Warm-up

Problem 1. Construct the Huffman tree of the following phrase: "data structure"

Problem 2. During the lecture we saw that the Fractional Knapsack algorithm's correctness depends on which greedy choice we make. We saw that there are instances where always picking the "highest benefit" or always picking the "smallest weight" doesn't give the optimal solution.

For each of these two strategies, give an infinite class of instances where the algorithm gives the optimal solution, thus showing that *you can't show correctness of an algorithm by using a finite number of example instances*.

Problem solving

Problem 3. Given vectors $a, b \in R^n$, consider the problem of finding a permutation a' and b' of a and b respectively minimizing their total difference

$$\sum_{1 \le i \le n} |a'_i - b'_i|.$$

Prove or disprove the correctness (i.e., that it always returns the optimal solution) of the following algorithm that greedily tries to pair up similar coordinates.

```
1: function GREEDY-MATCHING-V1(a, b)
2:     A ← multiset of values in a
3:     B ← multiset of values in b
4:     a' ← new empty list
5:     b' ← new empty list
6:     while A is not empty do
7:         x, y ← a pair in (A, B) minimizing |x − y|
8:         Append x to a'
9:         Append y to b'
10:        Remove x from A and y from B
11:    return (a', b')
```

Problem 4. Given vectors $a, b \in R^n$, consider the problem of finding a permutation a' and b' of a and b respectively that minimizes

$$\sum_{1 \le i \le n} |a'_i - b'_i|.$$

Prove or disprove the correctness (i.e., that it always returns the optimal solution) of the following algorithm that greedily tries to pair up similar coordinates.

```
1: function GREEDY-MATCHING-V2(a, b)
2:     A ← multiset of values in a
3:     B ← multiset of values in b
4:     a' ← new empty list
```

```
    b' ← new empty list
    while A is not empty do
        x ← smallest in A
        y ← smallest in B
        Append x to a'
        Append y to b'
        Remove x from A
        Remove y from B
    return (a', b')
```

Problem 5. Suppose we are to construct a Huffman tree for a string over an alphabet $C = c_1, c_2, \ldots, c_k$ with frequencies $f_i = 2^i$. Prove that every internal node in T has an external-node child.

Problem 6. Design a greedy algorithm for the following problem (see Figure 1): Given a set of n points $\{x_1, ..., x_n\}$ on the real line, determine the smallest set of unit-length intervals that contains all points.

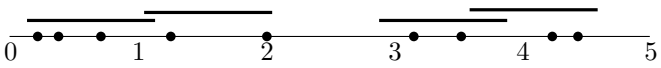


Figure 1: Covering points with unit intervals.

Problem 7. Suppose we are to schedule print jobs on a printer. Each job j has an associated weight $w_j > 0$ (representing how important the job is) and a processing time t_j (representing how long the job takes). A schedule σ is an ordering of the jobs that tells the printer in which order to process the jobs. Let C_j^σ be the completion time of job j under the schedule σ.

Design a greedy algorithm that computes a schedule σ minimizing the sum of weighted completion times, that is, minimizing $\sum_j w_j C_j^\sigma$.

假设不计算 space ——↘；但在实际做的时候要算上

Problem 1. Construct the Huffman tree of the following phrase: "data structure"

My Wrong answer

✓a:2 ✓r:2
✓c:1 ✓u:2
✓d:1
✓e:1
t:3
✓s:1

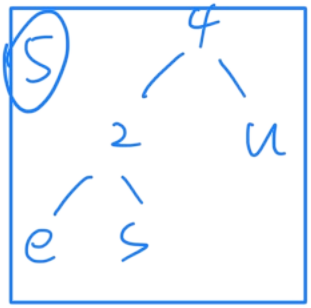
① c d e s r u a t
1 1 1 1 2 2 2 3

② 2 e s r u a t

③ 2 2 r u a t

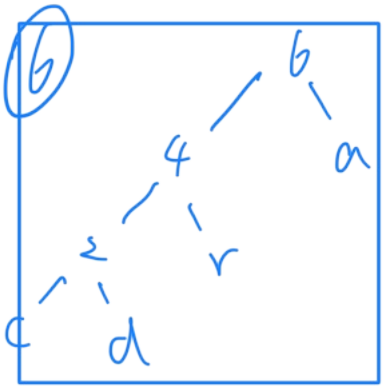
④ 4 2 u a t

⑤ 4 4 a t
2 3

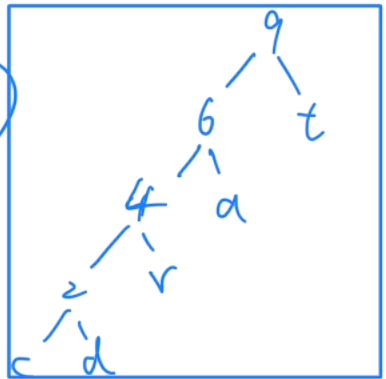


depth_T(a) < depth_T(b)
⇒ f(a) ≥ f(b)

