CS 6351 Data Compression | Fall 2022
Youssef

Solution to Homework 1

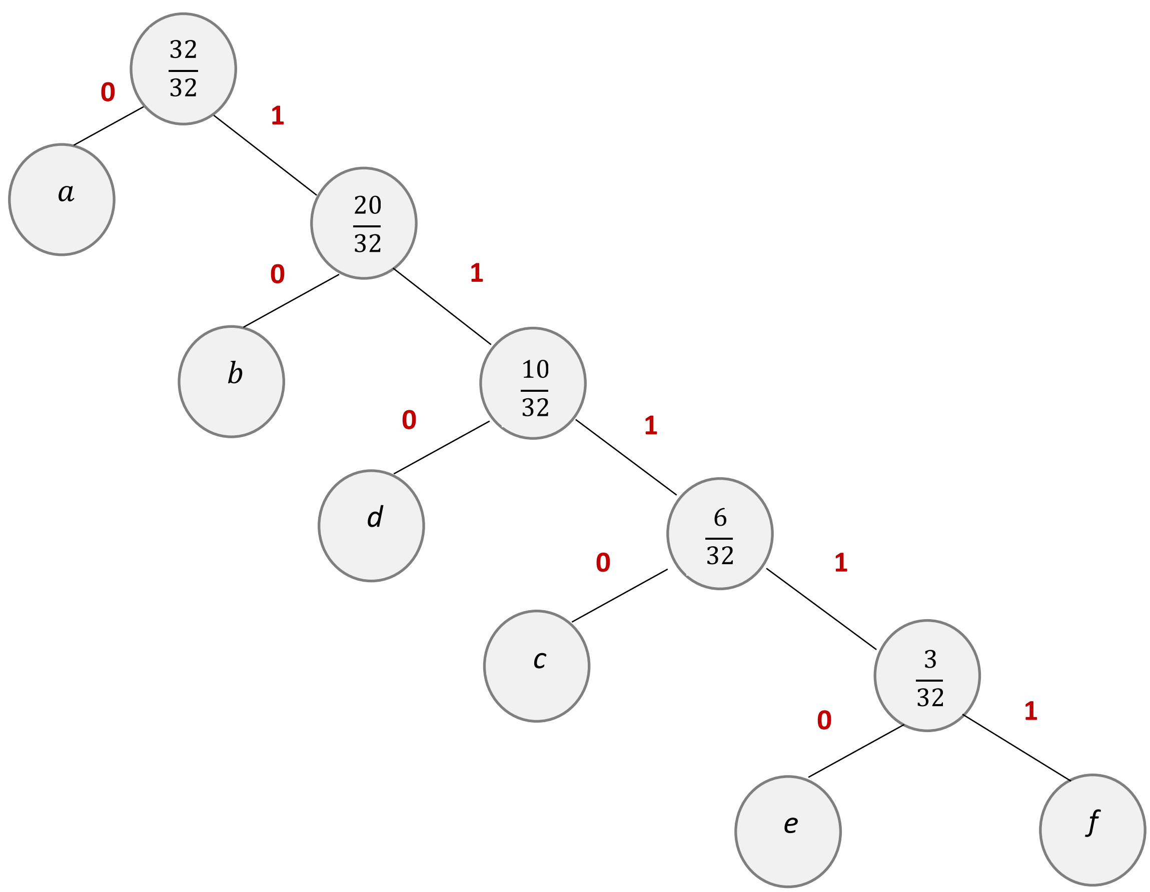
Problem 1: (20 points)

a.

$$H = -\sum_{\theta=a}^{f} P[\theta]\log P[\theta]$$
$$= -\left(\frac{12}{32}\log\frac{12}{32} + \frac{10}{32}\log\frac{10}{32} + \frac{3}{32}\log\frac{3}{32} + \frac{4}{32}\log\frac{4}{32} + \frac{1}{32}\log\frac{1}{32} + \frac{2}{32}\log\frac{2}{32}\right)$$
$$\approx 2.1$$

b.

(Note that the eligible Huffman tree is not unique.)



