Teoría de Códigos y Criptografía

Justo Peralta López
Juan Antonio López Ramos

Universidad de Almería
Departamento de Álgebra y Análisis Matemático

1. Código Bloque y Distancia Mínima
2. Relación entre la mínima distancia y la probabilidad de error
3. Algunas clasificaciones de códigos
   - Códigos sistemáticos
   - Códigos equivalentes
   - Códigos de repetición
   - Códigos lineales y cíclicos
4. Obtención de nuevos códigos a partir de otros ya existentes
   - Por extensión
   - Por punción
   - Por borrado y aumentación
   - Por recorte,suma directa y $(u, u+v)$ construcción

Definición

Si A es un alfabeto, una palabra de longitud n es una secuencia de símbolos de dicho alfabeto. Al conjunto de todas las palabras de longitud n sobre ese alfabeto lo denotamos por A^n.

Si A está formado por q símbolos, entonces podemos escoger entre q posibilidades en cada posición de la palabra a formar. Por lo tanto, el número total de palabras de longitud n será q^n.

Definición

Un (n, M)-código bloque C sobre un alfabeto A, es un subconjunto C de A^n, con $|C| = M$.

Normalmente, hablaremos de n como la longitud de C y de M como el tamaño o el número de palabras código de C.

Definición

El peso Hamming de una palabra del código, $x = (x_1, x_2, \ldots, x_n)$, al cual denotaremos por $w(x)$,es el número de componentes distintos de cero.

Definición

La distancia Hamming entre dos palabras del código, x e y, a la cual denotaremos por $d(x, y)$, es el número de posiciones en que dichas palabras difieren. Es decir

$$d(x, y) = w(x - y) = w(y - x)$$

Ejemplo

1. Sea $x = (10011)$ y $y = (01010)$ sobre $GF(2)$. Entonces $w(x) = 3$, $w(y) = 2$ y $d(x, y) = 3$.
2. Sea $x = (20120)$ y $y = (10221)$ sobre $GF(3)$. Entonces $w(x) = 3$, $w(y) = 4$ y $d(x, y) = 3 = w(x - y) = w(10202)$

Lema

La distancia Hamming es una métrica, es decir, verifica la siguientes propiedades: Para cualquier x, y, z

1. *$d(x, y)$ es un número real no negativo.*
2. *Si $d(x, y) = 0$, si y solo si $x = y$.*
3. *d(x,y)=d(y,x).*
4. *$d(x, y) \leq d(x, z) + d(z, y)$.*

Definición

La distancia mínima, d_{min}, de un código C es la mínima de las distancias entre todos los pares de las palabras del código.

A partir de ahora, un (n, M, d)-código representa un (n, M)-código bloque con mínima distancia d.

Teorema

Es necesario y suficiente que la distancia mínima de un código sea mayor o igual que d, para poder detectar $d - 1$ errores o menos.

Teorema

Un código C puede corregir t errores o menos si y solo si $d_{min} \geq 2t + 1$.

Teorema

Un código C puede corregir cualquier combinación de t errores y detectar d con $d \geq t$, si y solo si $d_{min} \geq t + d + 1$

Ejemplo

Sea $C = \{00, 11\}$. La distancia mínima es 2. Como se puede observar, si recibimos la palabra 01 o 10, podemos detectar el error, pero no corregirlo. (Ver figura 1).

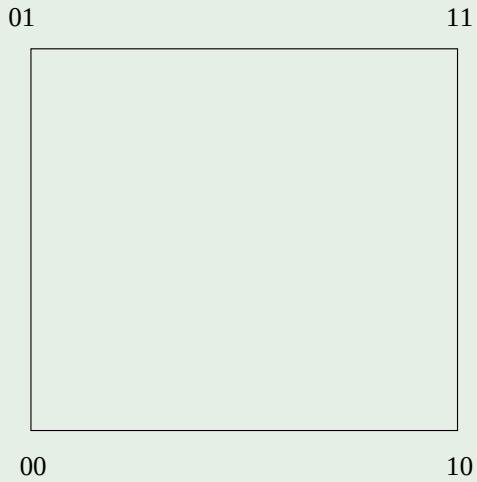

Figura: Ejemplo $C = \{00, 11\}$

Ejemplo

$C = \{000, 111\}$ Si un error ocurre, podemos corregir por aquella palabra del código más cercana. ($000 \to 010 \to 000$). Sin embargo, si dos errores ocurren ($000 \to 110 \to 111$), decodificamos por la más cercana de forma incorrecta. Es decir, en este caso, somos capaces de detectar y corregir un sólo error.
La otra posibilidad es usar el código sólo para detectar, en cuyo caso una palabra no contiene ningún error si y solo si pertenece al código, luego sólo podemos detectar dos errores.

