paritätsüberprüfungsmatrix baut mit folgender Eigenschaft:

- *Einer Höhe l welche der Anzahl an Paritätsbits ist*
- *Die Spalten sind alle möglichen unterschiedlichen Vektoren (aus 0 und 1), mit der Ausnahme des 0-Vektors.*

Satz 15. *Der generalisierte Hemming Code hat $d(C) \geq 3$.*

Proof. Da der Generalisierte Hemming Code Linear ist reicht es zu zeigen das es kein element in C gibt was genau 1 oder 2 nicht 0 Elemente hat.

Wäre ein element mit 1 nicht 0 element in C müsste P eine Triviale Spalte haben, was nicht der fall ist. Und wenn es ein Element mit 2 nicht 0 Elementen hätten müsste es 2 identische spalten geben. □

Theorem 16 (Decodieren)**.** *Wenn wir ein Wort w aus dem Hammingcode verschicken, kommt ein wort w' an. Wenn wir davon ausgehen das Maximal ein Fehler passiert gilt $w' = w$ oder $w' = w + e_i$, für $i \in \{1, 2, \ldots, n\}$.*

Wenn $w' = w$ dann gilt $Pw' = 0$, während für $w' = w + e_i$ gilt $Pw' = Pw + Pe_1 = Pe_1$ was ist equivalent zu der i-ten Spalte. Wenn maximal ein Fehler passiert ist, können wir sofort sehen wo dieser Fehler passiert ist.

Theorem 17. *Man kann auch den Hemmingcode auch in einer Tabelle nutzen, hier gibt es einige Eigenschaften die man nutzen kann.*

- *Man nummeriert die Bits von 1 beginnend.*
- *Statt am ende bringt man die Paritätsbits mittendrin unter, und zwar jeweils an Stelle* 2^{n-1}.
- *Wenn man die Tabelle erweitert (um mehr Bits zu haben) kann man sehr schnell sehen wann ein neuer Paritätsbit gebraucht wird*
- *Wenn ein Fehler passiert kann man einfach die Position der Falschen Paritätsbits addieren um die Position des fehlerhaften Bits zu erhalten.*

Beispiel für 0101:

Position		*Bits des Codewortes*
1_{10}	001_2	*0*
2_{10}	010_2	*1*
3_{10}	011_2	*0*
4_{10}	100_2	*0*
5_{10}	101_2	*1*
6_{10}	110_2	*0*
7_{10}	111_2	*1*