DS Lecture 28

Abhinav S Menon

Group Codes

A *code*, in coding theory, is simply the set of "codewords" that can be transmitted. Here, we will only consider codes whose codewords are n-tuples of binary values.

Any transmitted message, in real life, may be subject to errors (bit flips, in this case). When this happens, we would like the receiver to be able to identify the original message despite these errors. If the received word is outside the code, then we will consider the transmitted word to be that word in the code which is "closest" to the received message – this is called the *minimum-distance decoding criterion*. To this end, we will define a notion of distance and study its properties.

When the words in a binary code form a group under the XOR operation ($\oplus$), we call the code a *group code*. We will limit ourselves to the study of such codes.

Hamming Distance

Let C be a code, and let $x = \langle x_1, x_2, ..., x_n \rangle$ and $y = \langle y_1, y_2, ..., y_n \rangle$ be two codewords in it. We define the *Hamming distance* $H(x, y)$ as the number of bits in which x and y differ, *i.e.*, the number of errors needed to change one to the other. For example, $H(\langle 1, 0, 1, 0, 1 \rangle, \langle 0, 1, 1, 1, 0 \rangle) = 4$ (they are same only in the third bit).

In a group code C, the Hamming distance has the following properties:

1. $H(x, y) \geq 0$, $\forall x, y \in C$
2. $H(x, y) = 0 \Leftrightarrow x = y$, $\forall x, y \in C$
3. $H(x, y) = H(y, x)$, $\forall x, y \in C$ (commutative)
4. $H(x, z) \leq H(x, y) + H(y, z)$, $\forall x, y, z \in C$ (triangle inequality)

Further, for C, we define the *minimum Hamming distance* as $d_C = \min_{x,y \in C} H(x, y)$. This property of a group code is central to determining its error detection and error correction capacities.

Error Detection

First, let us consider how we might detect one or more errors in a message taken from a certain group code C. Clearly, if a received message is outside the group code, we can say that it has been subjected to at least one error (how many does not matter for now).

Now, let us suppose that we wish a certain group code to be able to detect (not necessarily correct) up to e errors. What does this mean? We would like that if a transmitted message undergoes e errors, we can say with absolute certainty that the received message is not a member of the group code anymore.

We note that the distance between the transmitted and received messages is at most e. If $d_C = e$, then consider the pair $c_1, c_2 \in C$ such that $H(c_1, c_2) = d_C$. If c_1 is transmitted, then it is entirely possible that e errors could change it to c_2, and we will be unable to detect any error. Similarly, if $d_C \leq e$, we cannot detect e errors.

Therefore, in order to be able to detect e errors, we see that a group code C must have $d_C > e$.

We also note that if C has $d_C > e$, it will be able to detect e errors or fewer, since that number of errors is never sufficient to change one member of the code to another.

We conclude that a code C can detect up to e errors if and only if $d_C > e$.

Error Correction

If we wish e errors to be corrected, then clearly the transmitted message must be outside the code (we cannot correct them without detecting them) and it should be closest to the received message (by the minimum-distance decoding criterion).

Let us say that we wish a group code C to be able to correct up to e errors. Then, as proved above $d_C > e$; but this is not a sufficient condition. Suppose $d_C = 2e$. Now, consider the pair $c_1, c_2 \in C$ such that $H(c_1, c_2) = 2e$. Since c_2 can be obtained from c_1 by flipping $2e$ bits, clearly it is possible to find some codeword c which is obtained from c_1 by flipping exactly e of these bits. Then, clearly $H(c_1, c) = e$. Also, e more bit-flips are needed to get c_2 from c, so we know that $H(c, c_2) = e$ also.

Hence, if c_1 were to be transmitted, e errors could change it into c, which (although outside C, as required) is equidistant from c_1 and c_2, meaning we cannot decode it.

Therefore, in order to be able to correct e errors, we see that a group code C must have $d_C > 2e$.

We also observe that if C has $d_C > 2e$, then e errors will not make any codeword more similar to a different one than the original, so it can correct e errors.

We conclude that a code C can correct up to e errors if and only if $d_C > 2e$.

Weight

The *weight* of a codeword x, denoted by $w(x)$, is simply the number of its bits having the value 1. From this, we can see that $w(x) = H(x, 0)$ (number of bit-flips from all zeroes = number of ones).

Further, consider $H(x, y)$ and $x \oplus y$. In $x \oplus y$, all those positions have the value 1 where x and y differ ($0 \oplus 1 = 1 \oplus 0 = 1$). But there are exactly $H(x, y)$ such positions, so clearly $H(x, y) = w(x \oplus y)$.

Consider the minimum distance d_C of a code C, and $c_1, c_2 \in C$ such that $H(c_1, c_2) = d_C$. We know that $w(c_1 \oplus c_2) = d_C$. If there existed $x \in C$ such that $w(x) = d < d_C$, then $H(x, 0) < d_C$, which is a contradiction; therefore d_C is also the minimum weight of all the codewords of C.

