Correctness of Huffman coding

Correctness

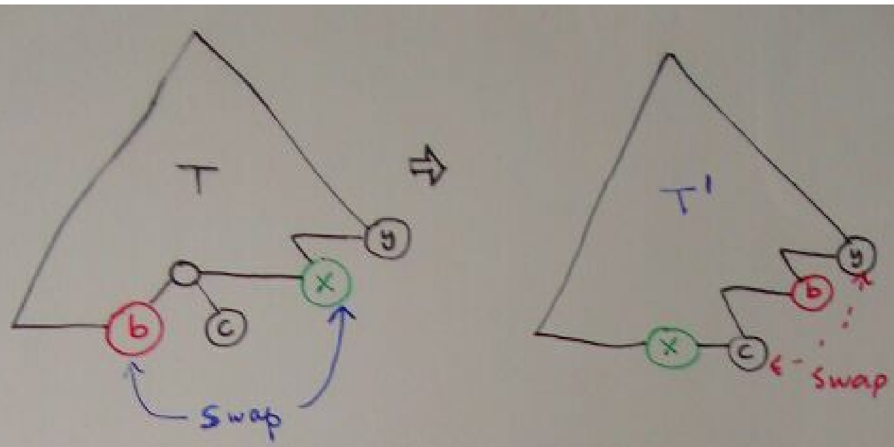
1. Show there is an optimal code tree agrees with our FIRST GREEDY CHOICE.

Proof:

Let T be an optimal code tree such that "b" and "c" appears at the LOWEST LEVEL and they are SIBLING LEAVES.
(兄弟)

Let "x" and "y" be the characters which have LOWEST FREQUENCIES.

THE GOAL is to construct an optimal tree T'' s.t. "x" and "y" appears in the LOWEST LEVEL and they are SIBLING LEAVES.



1. $\text{cost}(T) \leq \text{cost}(T')$

2. $$\text{cost}(T') = \text{cost}(T) - f(b)\,d(b) - f(x)\,d(x) + f(x)\,d(x) + f(b)\,d(x)$$

$$= \text{cost}(T) + \underbrace{\left[f(x) - f(b)\right]}_{\leq 0}\underbrace{\left[d(b) - d(x)\right]}_{\geq 0}$$

(∵ x has the SMALLEST freq.) (∵ b appears in the LOWEST LEVEL)

$\Rightarrow \text{cost}(T') \leq \text{cost}(T)$

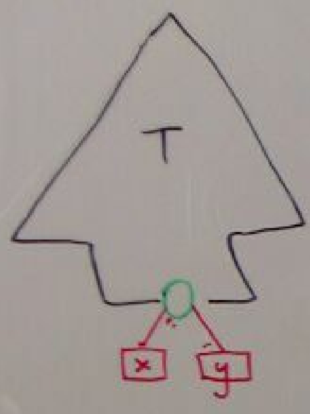
$\Rightarrow \text{cost}(T') = \text{cost}(T)$

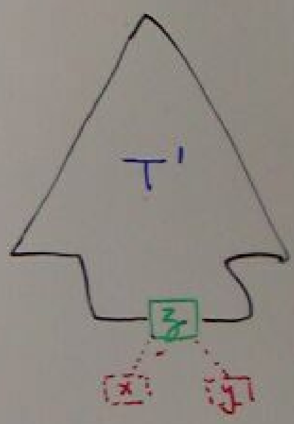
2. optimal substructure property

$$P = \{ A = \{ (x, f_x), (y, f_y), \cdots (a_k, f_k) \} \}$$

$\downarrow$ after the 1ST GREEDY CHOICE

$$P' = \{ A' = \{ (z, f_x + f_y), \cdots (a_k, f_k) \} \}$$





T is an optimal tree for P.

T' ??? an optimal tree for P'