Alphabet	Code
a	0
b	10
c	1110
d	110
e	11110
f	11111

c.

$$BR_{Huffman} = \frac{12}{32} \cdot 1 + \frac{10}{32} \cdot 2 + \frac{3}{32} \cdot 4 + \frac{4}{32} \cdot 3 + \frac{1}{32} \cdot 5 + \frac{2}{32} \cdot 5 \approx 2.2$$

d.

$$Input = a^5 b^3 d^3 bdeba^4 c^2 baf$$

$$Coded\ Bitstream = (0^5)(10)^3(110)^3(10)(110)(11110)(10)(0000)(1110)^2(10)0(11111)$$

$$BR = \frac{52}{24} = 2.1$$

Problem 2: (20 points)

a.

1.

$$L = 0$$
$$R = 1$$
$$L + P[1^{st} bit = 0]\Delta = \frac{1}{2}$$
$$I = \left[0, \frac{1}{2}\right)$$

2.

$$L = 0$$
$$R = \frac{1}{2}$$
$$L + P[0|0]\Delta = \frac{61}{64} \cdot \frac{1}{2}$$
$$I = \left[0, \frac{61}{2(64)}\right)$$

3.

$$L = 0$$
$$R = \frac{61}{2(64)}$$
$$L + P[0|0]\Delta = \frac{61}{2(64)} \cdot \frac{61}{64}$$
$$I = \left[0, \frac{61^2}{2(64)^2}\right)$$

4.

$$L = 0$$
$$R = \frac{61^2}{2(64)^2}$$
$$L + P[0|0]\Delta = \frac{24389}{65536}$$
$$I = \left[\frac{61^2}{2(64)^2}, \frac{61^3}{2(64)^3}\right)$$

5.

$$L = \frac{61^2}{2(64)^2}$$
$$R = \frac{61^3}{2(64)^3}$$
$$L + P[0|1]\Delta = \frac{14560273}{33554432}$$
$$I = \left[\frac{14560273}{33554432}, \frac{61^3}{2(64)^3}\right)$$

6.

$$I = \left[\frac{933900301}{214748364}, \frac{61^3}{2(64)^3}\right)$$

$$t = \lceil -\log(R - L) \rceil = \left\lceil -\log\left(\frac{41537523}{2147483648}\right) \right\rceil = 6 \text{ (the coded bitstream will be 6 bits)}$$

