CHAPTER 17 – INFORMATION SCIENCE

Binary and decimal numbers – a short review:

For decimal numbers we have 10 digits available (0, 1, 2, 3, … 9)

	4	7	3	1
10,000	1000	100	10	1
10^4	10^3	10^2	10^1	10^0

4731 = $4(1000) + 7(100) + 3(10) + 1(1)$

For binary numbers we have 2 digits available, 0 and 1.

1	0	1	1	0	0	0
	1	1	0	1	1	1
64	32	16	8	4	2	1
2^6	2^5	2^4	2^3	2^2	2^1	2^0

Express the following binary numbers as decimal numbers:

16 8 4 2 1
10101 = $1(16) + 1(4) + 1(1) = 16 + 4 + 1 = 21$

128 64 32 16 8 4 2 1
11100010 = $128 + 64 + 32 + 2 = 226$

Express the following decimal numbers as binary numbers:

55 = 110111_{TWO}

55
−32
23
−16
7
−4
3
−2
1

88 = 1011000_{TWO}

88
−64
24
−16
8
−8
0

An orbiting satellite can follow 16 different directions that are labeled 0 to 15 in binary (0000 to 1111). Each message is sent as the command along with 3 check digits. The check digits are arranged so that certain sums have even parity. These are called ***parity-check sums*** where the parity of a number refers to whether a number is even or odd. Even numbers have ***even parity*** and odd numbers have ***odd parity***.

For our satellite, the following sums must be even (0 mod 2)

$a_1 + a_2 + a_3 + c_1$, $a_1 + a_3 + a_4 + c_2$, and $a_2 + a_3 + a_4 + c_3$

1 0 0 +1 ← to make sum even

1 0 1 +0 already even sum

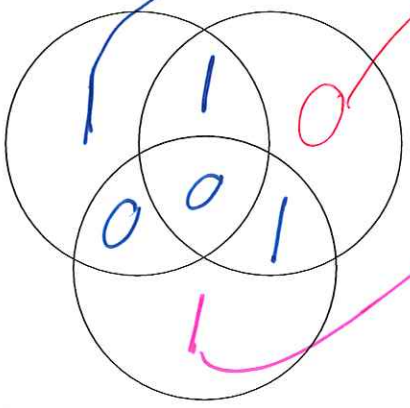
0 0 1 +1 to make sum even

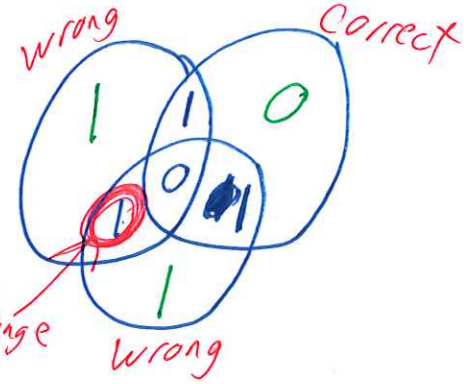
What are the check digits for command 9? 1001101

write 9 in binary

1	0	0	1
8	4	2	1

For this type of parity-check sum, we can use a Venn diagram to help find the check digits or find errors.





Fix the error in the code 1101101 if it is known only one digit has an error.

1001101

A set of words composed of 0's and 1's that has a message and parity check sums appended to the message is called a ***binary linear code***. The resulting strings are called ***code words***.

The process of determining the message you were sent is called ***decoding***. If you are sent a message x and receive the message as y, how can it be decoded?

The ***distance between two strings*** of equal length is the number of positions in which the strings differ.

(a) 10101 and 11101
distance of 1

(b) 111111 and 000000
distance of 6

The ***nearest neighbor decoding method*** decodes a message as the code word that agrees with the message in the most positions *provided there is only one such message*.

How good a code is at detecting and correcting errors is determined by the weight of the code. The ***weight of a binary code*** is the minimum number of 1's that occur among all non-zero code words of that code.

