The Greedy Method

- The greedy method is a general algorithm design paradigm, built on the following elements:
  - configurations: different choices, collections, or values to find
  - objective function: a score assigned to configurations, which we want to either maximize or minimize

- It works best when applied to problems with the greedy-choice property:
  - a globally-optimal solution can always be found by a series of local improvements from a starting configuration.
The Greedy Method

- The sequence of choices starts from some well-understood starting configuration, and then iteratively makes the decision that is best from all of those that are currently possible, in terms of improving the objective function.

The Standard Knapsack Problem

- Given \( n \) items of weights \( s_1, s_2, \ldots, s_n \) and values \( v_1, v_2, \ldots, v_n \) and weight \( C \), the knapsack capacity, the objective is to find integers \( x_1, x_2, \ldots, x_n \) in \( \{ 0, 1 \} \) that maximize the sum:

\[
\sum_{i=1}^{n} x_i v_i
\]

subject to the constraint:

\[
\sum_{i=1}^{n} x_i s_i \leq C
\]
The Standard Knapsack Problem

- Example:
- Optimal Solution: \( \{ 3, 4 \} \) has value 40.
- Greedy: repeatedly add item with maximum \( \frac{v_i}{w_i} \).
- Greedy Solution:
  - \( \{ 5, 2, 1 \} \) achieves only value = 35 \( \Rightarrow \) greedy not optimal.

<table>
<thead>
<tr>
<th>#</th>
<th>value</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>18</td>
<td>5</td>
</tr>
<tr>
<td>4</td>
<td>22</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>28</td>
<td>7</td>
</tr>
</tbody>
</table>

The Fractional Knapsack Problem

- Given \( n \) items of weights \( s_1, s_2, \ldots, s_n \) and values \( v_1, v_2, \ldots, v_n \) and weight \( C \), the knapsack capacity, the objective is to find nonnegative real numbers \( x_1, x_2, \ldots, x_n \) between 0 and 1 that maximize the sum

\[
\sum_{i=1}^{n} x_i v_i
\]

subject to the constraint

\[
\sum_{i=1}^{n} x_i s_i \leq C
\]
Example

- Given: A set $S$ of $n$ items, with each item $i$ having
  - $b_i$ - value
  - $w_i$ - weight
- Goal: Choose items with maximum total value but with weight at most $W$.

<table>
<thead>
<tr>
<th>Items</th>
<th>Weight</th>
<th>Value</th>
<th>Unit value ($/ml)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4 ml</td>
<td>$12</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>8 ml</td>
<td>$32</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>2 ml</td>
<td>$40</td>
<td>20</td>
</tr>
<tr>
<td>4</td>
<td>6 ml</td>
<td>$30</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>1 ml</td>
<td>$50</td>
<td>50</td>
</tr>
</tbody>
</table>

“knapsack”

Solution:
- 100% of 5 (1 ml)
- 100% of 3 (2 ml)
- 100% of 4 (6 ml)
- 12.5% of 2 (1 ml)

The Fractional Knapsack Problem

- This problem can easily be solved using the following greedy strategy:
  - For each item compute $v_i = \frac{b_i}{w_i}$, the ratio of its value to its weight.
  - Sort the items by decreasing ratio, and fill the knapsack with as much as possible from the first item, then the second, and so forth.

- This problem reveals many of the characteristics of a greedy algorithm discussed above: The algorithm consists of a simple iterative procedure that selects an item which produces the largest immediate gain while maintaining feasibility.
The Fractional Knapsack Algorithm

- Greedy choice: Keep taking item with highest value (per unit weight)
  - Run time: $O(n \log n)$. Why?
- Correctness: Suppose there is a better solution
  - there is an item $i$ with higher value than a chosen item $j$, but $x_i < 1$, $x_j > 0$ and $v_i > v_j$
  - If we substitute some $i$ with $j$, we get a better solution
  - How much of $i$:
    $$\min\{ w_i(1-x_i), w_jx_j \}$$
  - Thus, there is no better solution than the greedy one

