Lecture 15

Tuesday, April 08, 2014
9:52 PM

Topics:

1. Compression over blocks
2. Error Detection and Correction
   - Reed-Solomon
   - Parity bit
   - Hamming code
3. Secrecy (ran out of time in this lecture)

Lossless Compression over blocks:

We saw for Huffman: $H(X) \le E\,L_H(X) < H(X) + 1$ bits

One bit gap in bounds (overhead)

Encode over blocks:
1. The overhead becomes negligible.
2. Take advantage of correlations.

(t h)(e)(an)(g r)(y)(b)(ird)

e.g. Code: aa $\to$ 00100
ab $\to$ 010
⋮

Example without correlation:

$\Omega = \{a, b\}$, $\quad Prob(X=a) = 0.9$, $\quad Prob(X=b) = 0.1$

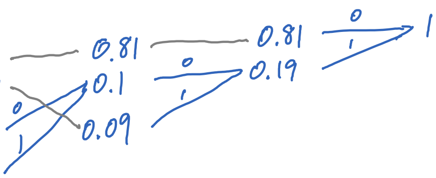
$$H(X) = 0.9 \cdot \log_2 \frac{1}{0.9} + 0.1 \cdot \log_2 10 \approx 0.469 \text{ bits}$$

Huffman code: Code: a $\to$ 0, b $\to$ 1 $\qquad E\,L_H(X) = 1$

Encode over pairs: $(X_1, X_2) \in \Omega$, $\quad \Omega = \{aa, ab, ba, bb\}$

$p(aa) = 0.9^2 = 0.81$ by independence
$p(ab) = p(ba) = 0.09$
$p(bb) = 0.01$

$C(X_1,X_2)$		
0	a a	0.81
1 1	a b	0.09
1 0 0	b a	0.09
1 0 1	b b	0.01



$$E\,L_H(X) = 0.81 \cdot 1 + 0.09 \cdot 2 + 0.09 \cdot 3 + 0.01 \cdot 3$$
$$\approx 1.2 \text{ bits}$$

≈ 1.3 bits.

$$\text{bits/symbol} \approx \frac{1.3}{2} = 0.65 \text{ bits}$$

Bounds for longer blocks:

Let L_n^* be the normalized expected length of the Huffman code for block length n.

$$L_n^* = \frac{1}{n} E L_H(X_1 X_2 \dots X_n)$$

$$n H(X) = H(X_1 X_2 \dots X_n) \le n L_n^* < H(X_1 X_2 \dots X_n) + 1 \text{ bits.}$$

assuming X_i are indendent and identically distributed.

$$\Rightarrow H(X) \le L_n^* < H(X) + \frac{1}{n} \text{ bits.}$$

In general, "entropy rate" is the fundamental compression limit.

Error Detection and Correction:

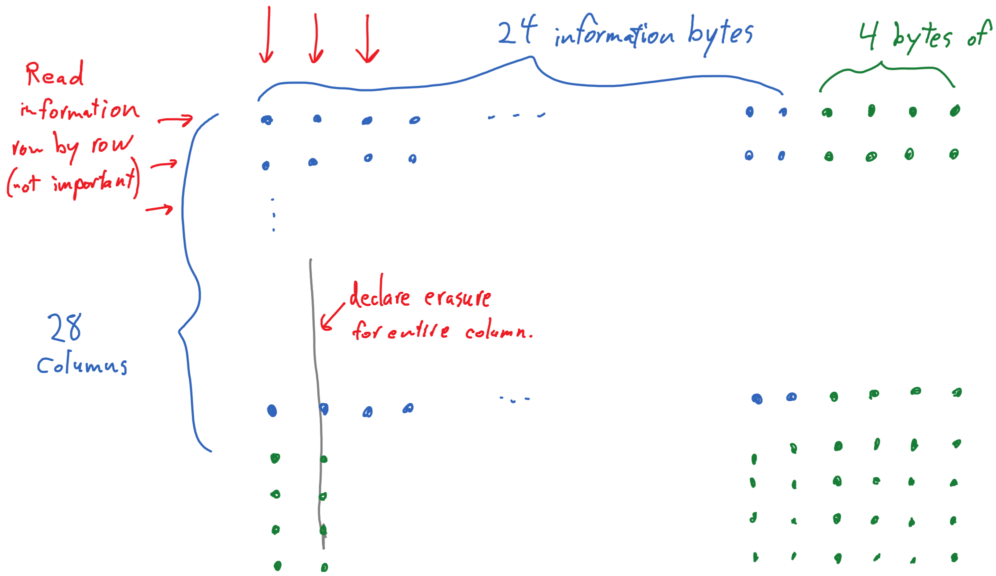
- Reed-Solomon

1. Many applications (1960).
2. Treats blocks of bits as symbols.
3. Take n symbols of information.
4. Add t symbols of redundancy. (used finite-field arithmatic)
5. Correct errors and erasures.
   $2 \cdot N_{errors} + N_{erasures} \le t$
6. If it cannot correct, it usually declares failure.
7. Decoder can be almost always understood with the Fourier transform.

