Readings

- Lecture Notes Chapter 15: Huffman Coding
- Lecture Notes Chapter 16: Graph Representations & BFS

Problems

Problem 1. What are pros and cons of Huffman Coding?

Problem 2. Construct an optimal Huffman coding for the following alphabet and frequency table:

<table>
<thead>
<tr>
<th>Character</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.4</td>
</tr>
<tr>
<td>B</td>
<td>0.3</td>
</tr>
<tr>
<td>C</td>
<td>0.15</td>
</tr>
<tr>
<td>D</td>
<td>0.1</td>
</tr>
<tr>
<td>E</td>
<td>0.05</td>
</tr>
</tbody>
</table>

What is the average encoded character length for the above encoding?

Problem 3. You have an alphabet with \( n > 2 \) letters and frequencies. You perform Huffman encoding on this alphabet, and notice that the character with the largest frequency is encoded by a 0. In this alphabet, symbol \( i \) occurs with probability \( p_i \); \( p_1 \geq p_2 \geq p_3 \geq ... \geq p_n \).

Given this alphabet and encoding, is it true or false that there exists an assignment of probabilities to \( p_1 \) through \( p_n \) such that \( p_1 < \frac{1}{3} \)?

Problem 4. Design an \( O(n \log k) \) time algorithm to find the \( k \)-th highest test score, where \( n \) is the total number of scores in the stream. Since PEFS provides minimal monetary resources, the 121-CIS staff have limited access to storage space and can only afford you \( O(k) \) space for your feature, where \( k \ll n \).

Additional Problems

Problem 5. Your task is to connect \( n \) ropes with a minimum cost. You are given \( n \) ropes of different lengths, and you need to connect these ropes into one rope. The cost to connect two ropes is equal to sum of their lengths. You need to connect the ropes with minimum cost.

For example if we are given 4 ropes of lengths 4, 3, 2 and 6. We can connect the ropes in following ways.

1. First connect ropes of lengths 4 and 3. Now we have three ropes of lengths 2, 7, 6.
2. Now connect ropes of lengths 2 and 7. Now we have two ropes of lengths 6 and 9.
3. Finally connect the two ropes and all ropes have connected.

Total cost for connecting all ropes is \( 7 + 9 + 15 = 29 \). Give an algorithm which always finds the minimum cost of attaching the rope.

Problem 6. Consider an indefinitely long stream of unsorted integers. We are interested in knowing the median (in sorted order) at any given time. How would we do this in an efficient manner?