Algorithm $\text{fractionalKnapsack}(S, W)$

Input: set $S$ of items w/ values $b_i$ and weight $w_i$; max. weight $W$

Output: amount $x_i$ of each item $i$ to maximize value w/ weight at most $W$

for each item $i$ in $S$

1. $x_i \leftarrow 0$
2. $v_i \leftarrow b_i / w_i$ \{value\}
3. $w \leftarrow 0$ \{total weight\}
4. while $w < W$ and $|S| > 0$
5.   remove item $i$ w/ highest $v_i$
6.   $x_i \leftarrow \min\{ w_i, W - w \} / w_i$
7.   $w \leftarrow w + \min\{ w_i, W - w \}$

Task Scheduling

- Job $j$ starts at $s_j$ and finishes at $f_j$.
- Two jobs compatible if they don't overlap.
- Goal: find minimal number of machines to process all jobs.
Task Scheduling

- Given: a set $T$ of $n$ tasks, each having:
  - a start time, $s_i$
  - a finish time, $f_i$ (where $s_i < f_i$)
- Goal: Perform all the tasks using a minimum number of “machines.”

Example

- Given: a set $T$ of $n$ tasks, each having:
  - A start time, $s_i$
  - A finish time, $f_i$ (where $s_i < f_i$)
  - $[1,4], [1,3], [2,5], [3,7], [4,7], [6,9], [7,8]$ (ordered by start)
- Goal: Perform all tasks on min. number of machines
Task Scheduling

- Job $j$ starts at $s_j$ and finishes at $f_j$.
- Goal: find minimum number of machines to schedule all jobs so that no two occur at the same time on the same machine.

**Ex:** This schedule uses 4 machines to schedule 10 jobs.

```
Time
9 9:30 10 10:30 11 11:30 12 12:30 1 1:30 2 2:30 3 3:30 4 4:30
```

Task Scheduling

- Job $j$ starts at $s_j$ and finishes at $f_j$.
- Goal: find minimum number of machines to schedule all jobs so that no two occur at the same time on the same machine.

**Ex:** This schedule uses only 3.

```
Time
9 9:30 10 10:30 11 11:30 12 12:30 1 1:30 2 2:30 3 3:30 4 4:30
```
Task Scheduling: Greedy Algorithms

- Greedy template. Consider jobs in some natural order.
  Take each job provided it's compatible with the ones already taken.
  - [Earliest start time] Consider jobs in ascending order of $s_j$.
  - [Earliest finish time] Consider jobs in ascending order of $f_j$.
  - [Shortest interval] Consider jobs in ascending order of $f_j - s_j$.
  - [Fewest conflicts] For each job $j$, count the number of conflicting jobs $c_j$. Schedule in ascending order of $c_j$.

Task Scheduling Algorithm

- Greedy choice: consider tasks by their start time and use as few machines as possible with this order.
  - Run time: $O(n \log n)$. Why?
- Correctness: Suppose there is a better schedule.
  - We can use $k-1$ machines
  - The algorithm uses $k$
  - Let $i$ be first task scheduled on machine $k$
  - Task $i$ must conflict with $k-1$ other tasks currently running
  - But that means there is no non-conflicting schedule using $k-1$ machines

Algorithm taskSchedule($T$)

Input: set $T$ of tasks w/ start time $s_j$ and finish time $f_j$.
Output: non-conflicting schedule with minimum number of machines

$m \leftarrow 0$ \hspace{1cm} \text{(no. of machines)}

while $T$ is not empty
  remove task $i$ w/ smallest $s_j$
  if there's a machine $j$ for $i$ then
    schedule $i$ on machine $j$
  else
    $m \leftarrow m + 1$
  schedule $i$ on machine $m$
Task Scheduling II

- Given: a set \( T \) of \( n \) tasks, start time, \( s_i \), and finish time, \( f_i \) (where \( s_i < f_i \))
- Goal: Perform a maximum number of compatible jobs on a single machine.

Task Scheduling II: Greedy Algorithms

- Greedy template. Consider jobs in some natural order. Take each job provided it's compatible with the ones already taken.

Counterexamples:
- Earliest start time
- Shortest interval
- Fewest conflicts
Lists and Iterators

Task Scheduling II: Greedy Algorithm

- Greedy algorithm. Consider jobs in increasing order of finish time. Take each job provided it's compatible with the ones already taken.