Consider a code of weight t,

- The code can *detect* $t-1$ or fewer errors.
- The code can *correct* half as many as it can detect (rounded down).
  - $\frac{t-1}{2}$ or fewer errors if t is odd.
  - $\frac{t-2}{2}$ or fewer errors if t is even.

distance 3 1 5 4

0001101 0001101 0001101 0001101

Consider the code C = {0000000, 0001111, 1111000, 1111111}

4 4 7

(a) What is the weight of the code? 4

(b) How many errors can this code detect?

weight − 1 = 4 − 1 = 3

(c) How many errors can this code correct?

Half as many as detect (rounded down) $\frac{3}{2} = 1.5$ round down to 1

or $\frac{4-2}{2} = \frac{2}{2} = 1$

(d) Decode the message received as 0001101.

Nearest Neighbor 0001111 b/c smallest distance

A ***compression algorithm*** converts data from an easy-to-use format to one that is more compact. jpg photo files use data compression as do most video and audio files.

Delta function encoding uses the beginning value and the differences in one value to the next to encode the data.

The data below is the closing price of the Dow Jones on Oct. 1, 2012 – Oct 5, 2012. Compress the data using delta function encoding and determine how much the data is compressed.

13610 13575 13495 13482 13515

13610 −35 −80 −13 33

do not write "+"

original 25 characters

compressed 16 characters

saved: 25 − 16 = 9 characters

as a percentage

$\frac{\text{Savings}}{\text{original}} = \frac{9}{25} = .36 = 36\%$ compression

uncompressed

Binary codes can also be compressed by assigning short codes to characters that occur frequently and longer codes to characters that occur rarely.

We have 5 symbols, A, B, C, D, and E. If we give all the symbols a code of the same length, we would need 3 binary digits (000 to 101). So a string of 6 symbols would be 6 x 3 = 18 characters long. Can we devise a different binary code if we knew how often each character occurred?

Use ***Huffman coding*** is a way to assign shorter code words to those characters that occur more often.

Step 1 Arrange these letters from least to most likely.

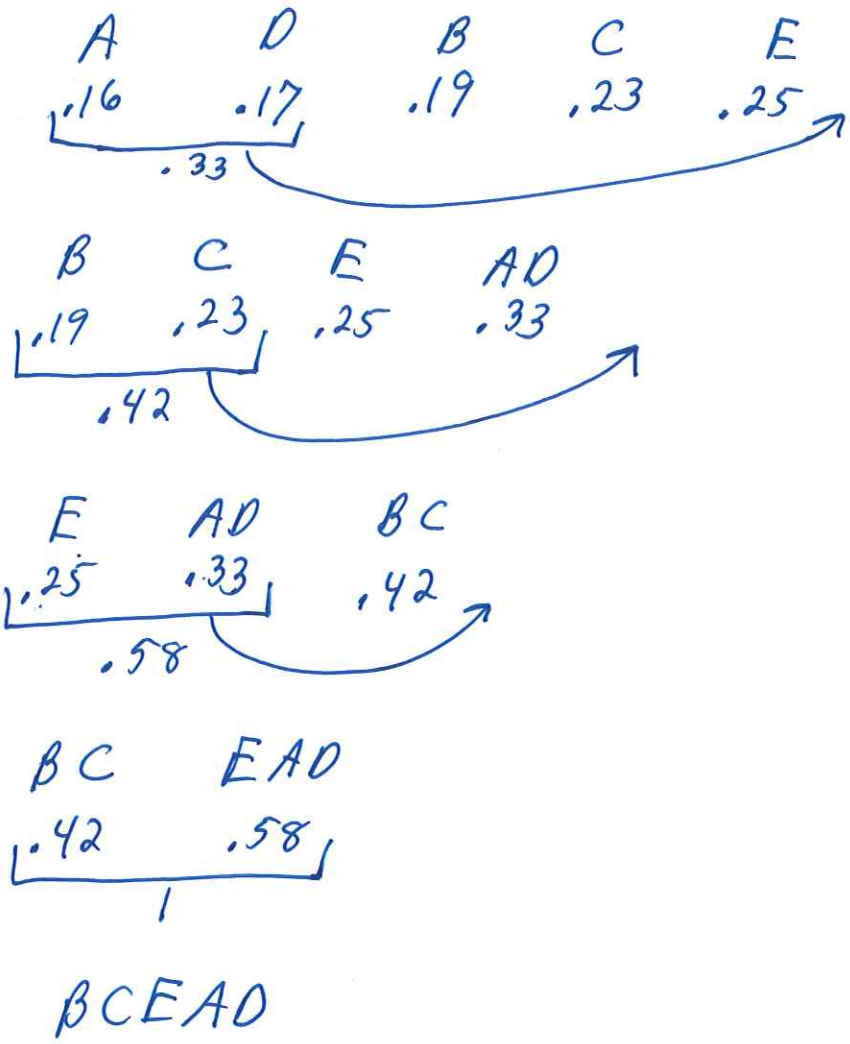
Step 2 Add the probabilities of the two least likely characters and combine them. Keep the letter with the smaller probability on the left. Arrange the new list from least to most likely.