As Seen on CD!

- symbols are bytes. (8 bits)
- 2 layers with interleaving.

Write to disk
column by column



Probability of error:
- need correct noise model.
  - good for bursty noise
  - not so good for scatted noise

Rate of this encoding: $\frac{24}{28} \cdot \frac{28}{32} = \frac{3}{4}$.

Error Detection:

Single bit error detection using parity bit.

Information bits: $\underbrace{0010111}_{\text{7 information bits}}$ 0 ← parity bit.

Parity check: $\sum_{i=1}^{\infty} b_i \bmod 2 \stackrel{?}{=} 0$

- Cannot detect 2 errors.
- Can detect odd numbers of errors.

Correction:

Repetition is simplest: Repetition code (3,1): Repeat each bit 3 times. Decode the majority.

Information	Codeword
0	000
1	111

Correct single errors.

Repetition (5,1):

0	00000
1	11111

Correct up to 2 errors.

Repetition (5,1):

0	00000
1	11111

Correct up to 2 errors.

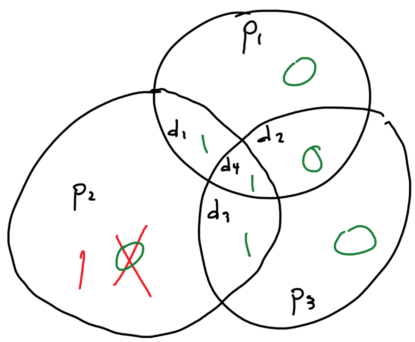
Hamming codes:

Correct 1 error.

Family of code.

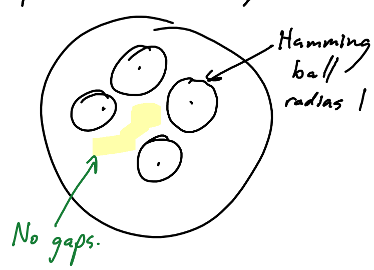
Hamming (7,4) code: — 4 bits of information.

d_1	d_2	d_3	d_4	p_1	p_2	p_3
1	0	1	1	0	$\cancel{0}$ 1	0



"Perfect code"

Space of 7-bit sequences:



Parity check

p_1	p_2	d_1	p_3	d_2	d_3	d_4	p	d	d	d			...
×		×		×		×		×		×			
	×	×			×	×			×	×			
			×	×	×	×					×	×	...
							×	×	×	×	×	×	...

Analyze prob. of error:

Under the "binary symetric channel" model.

Suppose each bit has error with prob. P_b.

Repetition (3,1): $P(\text{error}) = P(\text{two errors}) + P(3\text{ errors})$

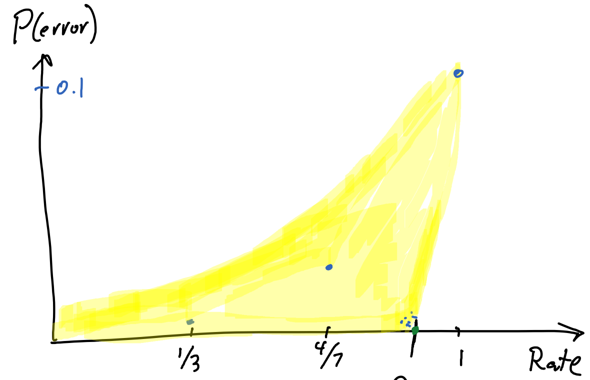
$$= 3 \cdot p_b^2 (1-p_b) + p_b^3$$

Hamming(7,4): $P(\text{error}) = 1 - P(0\text{ error}) - P(1\text{ error})$

$$= 1 - (1-p_b)^7 - 7\, p_b (1-p_b)^6 \approx 21\, p_b^2$$

$p_b = 0.01$

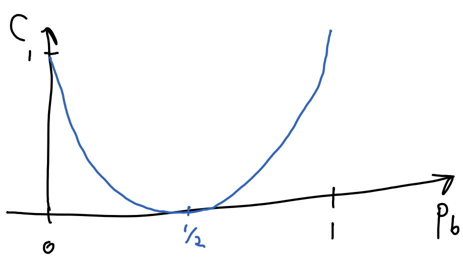
	P(error)
Hamming (7,4)	0.00203
Repetition (3,1)	0.000298
Uncoded (1,1)	0.01



$C = 0.919$ bits.

BSC:

$$C = 1 - p_b \log_2 \frac{1}{p_b} - (1-p_b)\log_2 \frac{1}{1-p_b}$$



Capacity formula is based on entropy.