  Sort jobs by finish times so that $f_1 \leq f_2 \leq \ldots \leq f_n$.
  $A \leftarrow \emptyset$
  for $j = 1$ to $n$ {
    if (job $j$ compatible with $A$)
      $A \leftarrow A \cup \{j\}$
  }
  return $A$

- Implementation: $O(n \log n)$.
  - Let job $j^*$ denote the job that was added last to $A$.
  - Job $j$ is compatible with $A$ if $s_j \geq f_{j^*}$.

Task Scheduling II: Analysis

- Theorem. The greedy algorithm is optimal.

- Proof. (by contradiction)
  - Assume greedy is not optimal, and let's see what happens.
  - Let $i_1, i_2, \ldots, i_k$ denote set of jobs selected by greedy.
  - Let $j_1, j_2, \ldots, j_m$ denote set of jobs in the optimal solution with $i_1 = j_1, i_2 = j_2, \ldots, i_k = j_k$ for the largest possible value of $r$.

Greedy: $i_1, i_2, i_3, i_4$... 
OPT: $j_1, j_2, j_3, j_4$... 

job $i_{r+1}$ finishes before $j_{r+1}$

why not replace job $j_{r+1}$ with job $i_{r+1}$?
Task Scheduling II: Analysis

- Theorem. The greedy algorithm is optimal.

- Proof. (by contradiction)
  - Assume greedy is not optimal, and let's see what happens.
  - Let $i_1, i_2, ..., i_k$ denote set of jobs selected by greedy.
  - Let $j_1, j_2, ..., j_m$ denote set of jobs in the optimal solution with
    $i_1 = j_1, i_2 = j_2, ..., i_r = j_r$ for the largest possible value of $r$.

Data Compression

- Given a string $X$, efficiently encode $X$ into a smaller string $Y$
  - Saves memory and/or bandwidth
- A good approach: **Huffman encoding**
  - Compute frequency $f(c)$ for each character $c$.
  - Encode high-frequency characters with short code words
  - No code word is a prefix for another code
  - Use an optimal encoding tree to determine the code words
Motivation

The motivations for data compression are obvious:

- reducing the space required to store files on disk or tape
- reducing the time to transmit large files.

Huffman savings are between 20% - 90%

Basic Idea:

Let the set of characters in the file be $C = \{c_1, c_2, \ldots, c_n\}$. Let also $f(c_i), 1 \leq i \leq n$, be the frequency of character $c_i$ in the file, i.e., the number of times $c_i$ appears in the file.

It uses a variable-length code table for encoding a source symbol (such as a character in a file) where the variable-length code table has been derived in a particular way based on the frequency of occurrence for each possible value of the source symbol.
**Example:**

Suppose you have a file with 100K characters.

For simplicity assume that there are only 6 distinct characters in the file from \(a\) through \(f\), with frequencies as indicated below.

We represent the file using a unique binary string for each character.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (in 100s)</td>
<td>45</td>
<td>13</td>
<td>12</td>
<td>16</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Fixed-length code-word</td>
<td>000</td>
<td>001</td>
<td>010</td>
<td>011</td>
<td>100</td>
<td>101</td>
</tr>
</tbody>
</table>

\[
\text{Space} = (45\times3 + 13\times3 + 12\times3 + 16\times3 + 9\times3 + 5\times3) \times 1000
\]

\[
= 300\text{K bits}
\]

Can we do better?? **YES!!**

By using **variable-length** codes instead of fixed-length codes.

Idea: Giving frequent characters short code-words, and infrequent characters long code-words.

i.e. The length of the encoded character is inversely proportional to that character's frequency.