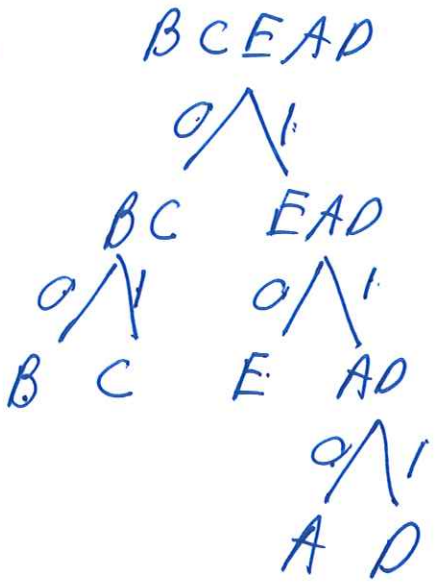
Step 3 Repeat Step 2 until all the letters have been combined into one group with probability of 1.

Step 4 To assign a binary code to each letter, display the information in a Huffman tree by undoing the process from Steps 2 and 3. Always keep the smaller probability on the left and assign a 0 to that branch. Assign a 1 to the branch with the higher probability.

Step 5 The 0's and 1's for each path determine the code word for that letter. Read from the top of the chart down to the letter.

A	B	C	D	E
0.16	0.19	0.23	0.17	0.25





A	B	C	D	E
110	00	01	111	10

Decipher a message that was encoded using this Huffman code:

10|1100|100|111|10|10|10|0001

E|A|C|B|D|A|E|E|C|B|C

The process of disguising data is called encryption. Cryptology is the study of making and breaking secret codes.

A ***Caesar cipher*** shifts the letters of the alphabet by fixed amount.

EXAMPLE

Create a Caesar cipher that shifts the alphabet by 10 letters and use it to encrypt the message THANKS.

DRKXUC

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J

TELEPHONE

The message HSZSDVCBS was created with a Caesar cipher with a shift of 14. What is the original message?

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	← orig
O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	← coded

A ***decimation cipher*** multiplies the position of each letter by a fixed number k (called the ***key***) and then uses modular arithmetic. To use a decimation cipher,

1. Assign the letters A – Z to the numbers 0 – 25.
2. Choose a value for the key, k, that is an odd integer from 3 to 25 but not 13 (why not?)
3. Multiply the value of each letter (i) by the key (k) and find the remainder when divided by 26.
4. To decrypt a message, the encrypted value x needs to be multiplied by the decryption letter j and then the remainder mod 26 is the original letter.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Use a decimation cipher with a key of 11 to encrypt THANKS : BZANGQ

	T	H	A	N	K	S
Position	19	7	0	13	10	18
(Pos)*(e.key) Mul by 11	209	77	0	143	110	198
Mod 26	1	25	0	13	6	16
Code	B	Z	A	N	G	Q

$26\overline{)209}$ = 8 R 1 ← (209 − 208)

$26\overline{)77}$ = 2 R 25 ← (77 − 52)

$26\overline{)0}$ = 0 R 0 ← (0 − 0)

77 − 26 = 51; 51 − 26 = 25

Check that 5 is decryption key for 21: Yes, 5 is ↑

$21 \cdot 5 \bmod 26 \equiv 1$

$105 \bmod 26$: $26\overline{)105}$ = 4 R 1 ← (105 − 104)

The message below was encrypted with a key of 21. The decryption key is 5. Decode the message.

