Summary on Lecture 17, October 8, 2017

Prefix Codes

We choose the prefix code 00, 01, 100, 1010, 1011, 11. This set could serve as a code for the letters in an alphabet $\Sigma = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ that has six letters. It is the set of leaves of a labeled binary tree in Fig. 1.
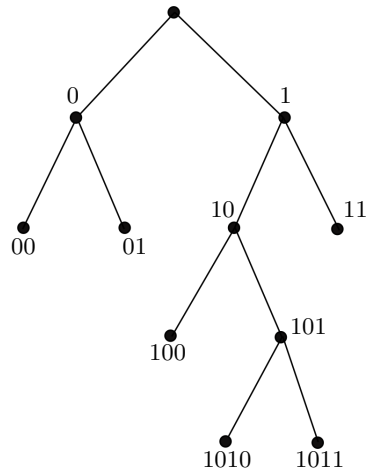

Fig. 1. Prefix code. Here $a_1 := 00$, $a_2 := 01$, $a_3 := 100$, $a_4 := 1010$, $a_5 := 1011$, $a_6 := 11$.

Now assume that we are given the binary string

$$1110101101100010 0111110010$$

We visit vertex 1, then vertex 11. Since vertex 11 is a leaf, we record 11 and return to the root. Next we visit vertices 1, 10, 101, and 1010. Since 1010 is a leaf, we record 1010 and return again to the root. Proceeding in this way, we obtain the sequence of code symbols

11, 1010, 11, 01, 100, 01, 00, 11, 11, 100, STOP

and have 10 left over which does not give any symbol encoded by the leaves. Now we suppose that we know how frequently each letter in $\Sigma = \{a_1, a_2, a_3, a_4, a_5, a_6\}$ is used in sending messages. Namely, we assume that the frequencies 10, 10, 15, 20, 20, 25 correspond to the letters $a_1, a_2, a_3, a_4, a_5, a_6$.

Then we would like to find the most efficient prefix code, i.e. we want to minimize the average length of a code message using (say, 100 letters) from Σ. Thus all we need is an **optimal binary tree for the weights** $\{10, 10, 15, 20, 20, 25\}$. Now we run Huffman algorithm and obtain the optimal tree, see Fig. 2 (b):
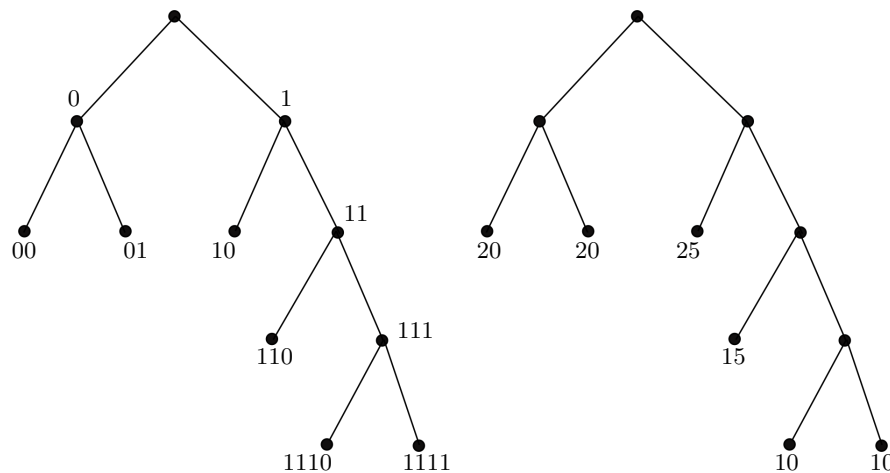

Fig. 2   (a)   (b)

1

Then the same tree (see Fig. 2 (b)) with binary labels gives new prefix code

$$a_1 := 1110, \ a_2 := 1111, \ a_3 := 110, \ a_4 := 00, \ a_5 := 01, \ a_6 := 10.$$

This is an **optimal prefix code for given frequences.**

Example. Here is a prefix code: $\{00, 010, 0110, 0111, 10, 11\}$ We construct a binary tree corresponding to this prefix code:
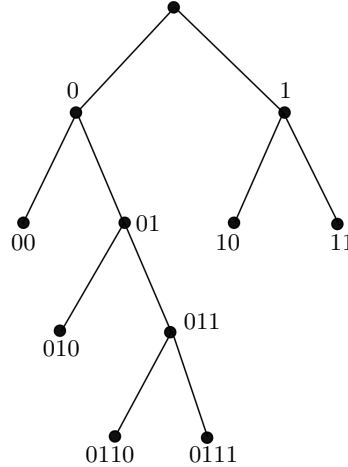

Fig. 3. Prefix code: here $00 :=$A, $10 :=$D, $11 :=$E, $010 :=$H, $0110 :=$M, and $0111 :=$' (the apostrophe)

Now we decode the string

$$01011011101100010110111011011011101100010$$

to get the string HE'MADE'ME'MAD (he made me mad).

Exercise 1. Assume that the letters A,D,E,H,M,' have frequences 30, 25, 27, 20, 33, 10 respectively. Find an optimal prefix code and encode the message "HEMADEMEMAD".

Exercise 2. Use the prefix code from Fig. 4 to decode the string
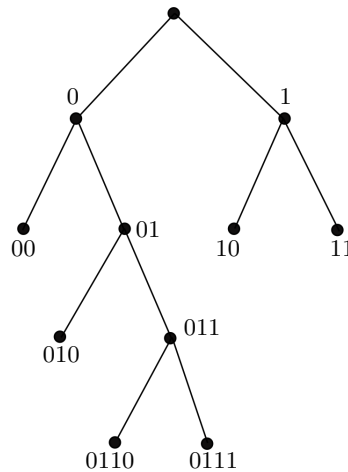
$$00100100001100001110001010001011$$

Fig. 4. Prefix code: here $00 :=$A, $10 :=$B, $11 :=$!, $010 :=$R, $0110 :=$C, and $0111 :=$D