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (in 1000s)</td>
<td>45</td>
<td>13</td>
<td>12</td>
<td>16</td>
<td>9</td>
<td>5</td>
</tr>
<tr>
<td>Fixed-length code-word</td>
<td>000</td>
<td>001</td>
<td>010</td>
<td>011</td>
<td>100</td>
<td>101</td>
</tr>
<tr>
<td>Variable-length code-word</td>
<td>0</td>
<td>101</td>
<td>100</td>
<td>111</td>
<td>1101</td>
<td>1100</td>
</tr>
</tbody>
</table>

\[
\text{Space} = (45\times1 + 13\times3 + 12\times3 + 16\times3 + 9\times4 + 5\times4) \times 1000
\]

\[
= 224\text{K bits} \quad (\text{Savings} = 25\%)
\]
**PREFIX CODES**: Codes in which no code-word is also a prefix of some other code-word.

("prefix-free codes" would have been a more appropriate name)

<table>
<thead>
<tr>
<th>Variable-length code-word</th>
<th>0</th>
<th>101</th>
<th>100</th>
<th>111</th>
<th>1101</th>
<th>1100</th>
</tr>
</thead>
</table>

It is very easy to encode and decode using prefix codes.

**No Ambiguity !!**

It is possible to show (although we won't do so here) that the optimal data compression achievable by a character code can always be achieved with a prefix code, so there is no loss of generality in restricting attention to prefix codes.

**Benefits of using Prefix Codes:**

Example:

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable-length code-word</td>
<td>0</td>
<td>101</td>
<td>100</td>
<td>111</td>
<td>1101</td>
<td>1100</td>
</tr>
</tbody>
</table>

Encoded as 1100 0 100 1101 = 110001001101

To decode, we have to decide where each code begins and ends, since they are no longer all the same length. But this is easy, since, no codes share a prefix. This means we need only scan the input string from left to right, and as soon as we recognize a code, we can print the corresponding character and start looking for the next code.

In the above case, the only code that begins with "1100.." is "f", so we can print "f" and start decoding "0100...", get "a", etc.
Benefits of using Prefix Codes:

Example:

To see why the no-common prefix property is essential, suppose that we encoded "e" with the shorter code "110"

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
</tr>
</thead>
<tbody>
<tr>
<td>code-word</td>
<td>0</td>
<td>101</td>
<td>100</td>
<td>111</td>
<td>1101</td>
<td>1100</td>
</tr>
<tr>
<td>code-word</td>
<td>0</td>
<td>101</td>
<td>100</td>
<td>111</td>
<td>110</td>
<td>1100</td>
</tr>
</tbody>
</table>

FACE = 11000100110

When we try to decode "1100"; we could not tell whether
1100 = "f"
or
1100 = 110 + 0 = "ea"

Representation:

The Huffman algorithm is represented as:

- binary tree
- each edge represents either 0 or 1
  - 0 means "go to the left child"
  - 1 means "go to the right child."
- each leaf corresponds to the sequence of 0s and 1s traversed from the root to reach it, i.e. a particular code.

Since no prefix is shared, all legal codes are at the leaves, and decoding a string means following edges, according to the sequence of 0s and 1s in the string, until a leaf is reached.
Labeling:
- leaf -> character it represents: frequency with which it appears in the text.
- internal node -> frequency with which all leaf nodes under it appear in the text (i.e. the sum of their frequencies).

Encoding Tree Summary

- A **code** is a mapping of each character of an alphabet to a binary code-word.
- A **prefix code** is a binary code such that no code-word is the prefix of another code-word.
- An **encoding tree** represents a prefix code:
  - Each external node stores a character.
  - The code word of a character is given by the path from the root to the external node storing the character (0 for a left child and 1 for a right child).
An optimal code for a file is always represented by a **proper binary tree**, in which every non-leaf node has two children.

The **fixed-length** code in our example is not optimal since its tree, is not a full binary tree: there are code-words beginning 10 . . . , but none beginning 11 ..

Since we can now restrict our attention to full binary trees, we can say that if C is the alphabet from which the characters are drawn, then the tree for an optimal prefix code has exactly |C| leaves, one for each letter of the alphabet, and exactly |C| - 1 internal nodes.

Given a tree T corresponding to a prefix code, it is a simple matter to compute the number of bits required to encode a file.

For each character c in the alphabet C,
- f(c) denote the frequency of c in the file
- d_T(c) denote the depth of c's leaf in the tree.