	Q	R	G	S	M	O	J	T	K
Position	16	17	6	18	12	14	9	19	10
(Pos)*(d.key) *5	80	85	30	90	60	70	45	95	50
Mod 26	2	7	4	12	8	18	19	17	24
Message	C	H	E	M	I	S	T	R	Y

A ***Vigenère cipher*** uses a ***key word*** to encode the characters.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

Use a ***Vigenère cipher*** with a key word of MINT to encode the message

	N	E	W		P	R	I	N	T	E	R
Position	13	4	22		15	17	8	13	19	4	17
Key Word	M 12	I 8	N 13		T 19	M 12	I 8	N 13	T 19	M 12	I 8
Sum	25	12	35		34	29	16	26	38	16	25
Mod 26	25	12	9		8	3	16	0	12	16	25
Code	Z	M	J		I	D	Q	A	M	Q	Z

A ***Vigenère cipher*** with a key word of LEX was used to encode the message below. Decode it. You will need to subtract rather than add.

	D	Y	M	P	V	J	L	R
Position	3	24	12	15	21	9	11	17
Key Word	L 11	E 4	X 23	L 11	E 4	X 23	L 11	E 4
Diff.	-8	20	-11	4	17	-14	0	13
Mod 26	18	20	15	4	17	12	0	13
Code	S	U	P	E	R	M	A	N

-8+26

To increase security, binary strings can be added together. If the result is even, enter 0. If the result is odd, enter 1. In other words, take the sum of the digits mod 2. *Note*: This is not the same as binary addition.
Add the binary strings:

(a) 10110
+ 00111
10001

(b) 10110
+ 10110
00000

SAMPLE EXAM QUESTIONS FROM CHAPTER 17

1. Convert the binary number 11001 to a decimal number.

(A) 3 (B) 25

(C) 6 (D) 31

16 8 4 2 1

16 + 8 + 1 = 25

2. What is the distance between received words 1100101 and 1010111?

(A) 1 (B) 2 (C) 3 (D) 4

(E) more than 4

1100101

1010111

3. Add the binary strings 1100101 and 1110001. How many 1s digits are in the sum?

(A) 1 (B) 2 (C) 3 (D) 4

(E) more than 4

1100101
+ 1110001
0010100

4. Use delta encoding to compress the data

1834 1831 1831 1825 1850.

1834 -3 0 -6 25

By how many characters is the data compressed?

(A) 9 (B) 10

(C) 11 (D) 13

original 20 characters

- Compressed data 11 characters

Saved 9 characters

5. Use the Venn diagram method to decode the received word 1101101.

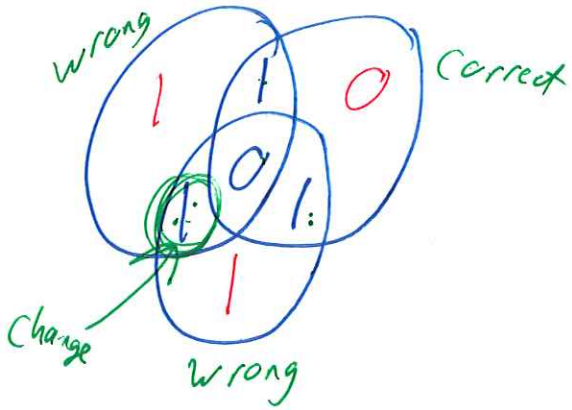
(A) 1001

(B) 0100

(C) 1101

(D) 1011

(E) None of these



2 2 2 2 2 2

Questions 6 and 7 use the code {1100, 1010, 1001, 0110, 0101, 0011}.