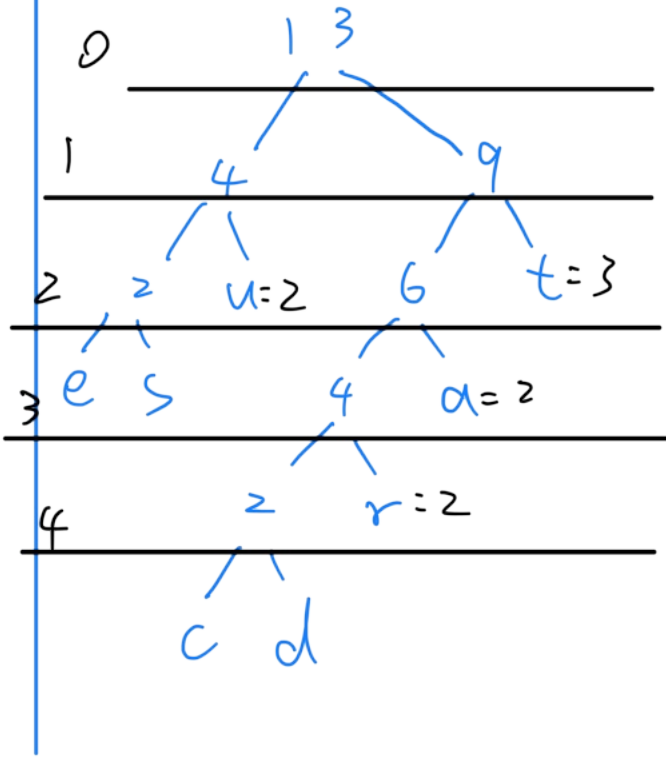
X ⑥ 6 4 t
3



X ⑦ 9 4



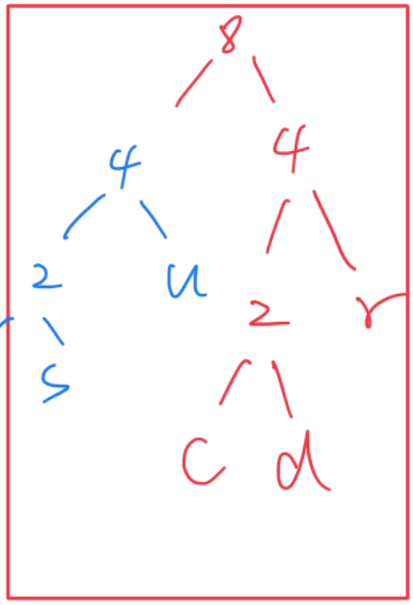
X ⑧



6

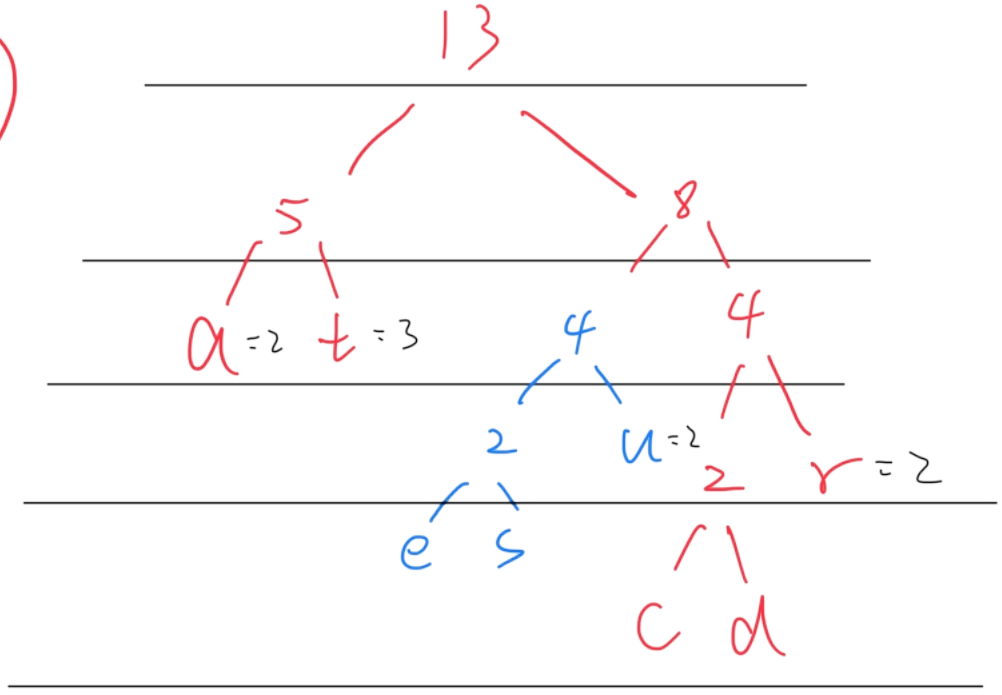
4 4 5

7



85

8



Problem 2. During the lecture we saw that the Fractional Knapsack algorithm's correctness depends on which greedy choice we make. We saw that there are instances where always picking the "highest benefit" or always picking the "smallest weight" doesn't give the optimal solution.