The number of bits required to encode a file is thus

$$B(T) = \sum_{c \in C} f(c) \cdot d_T(c)$$

which we define as the cost of the tree.
### Constructing a Huffman code

Huffman invented a greedy algorithm that constructs an optimal prefix code called a **Huffman code**. The algorithm builds the tree $T$ corresponding to the optimal code in a bottom-up manner.

*It begins with a set of $|C|$ leaves and performs a sequence of $|C| - 1$ "merging" operations to create the final tree.*

**Greedy Choice?**

The two smallest nodes are chosen at each step, and this local decision results in a globally optimal encoding tree.

In general, greedy algorithms use local minimal/maximal choices to produce a global minimum/maximum.

---

#### HUFFMAN($C$)

1. $n \leftarrow |C|$
2. $Q \leftarrow \text{BUILD-MIN-HEAP}(C)$ // using frequency $f[c]$ for $c \in C$
3. for $i \leftarrow 1$ to $n - 1$
   4. do $\text{ALLOCATE-NODE}(z)$ // create a new node $z$
   5. $\text{left}[z] \leftarrow x \leftarrow \text{EXTRACT-MIN}(Q)$
   6. $\text{right}[z] \leftarrow y \leftarrow \text{EXTRACT-MIN}(Q)$
   7. $f[z] \leftarrow f[x] + f[y]$ // frequency of $z$
   8. $\text{INSERT}(Q, z)$
9. return $\text{EXTRACT-MIN}(Q)$

$C$ is a set of $n$ characters: each character $c$ in $C$ is an object with a defined frequency $f[c]$.

A min-priority queue $Q$, keyed on $f$, is used to identify the two least-frequent objects to merge together and produce $z$. For $Q$, $z$ is a new character with frequency $f[z] = f[x] + f[y]$. For the tree, $z$ is a new internal node with children $x$ and $y$. 

---

*Lists and Iterators* 10/18/2016
The steps of Huffman's algorithm

1. **Heapify**
   - Initialize a min-priority queue $Q$ of frequency pairs.
   - Build a heap from the input frequencies.

2. **Extract Min**
   - Repeat for $n - 1$: Extract the two least-frequent nodes $x$ and $y$.
   - Combine $x$ and $y$ into a new node $z$.
   - Insert the new node back into the priority queue.

3. **Return**
   - Return the root of the tree, which represents the merged characters.

**Algorithm**

$$
\text{HUFFMAN}(C) = \begin{align*}
1 & n \leftarrow |C| \\
2 & Q \leftarrow \text{BUILD-MIN-HEAP}(C) \quad \text{// O(n)} \\
3 & \text{for } i \text{ from } 1 \text{ to } n - 1 \\
4 & \quad \text{do } \text{ALLOCATE-NODE}(z) \quad \text{// O(1)} \\
5 & \quad \quad \leftarrow x \leftarrow \text{EXTRACT-MIN}(Q) \quad \text{// O(log n)} \\
6 & \quad \quad \leftarrow y \leftarrow \text{EXTRACT-MIN}(Q) \quad \text{// O(log n)} \\
7 & \quad f[z] \leftarrow f[x] + f[y] \quad \text{// frequency of } z \quad \text{// O(1)} \\
8 & \quad \text{INSERT}(Q, z) \quad \text{// O(log n)} \\
9 & \text{return } \text{EXTRACT-MIN}(Q) \quad \text{// O(1)}
\end{align*}
$$

$C$ is a set of $n$ characters: each character $c$ in $C$ is an object with a defined frequency $f[c]$. A min-priority queue $Q$, keyed on $f$, is used to identify the two least-frequent objects to merge together and produce $z$. For $Q$, $z$ is a new character with frequency $f[z] = f[x] + f[y]$. For the tree, $z$ is a new internal node with children $x$ and $y$. 

---

Lists and Iterators 10/18/2016
Running Time Analysis

Assumes that $Q$ is implemented as a binary min-heap.

- For a set $C$ of $n$ characters, the initialization of $Q$ in line 2 can be performed in $O(n)$ time using the BUILD-MIN-HEAP procedure.

