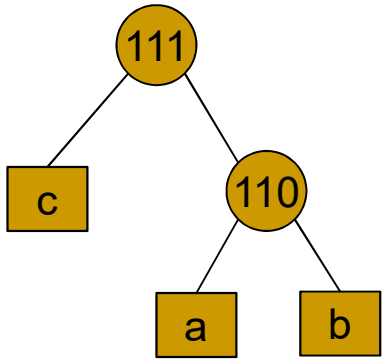
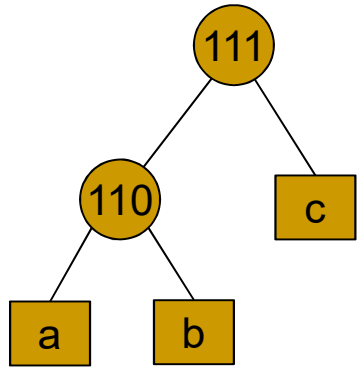
Uniqueness of Huffman Tree

- Is Huffman tree unique?
  - No. We can arbitrarily choose to make a node right or left child of the new parent.
  - E.g., let a, b, and c appear 100, 10, and 1 times, respectively. Both of two trees shown below are valid Huffman trees. Nonetheless, the external path lengths for two trees are the same.





Optimality of Huffman Tree

- (Theorem) Huffman coding tree gives the minimum external path weight
- (Proof)
  - (Lemma 1) An optimal tree should contain two characters with least frequency as sibling nodes whose depth is at least as deep as any other leaf nodes in the tree.
    - Proof by contradiction: Assume the conclusion of the Lemma is false. Let L be the set of two least frequency nodes. Let $y \in L$ is not the deepest, and Let $z \notin L$ be the deepest. Swapping y and z decreases the external path weight, thus contradiction.
  - Proof is by induction on n, the number of letters
  - Base case: for n = 2, Huffman tree is optimal.
  - Induction hypothesis: for n-1 letters Huffman tree is optimal.

Optimality of Huffman Tree

EPL: external path length

- (Proof: continued)
  - Induction step (proof by contradiction)
    1. Let T be a Huffman tree from n letters.
    2. Let x and y be letters with least frequencies in T. From Lemma1, x and y should be siblings whose depth is the deepest in T.
    3. Let v be the parent of x and y in T. Let T' be a tree by replacing v with a leaf node v' whose weight is w(x)+w(y). Note that T' is also a Huffman tree with n-1 letters. From I.H., T' is optimal.
    4. Assume T is not optimal; i.e., let Z be an optimal tree whose EPL is smaller than T. From Lemma1, we know that Z contains x and y as the deepest siblings. Create the tree Z' by replacing the parent of x and y with a new node v'' whose weight is w(x)+w(y). Then, EPL(Z) = EPL(Z') + w(x) + w(y) >= EPL(T') + w(x) + w(y) = EPL(T) , where the inequality use the result of step 3. This is a contradiction to the assumption that T is not optimal. Thus, T should be optimal.