CMPT 365 Written Assignment 2

Junchen Li

301385486

2021/3/18

Textbook Question

Section 7.8

Q1.

For calculate the entropy, using the formula $H = \sum_i -P(A_i) * \log_2 P(A_i)$ where $P(A_i)$ is the probability of A_i. Since we have a checkerboard image which have an image with half white and hale black. The P(black) = 0.5 and the P(white) = 0.5 as well. Then H = -0.5 * $\log_2 0.5$+ -0.5 * $\log_2 0.5$ $= \frac{1}{2} + \frac{1}{2} = 1$. So the entropy of a checkerboard image is 1.

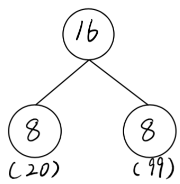
Q3.

The total numbers in this image are 64 and for each numbers (0,20,50,99) the probability of P(0) = $\frac{32}{64} = \frac{\cancel{32}\,1}{\cancel{64}\,2} = \frac{1}{2}$. P(20) = $\frac{8}{64} = \frac{\cancel{8}\,1}{\cancel{64}\,8} = \frac{1}{8}$, P(50) = $\frac{16}{64} = \frac{\cancel{16}\,1}{\cancel{64}\,4} = \frac{1}{4}$, P(99) = $\frac{8}{64} = \frac{\cancel{8}\,1}{\cancel{64}\,8} = \frac{1}{8}$. H = $-\frac{1}{2} * \log_2 \frac{1}{2} - 2(\frac{1}{8} * \log_2 \frac{1}{8}) - \frac{1}{4} * \log_2 \frac{1}{4} = \frac{1}{2} + \frac{3}{4} + \frac{1}{2} = \frac{7}{4}$. So the entropy of this image is 1.75.

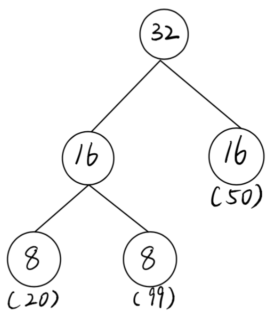
b) The summary table of all symbols frequency counts as follow: (Table sorted according to their frequency counts)

# (0)	# (50)	#(20)	#(99)
32	16	8	8

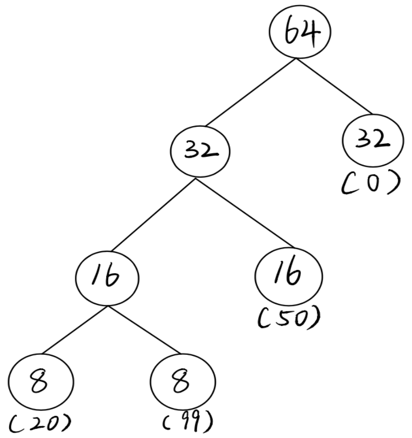
We first pick two symbols with lowest frequency counts and assign the sum of their children as a parent. We first pick # (20) and # (99) because of their frequency are 8 and sum their frequency as a new parent node



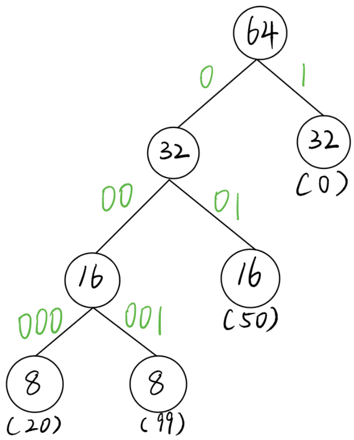
Then we compare between 32,16,16 and we choose two sixteen frequencies.



Then, we pick the last symbol 32 and combine them together:



Lastly, we set all left edges to “0” and all right edges to “1” for setting Huffman tree



We get a Huffman tree and the parallel table as follow:

0: “1”

50: “01”

20: “000”

99: “001”

c) The recall the P(0) = $\frac{1}{2}$ P(20) = $\frac{1}{8}$ P(50) = $\frac{1}{4}$ P(99) = $\frac{1}{8}$ The average Huffman code word length is

L = 0.5 * 1 + 0.125 * 3 + 0.25 * 2 + 0.125 * 3 = 0.5 + 0.375 + 0.5 + 0.375 = 1.75 bits/symbol.

