Shannon's Theorem; block codes

George McNinch

2024-02-22

Overview

So far, we have chosen to use the term *code* for a vector subspace C of $\mathbb{F}_q^n$. The idea is that we are interested in encoding some data by identifying it with vectors in $\mathbb{F}_q^k$.

If G is a *generator matrix* for our code in *standard form*, then we *encode* our data: given a vector $v \in \mathbb{F}_q^k$, the encoded version is

$$v \cdot G \in \mathbb{F}_q^n.$$

Note that – since G is in standard form – if $v = (v_1, \cdots, v_k)$ then

$$v \cdot G = (v_1, \cdots, v_k, w_{k+1}, \cdots, w_n)$$

for some scalars $w_j \in \mathbb{F}_q$.

Our intent is to "*transmit*" this encoded data $v \cdot G$, possibly through some noisy channels that may result in transmission errors. At the other end, some vector w in $\mathbb{F}_q^n$ is received, and the hope is to recover the vector v from the received vector w.

The field of *Information Theory* formulates a way to reason about such systems; we are going to sketch a rudimentary description.

Noisy Channels and Shannon's Theorem

We consider finite sets S and T (say $S = \{s_1, \cdots, s_n\}$ and $T = \{t_1, \cdots, t_m\}$).

We consider *random variables* X on S and Y on T. In particular, we may consider probabilities:

$$P(X = s) = p_s, \quad \text{the probability that the random var } X \text{ takes value } s \in S$$

$$P(Y = t) = q_t, \quad \text{the probability that the random var } Y \text{ takes value } t \in T$$

We view the elements of the set S as the data we *send* through some transmission channel, and T as the data we *receive*.

In the case of our linear codes $C \subset \mathbb{F}_q^n$ described above, S would be $\mathbb{F}_q^k$ and T would be $\mathbb{F}_q^n$.

Block codes

More generally, we consider *block codes* $C \subset A^n$. Here A is an *alphabet*, and the code words in C are just n-tuples of elements from A. We write $q = |A|$. We say that the *length* of the code C is n.

In this setting, for our transmission channel we take $S = C$ and $T = A^n$.

Channel

A channel Γ for transmission amounts to the following matrix with rows indexed by the set S and columns indexed by the set T:

$$p_{st} = P(Y = t \mid X = s)$$

i.e. the conditional probability that $Y = t$ given that $X = s$.

Example An example is the *binary symmetric channel*, where $S = T = \{0, 1\}$ and the channel matrix Γ is given by

$$(p_{st}) = \begin{bmatrix} \phi & 1-\phi \\ 1-\phi & \phi \end{bmatrix} \quad \text{for some } \phi \in [0, 1].$$

> Here ϕ represents the probability that 0 was received given that 0 was sent, and also the probability that 1 was received given that 1 was sent.

Note for example that if we know the channel matrix and if we know the probabilities for the random variable X, we find the probabilities for Y via

$$P(Y = t) = \sum_{s \in S} P(Y = t \mid X = s) \cdot P(X = s).$$

Example For the binary symmetric channel if we know that $P(X = 0) = p$, then

$$P(Y = 0) = \phi \cdot p + (1 - \phi)(1 - p).$$

Decoding

With notation as above, a *decoding* is a function $\Delta : T \to S$. Think of it this way: given that $t \in T$ was received, the decoding function "guesses" that $\Delta(t) \in S$ was sent.

The question is: how to identify a good decoding function.

Well, we can consider the probabilities that reflect how often a decoding Δ is correct:

$$q_{\Delta(t),t} = P(X = \Delta(t) \mid Y = t)$$

represents the probability that $\Delta(t)$ was sent given that t was received.

The *average probability of a correct decoding* is given by

$$P_{\text{COR}} = \sum_{t \in T} q_t \cdot q_{\Delta(t),t}$$

(remember that $q_t = P(Y = t)$)

Maximum likelihood decoding

A decoding $\Delta : T \to S$ is said to be a *maximal likelihood decoding* if

$$p_{\Delta(t),t} \geq p_{s,t}$$

for every $s \in S$ and every $t \in T$.

Under some circumstances, one achieves maximal likelihood decoding through *minimum distance decoding*.

For example, we have:

Lemma For the binary symmetric channel and $C \subset \mathbb{F}_2^n$, consider the assignment

$$\Delta(v) = u$$

> where u is the closest neighbor to v in C with respect to the Hamming distance.
>
> Then Δ is *maximal likelihood decoding*.

Transmission rate and capacity

For a block code $C \subset A^n$ with $|A| = q$, the *transmission rate* of C is defined to be

$$R = \frac{\log_q(|C|)}{n}$$

If $A = \mathbb{F}_q$ and C is a linear code with $k = \dim_{\mathbb{F}_q} C$, then

$$R = k/n.$$

There is a notion of the *capacity* of a channel Γ; it is a number which in some sense encodes the theoretical maximum rate at which information can be reliably transmitted over the channel; we omit the definition here.

Shannon's Theorem

Theorem (Shannon) Let $\delta > 0$ be given and let $0 < R$ with R less than the channel capacity.