- The for loop in lines 3-8 is executed exactly $n - 1$ times. Each heap operation requires time $O(\log n)$. The loop contributes $= (n - 1) * O(\log n) = O(n \log n)$

Thus, the total running time of HUFFMAN on a set of $n$ characters $= O(n) + O(n \log n) = O(n \log n)$

Correctness of Huffman's algorithm

To prove that the greedy algorithm HUFFMAN is correct, we show that the problem of determining an optimal prefix code exhibits the greedy-choice and optimal-substructure properties.
The Greedy-Choice Property

**Lemma 1**: Let $C$ be an alphabet in which each character $c$ in $C$ has frequency $f[c]$. Let $x$ and $y$ be two characters in $C$ having the lowest frequencies. Then there exists an optimal prefix code for $C$ in which the code words for $x$ and $y$ have the same length and differ only in the last bit.

**Proof Idea of Lemma 1**: The idea of the proof is to take the tree $T$ representing an arbitrary optimal prefix code and modify it to make a tree representing another optimal prefix code such that the characters $x$ and $y$ appear as sibling leaves of maximum depth in the new tree. If we can do this, then their code words will have the same length and differ only in the last bit.

**Proof of Lemma 1**

Let $a$ and $b$ be two characters that are sibling leaves of maximum depth in $T$, and $x$ and $y$ are the two characters of the minimum frequency. Without loss in generality, assume that $f[x] < f[y] < f[a] < f[b]$. Then we must have $d_T(x) = d_T(y) = d_T(a) = d_T(b)$.

The difference in cost between $T$ and $T'$ is

$$B(T) - B(T') = \sum f(c) d_T(c) - \sum f(c) d_{T'}(c)$$

$$= f[x] d_T(x) + f[a] d_T(a) - f[y] d_T(y) - f[a] d_T(a)$$

$$= f[x] d_T(x) + f[a] d_T(a) - f[x] d_T(x) - f[a] d_T(x)$$

$$= (f[a] - f[x])( d_T(a) - d_T(x)) \geq 0 \quad \text{// } f[x] \text{ is min and } d_T(a) \text{ is max}$$

On the other hand, $B(T) - B(T') \leq 0$, because $B(T)$ is minimal. So $B(T) - B(T') = 0$ and $d_T(a) = d_T(x)$. 

Proof: Exchange the positions of $a$ and $x$ in $T$, to produce $T'$. 

**Proof**: Exchange the positions of $a$ and $x$ in $T$, to produce $T'$. 

Let $a$ and $b$ be two characters that are sibling leaves of maximum depth in $T$, and $x$ and $y$ are the two characters of the minimum frequency. Without loss in generality, assume that $f[x] < f[y] < f[a] < f[b]$. Then we must have $d_T(x) = d_T(y) = d_T(a) = d_T(b)$.

The difference in cost between $T$ and $T'$ is

$$B(T) - B(T') = \sum f(c) d_T(c) - \sum f(c) d_{T'}(c)$$

$$= f[x] d_T(x) + f[a] d_T(a) - f[y] d_T(y) - f[a] d_T(a)$$

$$= f[x] d_T(x) + f[a] d_T(a) - f[x] d_T(x) - f[a] d_T(x)$$

$$= (f[a] - f[x])( d_T(a) - d_T(x)) \geq 0 \quad \text{// } f[x] \text{ is min and } d_T(a) \text{ is max}$$

On the other hand, $B(T) - B(T') \leq 0$, because $B(T)$ is minimal. So $B(T) - B(T') = 0$ and $d_T(a) = d_T(x)$. 

Proof: Exchange the positions of $a$ and $x$ in $T$, to produce $T'$. 

Let $a$ and $b$ be two characters that are sibling leaves of maximum depth in $T$, and $x$ and $y$ are the two characters of the minimum frequency. Without loss in generality, assume that $f[x] < f[y] < f[a] < f[b]$. Then we must have $d_T(x) = d_T(y) = d_T(a) = d_T(b)$.

