Error Correction

Anubhab Das

October 2022

Coding Theory is the study of codes and I am specifically tackling error correction. Today generally I will be going over the fundamental model of Coding Theory and a broad overview of error-correction. Explain the general model of channel coding.
Basic definitions: Let $A = \{a_1, a_2 ... a_q\}$ be a set of size q, which we refer to as a code alphabet and whose elements are called code symbols.

1. A q-ary word of length n over A is a sequence $w = w_1 w_2 ... w_n$ with each $w_i \in A$ for all i. Equivalently, w may also be regarded as the vector $(w_1, ..., w_n)$.
2. A q-ary block code of length n over A is a nonempty set C of q -ary words having the same length n.
3. An element of C is called a codeword in C.
4. The number of codewords in C, denoted by $|C|$, is called the size of C.
5. The (information) rate of a code C of length n is defined to be $(\log_q |C|)/n$
6. A code of length n and size M is called an (n, M)

Begin stats:

- There is a list of forward channel probabilities that can be expressed with $\sum_{j=1}^{q} P(a_j \textbf{ received } | \; a_i \textbf{ sent})$ for all a_i.
- memoryless = independent and symmetric means errors are equally likely and $< \frac{1}{2}$ probability.
- With p = 0.05 in the code $C = \{000, 111\}$ we would have (write some eq examples)

Maximum likelihood decoding is essentially where we use the probabilities and determine which codeword seems most likely.
Hamming Distance for the words $x = (x_1, x_2 ... x_n)$ and $x = (y_1, y_2 ... y_n)$ can be easily calculated using the formula:

$$d(x, y) = |\{i \mid x_i y_i\}|.$$

Exercise: Prove Triangle inequality $d(x, z) \leq d(x, y) + d(x, z)$.

- Nearest neighbor/minimum distance decoding can be explained pretty simply and follows in a similar vein to maximum likelihood decoding. It will decode x to c_x if $(x, c_X) = min_{c \in C}(x, c)$.
- A code C is u-error-detecting if $min_{c \in C}(x, c) \geq u + 1$ and is v-error-correcting if $min_{c \in C}(x, c) \geq 2v + 1$.