In general, we say the average Huffman code length for an information source S is strictly less than entropy+1: 1.75 ≤ 1.75 < 2.75 → H(S) ≤ L < H(S) + 1.

Q13

We only know the input 0 has the output “0” and input 1 has the output “1”.

We will apply the LZW compression for the input “ 0 1 1 0 0 1 1”. The next table will show the process of LZW compressing.

S	C	Output	Code	String
0	1	0	2	“01”
1	1	1	3	“11”
1	0	1	4	“10”
0	0	0	5	“00”
0	1	Already exist in the dictionary		
01	1	2	6	“011”
1	EOF	1		

So the output is “ 0 1 1 0 2 1” Instead of seven characters, only six codes need to be sent.

EX1:

(1) We find second-order entropy for “0110110011001110001101110110011 1” and for each two pairs symbol we have four different combinations “00”, “01”, “10”, “11”. We change sequence to “01 10 11 00 11 00 11 10 00 11 01 11 01 10 01 11” there are total sixteen pairs. So we have P(00) = $\frac{3}{16}$ P(01) = $\frac{\cancel{4}\ 1}{\cancel{16}\ 4} = \frac{1}{4}$ P(10) = $\frac{3}{16}$ P(11) =$\frac{\cancel{6}\ 3}{\cancel{16}\ 8} = \frac{3}{8}$. H = $-\frac{1}{4} * \log_2 \frac{1}{4} - 2(\frac{3}{16} * \log_2 \frac{3}{16}) - \frac{3}{8} * \log_2 \frac{3}{8} = \frac{1}{2} + 0.9056 + 0.53063 = 1.93627$ So the entropy of this image is 1.93627.

(2) The recall the P(00) = $\frac{3}{16}$ P(01) = $\frac{1}{4}$ P(10) = $\frac{3}{16}$ P(11) = $\frac{3}{8}$ The average Huffman code word length is $L = \frac{3}{16}$* 2 +$\frac{1}{4}$ * 2 + $\frac{3}{16}$ * 2 + $\frac{3}{8}$ * 2 = 0.375 + 0.5 + 0.375 + 0.75 = 2 bits/pair symbols. And we should divide it by two for getting the average code word length (per symbol). 2/2 = 1 bits/symbol.

EX2:

(1)

S	C	Output	Code	String
			0	a
			1	b
			2	c
a	c	0	3	ac
c	b	2	4	cb

b	b	1	5	bb
b	a	1	6	ba
a	c	Already exist in the dictionary		
ac	a	3	7	aca
a	EOF	0		

The output sequence for input "a c b b a c a" by using LZW coding is 0 2 1 1 3 0.

(2) The dictionary in LZW coding:

The dictionary in LZW coding								
String	"a"	"b"	"c"	"ac"	"cb"	"bb"	"ba"	"aca"
Code	0	1	2	3	4	5	6	7

EX3

We have sequence "a a b b b a" and we can set the P(a) = $\frac{\cancel{3}1}{\cancel{6}2} = \frac{1}{2}$, P(b) = $\frac{\cancel{3}1}{\cancel{6}2} = \frac{1}{2}$. We set the range of "a" as $[0,0.5)$ and "b" as $[0.5,1)$. Whole process steps as follow:

Initial			High	Low
	0 a 0.5 b 1		1	0
"a"	0 a 0.25 b 0.5		0.5	0
"a"	0 a 0.125 b 0.25		0.25	0
"b"	0.125 a 0.1875 b 0.25		0.25	0.125
"b"	0.1875 a 0.21875 b 0.25		0.25	0.1875
"b"	0.21875 a 0.234375 b 0.25		0.25	0.21875
"a"	0.21875 a 0.2265625 b 0.234375		0.234375	0.21875

So the lower bound is 0.21875 and higher bound is 0.234375. So the final bound of final interval in arithmetic coding is [0.234375,0.21875].