The difference in cost between $T$ and $T'$ is

$$B(T) - B(T') = \sum f(c) d_T(c) - \sum f(c) d_{T'}(c)$$

$$= f[x] d_T(x) + f[a] d_T(a) - f[y] d_T(y) - f[a] d_T(a)$$

$$= f[x] d_T(x) + f[a] d_T(a) - f[x] d_T(x) - f[a] d_T(x)$$

$$= (f[a] - f[x])( d_T(a) - d_T(x)) \geq 0 \quad \text{// } f[x] \text{ is min and } d_T(a) \text{ is max}$$

On the other hand, $B(T) - B(T') \leq 0$, because $B(T)$ is minimal. So $B(T) - B(T') = 0$ and $d_T(a) = d_T(x)$. 

Proof: Exchange the positions of $a$ and $x$ in $T$, to produce $T'$. 

Let $a$ and $b$ be two characters that are sibling leaves of maximum depth in $T$, and $x$ and $y$ are the two characters of the minimum frequency. Without loss in generality, assume that $f[x] < f[y] < f[a] < f[b]$. Then we must have $d_T(x) = d_T(y) = d_T(a) = d_T(b)$.

The difference in cost between $T$ and $T'$ is

$$B(T) - B(T') = \sum f(c) d_T(c) - \sum f(c) d_{T'}(c)$$

$$= f[x] d_T(x) + f[a] d_T(a) - f[y] d_T(y) - f[a] d_T(a)$$

$$= f[x] d_T(x) + f[a] d_T(a) - f[x] d_T(x) - f[a] d_T(x)$$

$$= (f[a] - f[x])( d_T(a) - d_T(x)) \geq 0 \quad \text{// } f[x] \text{ is min and } d_T(a) \text{ is max}$$

On the other hand, $B(T) - B(T') \leq 0$, because $B(T)$ is minimal. So $B(T) - B(T') = 0$ and $d_T(a) = d_T(x)$. 

Proof: Exchange the positions of $a$ and $x$ in $T$, to produce $T'$. 

Let $a$ and $b$ be two characters that are sibling leaves of maximum depth in $T$, and $x$ and $y$ are the two characters of the minimum frequency. Without loss in generality, assume that $f[x] < f[y] < f[a] < f[b]$. Then we must have $d_T(x) = d_T(y) = d_T(a) = d_T(b)$.

The difference in cost between $T$ and $T'$ is

$$B(T) - B(T') = \sum f(c) d_T(c) - \sum f(c) d_{T'}(c)$$

$$= f[x] d_T(x) + f[a] d_T(a) - f[y] d_T(y) - f[a] d_T(a)$$

$$= f[x] d_T(x) + f[a] d_T(a) - f[x] d_T(x) - f[a] d_T(x)$$

$$= (f[a] - f[x])( d_T(a) - d_T(x)) \geq 0 \quad \text{// } f[x] \text{ is min and } d_T(a) \text{ is max}$$

On the other hand, $B(T) - B(T') \leq 0$, because $B(T)$ is minimal. So $B(T) - B(T') = 0$ and $d_T(a) = d_T(x)$.
Similarly exchanging the positions of \( b \) and \( y \) in \( T' \), to produce \( T'' \) does not increase the cost,

\[
B(T') - B(T'') = 0.
\]

Since \( T \) is optimal, so is \( T' \) and \( T'' \).

Thus, \( T'' \) is an optimal tree in which \( x \) & \( y \) appear as sibling leaves of maximum depth from which **Lemma 1** follows.

**Lemma 2**: Let \( C \) be a given alphabet with frequency \( f[c] \) defined for each character \( c \in C \). Let \( x \) and \( y \) be two characters in \( C \) with minimum frequency. Let \( C' \) be the alphabet \( C \) with characters \( x,y \) removed and (new) character \( z \) added, so that \( C' = C - \{x,y\} \cup \{z\} \); define \( f \) for \( C' \) as for \( C \), except that \( f[z] = f[x] + f[y] \). Let \( T' \) be any tree representing an optimal prefix code for the alphabet \( C' \). Then the tree \( T \), obtained from \( T' \) by replacing the leaf node for \( z \) with an internal node having \( x \) and \( y \) as children, represents an optimal prefix code for the alphabet \( C \).