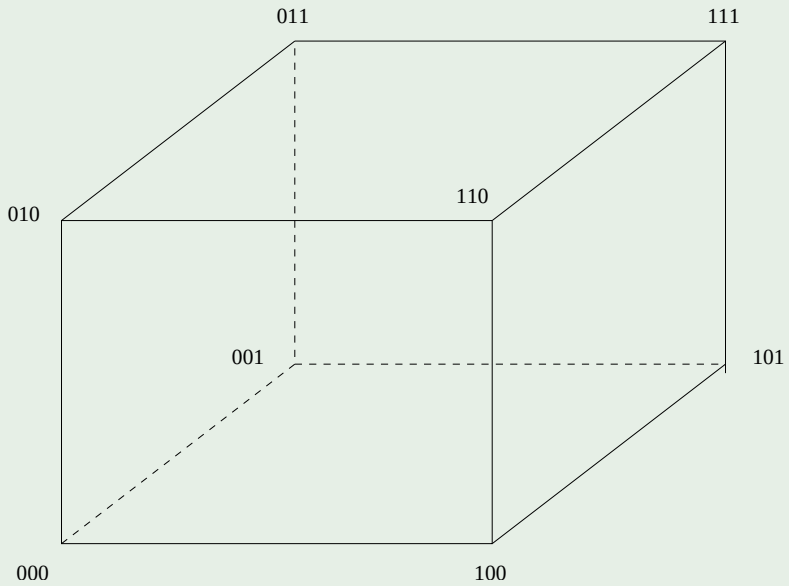

Figura: Ejemplo $C = \{000, 111\}$

Ejemplo

Sea $C = \{000000, 111111\}$ En este caso la distancia mínima es 6 y tenemos las siguientes posibilidades.

Casos	d	t
1	3	2
2	4	1
3	5	0

Ejemplo (Canal BEC)

$C = \{000, 111\}$ La decodificación viene dada por la siguiente tabla

000	111
$x00$	$x11$
$0x0$	$1x1$
$00x$	$11x$
$xx0$	$xx1$
$x0x$	$1x1$
$0xx$	$1xx$

Teorema

Si C es un código con distancia mínima d en un BEC. Entonces podemos corregir d – 1 errores (si sólo se utiliza para corregir), y detectar d errores (si sólo se utiliza para detectar).

Lema

Si utilizamos un código bloque de longitud n sobre un canal simétrico binario (BSC), donde la probabilidad de que se produzca un error es p. Entonces, la probabilidad de que se produzcan un error de peso k (o en k posiciones), es $p^k(1-p)^{n-k}$. *Mientras que la probabilidad de que se produzca algún error de peso k, será* $\binom{n}{k}p^k(1-p)^{n-k}$

Definición

Un (n, q^k)-código es llamado sistemático, si existen k posiciones de las palabras del código, $i_1, i_2, \ldots, i_n$, tal que observando esas posiciones en todas las palabras del código, obtenemos todas las q^k posibles palabras, de longitud k sobre el alfabeto de q elementos.

Ejemplo

1. Para $C = \{0000, 0110, 1001, 1010\}$, si seleccionamos la primera y la tercera posición en todas las palabras del código, obtenemos el conjunto $\{00, 01, 10, 11\}$, es decir, todas las posibles palabras de longitud 2 sobre el alfabeto binario.
2. Para el caso $C = \{000, 100, 010, 001\}$, el código no es sistemático.

Definición

Decimos que dos (n, M)-códigos, C_1 y C_2, son equivalentes, si existe una permutación σ de las coordenadas o posiciones de las palabras del código, y permutaciones $\pi_1, \pi_2, \ldots, \pi_n$ de los símbolos del alfabeto tal que $c_1 c_2 \ldots c_n \in C_1$ si y sólo si $\pi_1(c_{\sigma(1)})\pi_2(c_{\sigma(2)} \ldots \pi_n(c_{\sigma(n)}) \in C_2$

Definición

Dos (n, M)-códigos, C_1 y C_2, sobre $GF(q)$, se dicen que son equivalentes múltiplo por un escalar, si podemos obtener uno a partir del otro multiplicando los símbolos del alfabeto por un escalar.