6. What is the weight of this code? counting # ones look for min

(A) 0 (B) 1 (C) 2 (D) 3 (E) 4

7. Which one of the following is a true statement about this code?

(A) This code can detect and correct two errors

(B) This code can detect two errors and correct 1 error

(C) This code can detect and correct one error.

(D) This code can detect one error and correct 0 errors

(E) None of these

detect $2-1=1$ error

correct $\frac{2-2}{2}=\frac{0}{2}=0$

or $\frac{1}{2}$ rounded down $=0$

8. Given binary codes A → 0, C → 10, I → 110, S → 1110, B →11110.

(a) Encode the message CASSI

10011101110110

(b) Decode the message 11110011101101011110

B A S I C S

9. What is the code word for the message 110, if the code word is the message appended with three check digits found using the parity-check sums $a_1 + a_2 + a_3$, $a_1 + a_3$ and $a_2 + a_3$?

110011

1 1 0 | 1 0 | 1 0

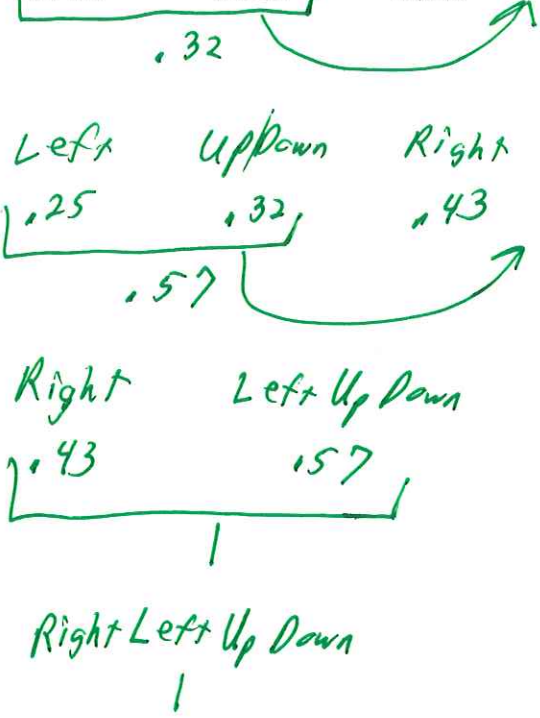
even so 1st check digit is 0

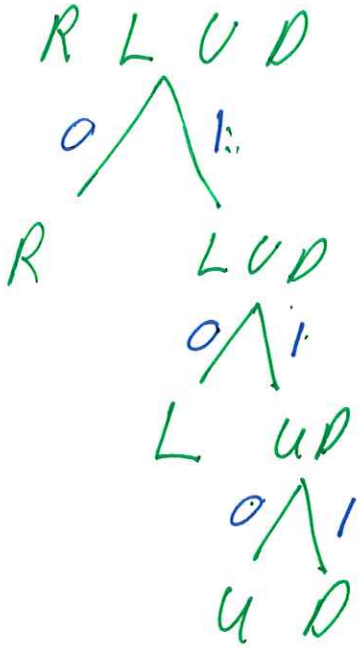
odd so 2nd check digit is 1

odd so 3rd check digit is 1

10. Use a Huffman code to assign binary codes to the directions that occur with the probabilities given below.

Up	Down	Left	Right
0.12	0.20	0.25	0.43





R = 0

L = 10

U = 110

D = 111

ELQDUB

11. Use a Caesar cipher with a shift of 3 to encode the word BINARY.

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

12. Use a decimation cipher with key 9 to encode the word CABLE.

	C	A	B	L	E
Position	2	0	1	11	4
Mul by 9	18	0	9	99	36
Mod 26	18	0	9	21	10
Code	S	A	J	V	K

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z

13. Use the Vigenere cipher with the key word PEN to encode BASEBALL.

	B	A	S	E	B	A	L	L
Position	1	0	18	4	1	0	11	11
Code word	P 15	E 4	N 13	P 15	E 4	N 13	P 15	E 4
Sum	16	4	31	19	5	13	26	15
Mod 26	16	4	5	19	5	13	0	15
Code	Q	E	F	T	F	N	A	P