$$\frac{L + R}{2} = 0.44 = (0.\mathbf{011100}01)_2$$

$$\therefore Code = 011100$$

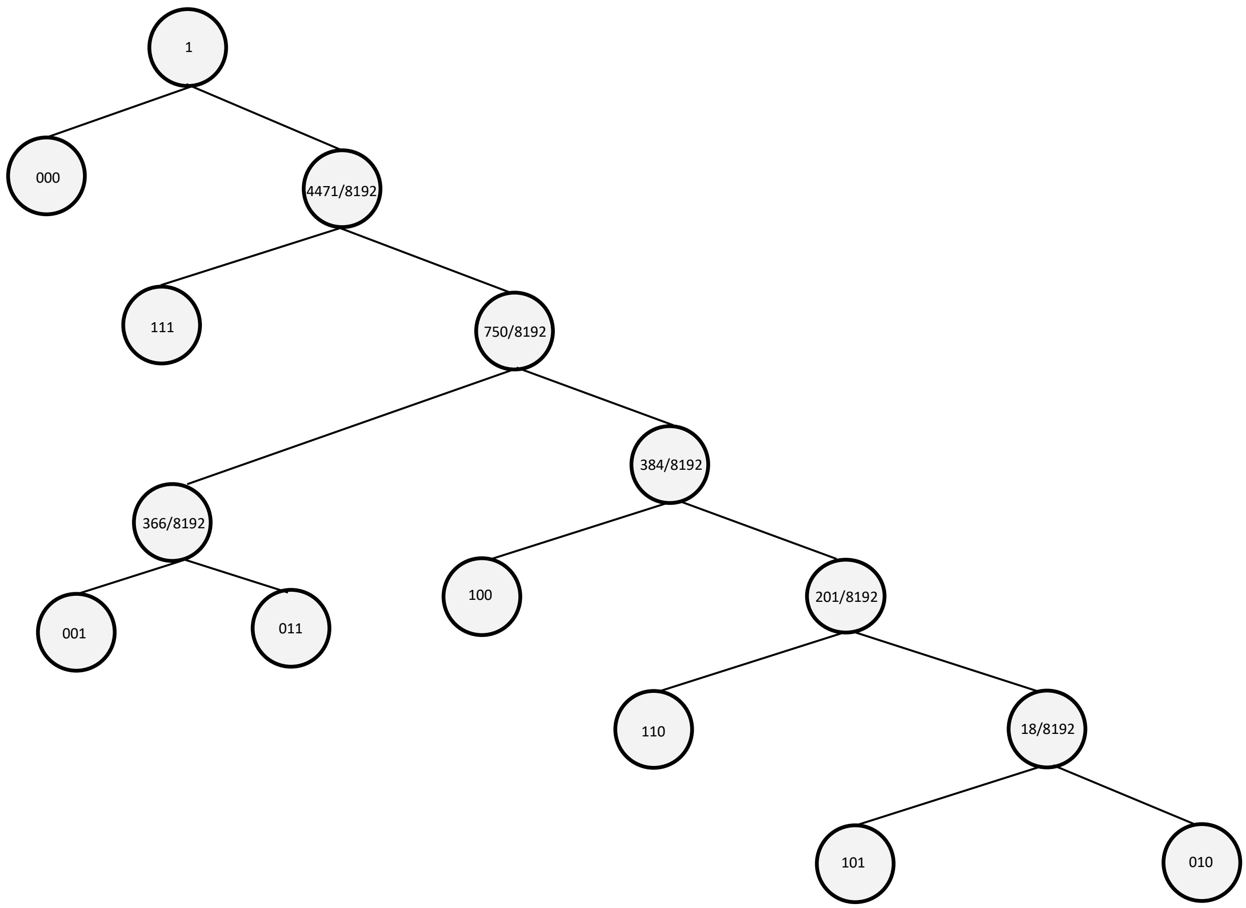
$$BR_a = \frac{6}{6} = 1$$

b.

$$P[000] = P[0]P[0|0]P[0|0] = \frac{61^2}{8192}$$
$$P[001] = P[0]P[0|0]P[1|0] = \frac{183}{8192}$$
$$P[010] = P[0]P[1|0]P[0|1] = \frac{9}{8192}$$
$$P[011] = P[0]P[1|0]P[1|1] = \frac{183}{8192}$$
$$P[100] = P[1]P[0|1]P[0|0] = \frac{183}{8192}$$

$$P[101] = P[1]P[0|1]P[1|0] = \frac{9}{8192}$$
$$[110] = P[1]P[1|1]P[0|1] = \frac{183}{8192}$$
$$P[111] = P[1]P[1|1]P[1|1] = \frac{61^2}{8192}$$

(Note that the eligible Huffman tree is not unique.)



Block	Code
000	0
111	10
001	1100
011	1101
100	1110
110	11110
010	111111
101	111110

$$\therefore Coded\ Bitstream = 0\,10$$

$$BR_b = \frac{3}{6} < BR_a = \frac{6}{6}$$

Problem 3: (20 points)

a. (0, 1111) (1, 1011) (0, 1010) (1, 1101) (0, 1110) (1, 1001)
$Coded\ Bitstream =$ 111110111010110111101001

b.

$$p(0) = \frac{39}{72}$$
$$\therefore MPB = 0$$
$$\frac{p\ln 2}{1-p} = \frac{39}{72}\ln 2 \cdot \frac{33}{72} \approx 0.8$$
$$\therefore m = 1, log_2 m = 0$$

$0^{15}1$: q = 15 r = 0 =>$1^{15}0$
$0^{0}1$: q = 0 r = 0 => $1^{0}0$
$0^{10}1$: q = 10 r = 0 => $1^{10}0$
$0^{14}1$: q = 14 r = 0 => $1^{14}0$