This can be proved in the other direction as well: consider the minimum weight w_C of a code C, and $c \in C$ such that $w(c) = w_C$. If there existed $x, y \in C$ such that $H(x, y) = w < w_C$, then $w(x \oplus y) < w_C$, which is a contradiction because $x \oplus y \in C$. Therefore w_C is also the minimum distance of C.

Null Spaces and Parity-Check Matrices

Every group code C has an associated matrix H of binary values called a *parity-check matrix*. However, obtaining H from C has not been covered; it can be done using field theory.

Obtaining C from H, on the other hand, is fairly straightforward. If H has dimensions $r \times n$, then C is defined as

$$C = \{x | x \cdot H^T = \mathbf{0}_r\}.$$

Here, x is a row matrix of dimensions $1 \times n$, and $\mathbf{0}_r$ is the row matrix consisting of r zeroes. If H is C's parity-check matrix, then C is called the *null space* of H and sometimes denoted by $N(H)$. It can be proved that for any H, $N(H)$ forms a group under $\oplus$:

Closure. $x, y \in N(H) \Rightarrow x \cdot H^T = \mathbf{0}_r, y \cdot H^T = \mathbf{0}_r \Rightarrow x \cdot H^T \oplus y \cdot H^T = \mathbf{0}_r \Rightarrow (x \oplus y) \cdot H^T = \mathbf{0}_r \Rightarrow (x \oplus y) \in N(H)$.

Associativity. We know that $\oplus$ is associative.

Existence of Identity. $\mathbf{0}_n$ is the identity in $N(H)$, since $x \oplus \mathbf{0}_n = \mathbf{0}_n \oplus x = x$.

Existence of Inverse. Since $x \oplus x = \mathbf{0}_n$, $\forall x \in C$, each x is its own inverse.

We note that $N(H)$ is an abelian group.

Also, H can always be expressed in the form $[P \| I_r]$, where I_r is the $r \times r$ identity matrix and P is an arbitrary $r \times (n-r)$ matrix. This is called the *canonical* or *standard* form.

Theorem. For an $r \times n$ matrix H, $N(H)$ has minimum weight d if and only if d is the minimum number of columns that sum to $\mathbf{0}_r$.

Proof. We will prove this in both directions.

(i) $N(H)$ has minimum weight $d \Rightarrow d$ is the minimum number of columns that sum to $\mathbf{0}_r$.

Assuming that $N(H)$ has minimum weight d and that there exist $s < d$ columns (call them $h_{i_1}, h_{i_2}, ..., h_{i_s}$) of H that sum to $\mathbf{0}_r$, we will derive a contradiction.

Construct a codeword c such that $c_{i_1} = c_{i_2} = \cdots = c_{i_s} = 1$, and all the remaining components of c are 0. Then, consider $(H \cdot c^T)$. [c is a row matrix; c^T is an $n \times 1$ column matrix].

In the first element of the product (an $r \times 1$ column matrix), the first element of each column h_i is multiplied by the corresponding component c_i of c. But all components except $c_{i_1}, c_{i_2}, ..., c_{i_s}$ are 0; hence this sum is simply the sum of the first elements of $h_{i_1}, h_{i_2}, ..., h_{i_s}$. This we know to be 0; hence the first element of the product is 0. Similarly, all other elements can be shown to be 0, and therefore $H \cdot c^T$ is simply the column vector consisting of r zeroes. Taking the transpose on both sides, $c \cdot H^T = \mathbf{0}_r$, and so $c \in N(H)$.

But by construction, $w(c) = s < d$, which is a contradiction. Therefore, d is the minimum number of columns of H that can sum to $\mathbf{0}_r$.

(ii) d is the minimum number of columns that sum to $\mathbf{0}_r \Rightarrow N(H)$ has minimum weight d.

Again, assuming that d is the minimum number of columns that sum to $\mathbf{0}_r$ and that some codeword $c \in N(H)$ has weight $w(c) = s < d$, we will derive a contradiction.

We know that $c \cdot H^T = \mathbf{0}_r$, so $H \cdot c^T$ is the column matrix consisting of r zeroes. Suppose c has 1 in positions $i_1, i_2, ..., i_s$. Then the first elements of the columns $h_{i_1}, h_{i_2}, ...h_{i_s}$ sum to 0. Similarly, each of the other sets of s elements sums to 0.

Therefore the s columns $h_{i_1}, h_{i_2}, ...h_{i_s}$ sum to 0, which is a contradiction. Hence d is the minimum weight of $N(H)$. This completes the proof.

$$H \cdot c^T = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

From this theorem, we see that if a group code C has a minimum weight 3 (which is required to correct one error or detect 2), its parity-check matrix H must satisfy the following conditions:

1. no column of H is all zeroes, and no two columns of H are identical [so that no two columns sum to 0]
2. there exist three columns $h_{i_1}, h_{i_2}, h_{i_3}$ in H whose sum is 0 [so 3 is the minimum number of columns that sum to 0].