For every sufficiently large natural number n, there is a code of length n and transmission rate $\geq R$ such that, using maximum likelihood decoding, the probability P_{COR} of a correct decoding satisfies

$$P_{\text{COR}} > 1 - \delta.$$

This shows that, given a channel with non-zero capacity, there are codes which allow us to communicate using the channel and decode with a probability of a correct decoding arbitrary close to 1.

It is not constructive, of course – in some sense, this motivates the subject: how to find the codes that work well?

Bounds for block codes

Here we consider an *alphabet* A – thus A is just a finite set, of order $|A| = q$ – and a set of codewords $C \subset A^n$; we call n the *length* of the codewords.

We consider the Hamming distance on A^n: for $u, v \in A^n$,

$$\text{dist}(u, v) = \#\{i \mid u_i \neq v_i\}$$

where e.g. $u = (u_1, \cdots, u_n) \in A^n$. In words, the distance between u and v is the number of coordinates in which the tuples differ.

It is straightforward to check that dist is a *metric* on the finite set A^n; in particular, the *triangle inequality* holds: for every $u, v, w \in A^n$ we have

$$\text{dist}(u, v) \leq \text{dist}(u, w) + \text{dist}(w, v).$$

The *minimal distance* of C is given by

$$d = \min\{\text{dist}(u, v) \mid u, v \in C, u \neq v\}.$$

Lemma Using nearest neighbor decoding, a block code of minimal distance d can correct up to $(d-1)/2$ errors.

Proof For every $w \in A^n$ and every $u, v \in C$ we have

$$(*) \quad d \leq \text{dist}(u, v) \leq \text{dist}(u, w) + \text{dist}(w, v).$$

Now, if $\text{dist}(u, w) \leq (d-1)/2$ and $\text{dist}(w, v) \leq (d-1)/2$ then $\text{dist}(u, v) \leq d - 1$, contrary to $(*)$. Thus for any w there is at most one codeword $u \in C$ for which $\text{dist}(u, w) \leq (d-1)/2$.

From the point of view of code transmission, if $w \in A^n$ is *received* and no more than $(d-1)/2$ of the components of $w = (w_1, w_2, \cdots, w_n)$ are erroneous, then nearest neighbor decoding will find the codeword in C that was transmitted.

Example (*Repetition code*) Consider a finite alphabet A and the the codewords $C = \{(a, a, \cdots, a) \in A^r \mid a \in A\}$. Thus the data $a \in A$ is encoded by the redundant codeword $(a, a, \cdots, a)$.

The minimal distance between distinct codewords in C is r, so the Lemma shows that using nearest neighbor decoding, this code can correct up to $(r-1)/2$ errors.

(Note that in this case nearest neighbor decoding amounts to decoding

$$(a_1, a_2, \cdots, a_r)$$

by "majority vote"; in other words, view $\{a_1, a_2, \cdots, a_r\}$ as a *multi-set* and choose an element with maximal multiplicity.)

Counting codes with given parameters

Write $A_q(n, d)$ for the maximum size $|C|$ of a block code $C \subset A^n$ having minimal distance d.

We are going to prove some results about the quantity $A_q(n, d)$.

We first compute the size of a "ball" in the metric space (A^n, dist).

Lemma For $u \in A^n$ and a natural number m write:

$$B_m(u) = \{v \in A^n \mid \text{dist}(u, v) \leq m\}.$$

Let

$$\delta(m) = 1 + \binom{n}{1}(q-1) + \binom{n}{2}(q-1)^2 + \cdots + \binom{n}{m}(q-1)^m = \sum_{j=0}^{m} \binom{n}{j}(q-1)^j.$$

Then $|B_m(u)| = \delta(m)$.

Remark Note that if $k \in \mathbb{N}$, $k > n$ then we insist that $\binom{n}{k} = 0$; in this case "there are 0 ways of choosing precisely k elements from a set of size n."

This is consistent e.g. with the formula

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!}$$

since the factor $(n-n) = 0$ appears in the numerator.

Proof of Lemma For each $j = 0, 1, \cdots, m$ there are $\binom{n}{j} \cdot (q-1)^j$ elements of A^n at distance precisely j from u.

We may now state and prove the following:

Theorem (*Gilbert-Varshamov Bound*)

$$A_q(n, d) \cdot \delta(d-1) \geq q^n.$$

Proof Suppose that $C \subset A^n$ is a code with minimal distance d for which $|C| = A_q(n, d)$.

Notice that $|C| \cdot \delta(d-1)$ is the size of the disjoint union

$$\bigsqcup_{u \in C} B_{d-1}(u).$$

.

Thus, if

$$|C| \cdot \delta(d-1) < q^n = |A^n|$$

then

$$\bigcup_{u \in C} B_{d-1}(u) \subsetneq A^n;$$

thus, there is some element $v \in A^n$ for which

$$v \notin \bigcup_{u \in C} B_{d-1}(u).$$

We then have dist$(u, v) \geq d$ for every $u \in C$.

This shows that $C \cup \{v\} \subset A^n$ is a code having minimal distance d, contradicting the assumption that $|C| = A_q(n, d)$; i.e. contradicting the assumption that C has maximal size among codes $C' \subset A^n$ with minimal distance d.

Bibliography