$Tail\ bit$ = 1

$$Coded\ Bitstream = 0\ 1^{15}0^{11}\ 1^{10}0^{13}\ 1^{14}\ 0^{9}\ 1$$

c.

$$z' = 0^{15}10^{10}10^{9}10^{12}10^{13}10^{8}$$
$$MPB = 0, p = \frac{67}{72}$$
$$\frac{p\ln 2}{1-p} \approx 9.28$$
$$\therefore m = 8, \log m = 3$$

$Tail\ bit$=0

$$Coded\ Bitstream = \ 0\ 10111\ 10010\ 10001\ 10100\ 10101\ 10000\ 0$$
$$BR_{RLE} = \frac{24}{72} < BR_{DG} = \frac{32}{72} < BR_{G} = \frac{74}{72}$$

Problem 4: (20 points)

a.

i	j	$Dict[i]$
1	*empty*	0
2	1=1	0^2
3	10	0^3
4	11	0^4
5	100	0^5
6	000	1
7	110	1^2
8	111	1^3
9	1000	1^4
10	0110	10
11	0101	0^5
12	0011	$0^3 1$
13	1001	1^5
14	1101	1^6
15	1010	100
16	1011	0^7
17	00101	$0^5 1$
18	01110	1^7
19	00110	*empty*

$Coded\ Bitstream$: 0101001101000000111011111100010110001010001111001111011110100101100010110111010011 0

Coded Bitstream: $0101001101000000111011111100010110001010001111001111011110100101$
$100010110111010 0110$

$$\#\ bits = 82 \quad BR = \frac{82}{72} > BR_{RLE} = \frac{24}{72}$$

b.

$$y = a^4 b^4 a b^3 a b a^2 b a^2 b^2 a^3$$

i	$Dict[i]$
1	a
2	a^2
3	ab
4	b
5	bb
6	ab^2
7	ba

8	baa
9	$baab$
10	ba^3

$$Code: a(1,a)(01,b)(00,b)(100,b)(011,b)(100,a)(111,a)(1000,b)(1000,a)$$

$Let\ a = 0\ and\ b = 1:$

$$Coded\ Bitstream = 0\ 1\ 0\ 011\ 001\ 1001\ 0111\ 1000\ 1110\ 10001\ 10000$$

$\#\ bits = 35$ $\quad BR = \frac{35}{25}$

Problem 5: (20 points)

The key solution is to subdivide the unit interval [0 1] into 3 subintervals, of relative lengths P[a|y], P[b|y] and P[c|y], respectively, and to correctly select the corresponding sub-interval for every bit in the input stream.

Input: Input string x of alphabet = {a,b,c}

Output: a coded bitstream

Method:

1. Let $I = [L, R)$ where initially $L = 0, R = 1$
2. For $i = 1$ to n do

    Let $P_a = \Pr[a|x_1x_2 \ldots x_{i-1}]$, $P_b = \Pr[b|x_1x_2 \ldots x_{i-1}]$; $P_c = \Pr[c|x_1x_2 \ldots x_{i-1}]$.

    Let $\Delta = R - L$;

    D1= $L + P_a\Delta$

    D2= $\mathrm{D1} + P_b\Delta$

    Divide interval into 3 subintervals:

    $[L, \mathrm{D1})$, $[\mathrm{D1}, \mathrm{D2})$ and $[\mathrm{D2}, \mathrm{R})$;

    Interval splitting: the algorithm splits the interval [L R) into three subintervals

    If $(x_i == a)$, reduce I to $[L, \mathrm{D1})$;

    Else if $(x_i == b)$, reduce I to $[\mathrm{D1}, \mathrm{D2})$;

    Else, reduce I to $[\mathrm{D2}, \mathrm{R})$;

    Subinterval selection (update of L and R)

3. Let $t = \lceil -\log(R - L) \rceil$, and r= (L+R)/2 expressed in binary as $0.r_1r_2 \ldots r_t \ldots$ (stop at r_t)

Output $= r_1r_2 \ldots r_t$