Proof:

We first express \( B(T) \) in terms of \( B(T') \)

\[
\forall c \in C - \{x,y\} \text{ we have } d_T(c) = d_{T'}(c) \text{, and hence } f[c]d_T(c) = f[c]d_{T'}(c).
\]
**Claim:** If $T'$ is optimal, so is $T$.

Since $d_r(x) = d_r(y) = d_r(z) + 1$, we have

$$f[x]d_r(x) + f[y]d_r(y) = (f[x] + f[y])(d_r(z) + 1) = f(z)d_r(z) + (f[x] + f[y])$$

Or $f[x]d_r(x) + f[y]d_r(y) - f(z)d_r(z) = (f[x] + f[y])$

From which we conclude that

$$B(T) = B(T') + (f[x] + f[y]) \quad \text{or} \quad B(T') = B(T) - (f[x] - f[y])$$

**Proof of Claim by contradiction**

Suppose that $T$ does not represent an optimal prefix code for $C$. Then there exists a tree $Opt$ such that $B(Opt) < B(T)$.

Without loss in generality (by Lemma 1) Opt has $x$ & $y$ as siblings. Let $T''$ be the tree $Opt$ with the common parent of $x$ & $y$ replaced by a leaf $z$ with frequency $f[z] = f[x] + f[y]$.

Then, $B(T'') = B(Opt) - (f[x] - f[y])$

$$< B(T) - (f[x] - f[y]) \quad \text{(assume B(Opt) < B(T))}$$

$$= B(T')$$

Yielding a contradiction to the assumption that $T'$ represents an optimal prefix code for $C'$. Thus, $T$ must represent an optimal prefix code for the alphabet $C$. 
**Theorem**: Huffman Code is optimal for n characters.

Proof: Induction on n.

Base case: $n = 2$ and one character is 0 and the other is 1. Optimal.

Inductive hypothesis: Huffman Code is optimal for $n - 1$ characters.

Induction case: We have $n$ characters in C. Let $x$ & $y$ be the least frequent characters in C. We replace $x$ & $y$ by $z$ with $f(z) = f(x) + f(y)$ to obtain $C' = C - \{ x, y \} U \{ z \}$. By induction hypothesis, we have optimal code for $C'$. Let code($z$) = c. Then let code of $x$ be c0 and code of $y$ be c1. By Lemma 2, the resulting code is optimal for $C$.

**Drawbacks**

The main disadvantage of Huffman’s method is that it makes two passes over the data:

- one pass to collect frequency counts of the letters in the message, followed by the construction of a Huffman tree and transmission of the tree to the receiver; and
- a second pass to encode and transmit the letters themselves, based on the static tree structure.

This causes delay when used for network communication, and in file compression applications the extra disk accesses can slow down the algorithm.

We need one-pass methods, in which letters are encoded “on the fly”.
Example

\( X = \text{abracadabra} \)

Frequencies

<table>
<thead>
<tr>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Huffman Code Example

String: a fast runner need never be afraid of the dark

<table>
<thead>
<tr>
<th>Character</th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>d</th>
<th>e</th>
<th>f</th>
<th>h</th>
<th>i</th>
<th>k</th>
<th>n</th>
<th>o</th>
<th>r</th>
<th>s</th>
<th>t</th>
<th>u</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency</td>
<td>9</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>7</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
Summary

- Given a string $X$, Huffman’s algorithm constructs a prefix code that minimizes the weight of the encoding of $X$.
- It runs in time $O(m + n \log n)$, where $m$ is the weight of $X$ and $n$ is the number of distinct characters of $X$.
- A heap-based priority queue is used as an auxiliary structure.

How to Show a Greedy Method is Optimal?

- In general, a greedy method is simple to describe, efficient to run, but difficult to prove.
- To show a greedy method is not optimal, we need to find a counterexample.
- To show a greedy method is indeed optimal, we use the following proof strategy:
  - Suppose $S$ is the solution found by the greedy method and $Opt$ is an optimal solution that differs from $S$ minimally. If $S = Opt$, we are done. If not, we “modify” $Opt$ to obtain another optimal solution $Opt'$, such that $Opt'$ has less difference than $Opt$ comparing to $S$. That’s a contradiction to the assumption that $Opt$ differs from $S$ minimally.