Definición

Un r-código C se dice que es de repetición si todas las posiciones de todas las palabras códigos tienen el mismo símbolo.

Ejemplo

Sea C es siguiente q-código.

$$C = \{00\ldots0, 11\ldots1, \ldots, (q-1)(q-1)\ldots(q-1)\}$$

Obsérvese, que la razón del código es $R = 1/n$, donde n es la longitud de palabra del código.

Definición

Un r-código C se dice que es lineal, si para toda $x, y \in C$, $x + y \in C$.

Si C es un código lineal, lo notaremos por un (n, k, d)-código lineal donde n es la longitud de palabra del código, k la dimensión del código y d su distancia mínima.

Definición

Si $x = x_1 x_2 \dots X_n$ e $y = y_1 y_2 \dots y_n$ son palabras de un código binario, entonces definimos la intersección de x con y por

$$x \wedge y = (x_1 y_1, x_2 y_2, \dots, x_n y_n)$$

Lema

Para todo $x, y \in V(n, 2)$,

$$d(x, y) = w(x) + w(y) - 2w(x \wedge y)$$

Definición

Un r-código C es cíclico si es lineal y si para toda palabra $c = (c_1, c_1, \dots, c_n) \in C$, entonces $c_n, c_1, \dots, c_{n-1}) \in C$.

Definición

Si C es un (n, M, d)-código sobre el alfabeto $GF(q)$, el código extendido se obtiene como sigue

$$\hat{C} = \{c_1 c_2 \dots c_n c_{n+1} | c_1 c_2 \dots c_n \in C \text{y} \sum_{k=1}^{n+1} c_k = 0 \ mod \ q\}$$

Teorema

Si C es un (n, M, d)-código, entonces $\hat{C}$ es un $(n+1, M, d \ o \ d+1)$-código.

Ejemplo

Sea $C = \{00, 01, 10, 11\}$, entonces $\hat{C} = \{000, 011, 101, 110\}$. Nótese que C posee distancia mínima 1 y $\hat{C}$ distancia mínima 2.

Definición

Es el proceso contrario al de extensión. En este caso, una o más posiciones son borradas en todas las palabras del código.

Teorema

Si C es un (n, M, d)-código, entonces C^* es un $(n-1, M, d$ o $d-1$

Teorema

Un código $(n, M, 2t+1)$-código binario existe si y sólo si existe un $(n+1, M, 2t+2)$-código binario.

Definición

Un nuevo código por borrado consiste en eliminar algunas de las palabras del código

Definición

Un nuevo código por aumentación consiste en añadir algunas las palabras al código

Una forma es añadir las palabras complementarias del código para el caso binario. Por ejemplo, si $c = 010111$, su complementario $c^c = 101000$.

Lema

Si x, y pertenecen a $V(n, 2)$ (el espacio vectorial formado por todas las palabras binarias de longitud n), entonces $d(x, y^c) = n - d(x, y)$

Teorema

Si C es un (n, M, d)-código binario, entonces

$$d(C \cup C^c) = min\{d, n - d_{max}\}$$

donde d_{max} es la distancia máxima de las palabras de C.

Definición

Consiste en mantener sólo las palabras que poseen un determinado símbolo en una determinada posición. Al símbolo y la posición en cuestión se le llama sección de corte $(x_i = s)$.

Teorema

Si C es un (n, M, d)-código lineal binario, el código recortado con sección de corte $x_1 = 0$, es un $(n-1, 1/2M, d)$-código.

Definición

Si C_1 es un (n_1, M_1, d_1)-código y C_2 un (n_2, M_2, d_2)-código, entonces la suma directa de C_1 y C_2 viene dado por

$$C_3 = \{cd | c \in C_1, d \in C_2\}$$

donde cd es la concatenación de esas dos palabras. Y C_3 será un $(n_1 + n_2, M_1 M_2, min\{d_1, d_2\})$-código.

Definición

Es una variación de la anterior. Si C_1 es un (n, M_1, d_1)-código y C_2 un (n, M_2, d_2)-código, ambos con la misma longitud, entonces $C_1 \oplus C_2 = \{c(c+d) | c \in C_1, d \in C_2\}$ con parámetros $(2n, M_1 M_2, min\{2d_1, d_2\})$.