For each of these two strategies, give an infinite class of instances where the algorithm gives the optimal solution, thus showing that *you can't show correctness of an algorithm by using a finite number of example instances.*

highest benefit:

(1) pick 5 items from N items, with the highest total price. Items cannot break into pieces. Items have same weight

(2) if sack is big enough to hold everything

smallest weight:

(1) pick 5 items from N items, with the smallest total weight. Items cannot break into pieces, items have same benefit

(2) if sack is big enough to hold everything

(3) when benefit = weight for each item

Problem 3. Given vectors $a, b \in R^n$, consider the problem of finding a permutation a' and b' of a and b respectively minimizing their total difference

$$\sum_{1 \le i \le n} |a'_i - b'_i|.$$

Prove or disprove the correctness (i.e., that it always returns the optimal solution) of the following algorithm that greedily tries to pair up similar coordinates.

```
1:  function GREEDY-MATCHING-V1(a, b)
2:      A ← multiset of values in a
3:      B ← multiset of values in b
4:      a' ← new empty list
5:      b' ← new empty list
6:      while A is not empty do
7:          x, y ← a pair in (A, B) minimizing |x − y|
8:          Append x to a'
9:          Append y to b'
10:         Remove x from A and y from B
11:     return (a', b')
```

Incorrect:

$A = \{1, 4\}$, $B = \{3, 6\}$ Counter example

$\begin{cases} |4-3| = 1 \\ |1-6| = 5 \end{cases} \Rightarrow 6$

$\begin{cases} |1-3| = 2 \\ |4-6| = 2 \end{cases} \Rightarrow 4$

Prove: $|x_1 - y_1| + |x_2 - y_2| + \cdots |x_n - y_n|$?

Problem 4. Given vectors $a, b \in R^n$, consider the problem of finding a permutation a' and b' of a and b respectively that minimizes

$$\sum_{1 \le i \le n} |a'_i - b'_i|.$$

same Goal as Q3

Prove or disprove the correctness (i.e., that it always returns the optimal solution) of the following algorithm that greedily tries to pair up similar coordinates.

```
1:  function GREEDY-MATCHING-V2(a, b)
2:      A ← multiset of values in a
3:      B ← multiset of values in b
4:      a' ← new empty list
```



```
5:      b' ← new empty list
6:      while A is not empty do
7:          x ← smallest in A
8:          y ← smallest in B
9:          Append x to a'
10:         Append y to b'
11:         Remove x from A
12:         Remove y from B
13:     return (a', b')
```

correct

?

- Both A and B is sorting, since we pick least element in both sets, say A_i, B_i

$$|A_i - B_i| + |A_{i+1} - B_{i+1}| + \cdots \le |A_i - B_j| + |A_{i+1} - B_j| + \cdots$$

Say $\begin{cases} A_i + y = A_j & A_{i+1} \ge A_i \Rightarrow y \ge 0 \\ B_i + x = B_j & B_{i+1} \ge B_i \Rightarrow x \ge 0 \end{cases}$ | y, x are arbitrary positive number; B_j are random element in B

$$|A_i - B_i| + |A_{i+1} - B_i - x| + \cdots \le |A_i - B_i - x| + |A_i - B_i - x| + \cdots$$

不对，因为 X 互不相等

Problem 5. Suppose we are to construct a Huffman tree for a string over an alphabet $C = c_1, c_2, \ldots, c_k$ with frequencies $f_i = 2^i$. Prove that every internal node in T has an external-node child.

$f_1 = 2$

$f_2 = 4$

$f_3 = 8$

$f_4 = 16$

- Since we always choose the least two "nodes" from PQ
- and $f_i + f_{i+1} \leq f_{i+2}$
- Hence we have

Problem 6. Design a greedy algorithm for the following problem (see Figure 1): Given a set of n points $\{x_1, ..., x_n\}$ on the real line, determine the smallest set of unit-length intervals that contains all points.

length = 1

Figure 1: Covering points with unit intervals.

① 从最左 point 开始，用一个 interval

② 找到 interval 外的下一个 point 再用 interval

③ repeat ②

Problem 7. Suppose we are to schedule print jobs on a printer. Each job j has an associated weight $w_j > 0$ (representing how important the job is) and a processing time t_j (representing how long the job takes). A schedule σ is an ordering of the jobs that tells the printer in which order to process the jobs. Let C_j^σ be the completion time of job j under the schedule σ.

Design a greedy algorithm that computes a schedule σ minimizing the sum of weighted completion times, that is, minimizing $\sum_j w_j C_j^\sigma$.

(1) $\frac{w_j}{t_j}$

(2) pick job with $\frac{w_j}{t_j}$ in desc order