CS 473: Algorithms, Fall 2018

Compression, Information and Entropy – Huffman's coding

Lecture 25

December 1, 2018

25.1: Huffman coding

Entropy...

Clicker question

1. I know what entropy of information is.
2. I know what entropy of information is - I use every day to wash my dishes.
3. I do **not** know what entropy of information is.
4. I do **not** know what entropy of information is, but I know it increases.

Huffman's trees...

Clicker question

1. I know what Huffman's trees are.
2. I know what Huffman's trees are, and I know how to build them.
3. I do **not** know what Hufmman's trees are.
4. I know what Huffman's trees are - I use them every day to dry my dishes.
5. I am going to take the fifth on this question.

Codes...

1. Σ: alphabet.
2. ***binary code***: assigns a string of 0s and 1s to each character in the alphabet.
3. each symbol in input $=$ a codeword over some other alphabet.
4. Useful for transmitting messages over a wire: only $0/1$.
5. receiver gets a binary stream of bits...
6. ... decode the message sent.
7. ***prefix code***: reading a prefix of the input binary string uniquely match it to a code word.
8. ... continuing to decipher the rest of the stream.
9. binary/prefix code is ***prefix-free*** if no code is a

Codes...

1. Σ: alphabet.
2. ***binary code***: assigns a string of 0s and 1s to each character in the alphabet.
3. each symbol in input = a codeword over some other alphabet.
4. Useful for transmitting messages over a wire: only $0/1$.
5. receiver gets a binary stream of bits...
6. ... decode the message sent.
7. ***prefix code***: reading a prefix of the input binary string uniquely match it to a code word.
8. ... continuing to decipher the rest of the stream.
9. binary/prefix code is ***prefix-free*** if no code is a

Codes...

1. Σ: alphabet.
2. ***binary code***: assigns a string of 0s and 1s to each character in the alphabet.
3. each symbol in input $=$ a codeword over some other alphabet.
4. Useful for transmitting messages over a wire: only $0/1$.
5. receiver gets a binary stream of bits...
6. ... decode the message sent.
7. ***prefix code***: reading a prefix of the input binary string uniquely match it to a code word.
8. ... continuing to decipher the rest of the stream.
9. binary/prefix code is ***prefix-free*** if no code is a

Codes...

1. Σ: alphabet.
2. ***binary code***: assigns a string of 0s and 1s to each character in the alphabet.
3. each symbol in input = a codeword over some other alphabet.
4. Useful for transmitting messages over a wire: only $0/1$.
5. receiver gets a binary stream of bits...
6. ... decode the message sent.
7. ***prefix code***: reading a prefix of the input binary string uniquely match it to a code word.
8. ... continuing to decipher the rest of the stream.
9. binary/prefix code is ***prefix-free*** if no code is a

Codes...

1. Σ: alphabet.
2. ***binary code***: assigns a string of 0s and 1s to each character in the alphabet.
3. each symbol in input $=$ a codeword over some other alphabet.
4. Useful for transmitting messages over a wire: only $\mathbf{0/1}$.
5. receiver gets a binary stream of bits...
6. ... decode the message sent.
7. ***prefix code***: reading a prefix of the input binary string uniquely match it to a code word.
8. ... continuing to decipher the rest of the stream.
9. binary/prefix code is ***prefix-free*** if no code is a

Codes...

1. $\boldsymbol{\Sigma}$: alphabet.
2. ***binary code***: assigns a string of $\mathbf{0}$s and $\mathbf{1}$s to each character in the alphabet.
3. each symbol in input = a codeword over some other alphabet.
4. Useful for transmitting messages over a wire: only $\mathbf{0/1}$.
5. receiver gets a binary stream of bits...
6. ... decode the message sent.
7. ***prefix code***: reading a prefix of the input binary string uniquely match it to a code word.
8. ... continuing to decipher the rest of the stream.
9. binary/prefix code is ***prefix-free*** if no code is a

Codes...

1. Σ: alphabet.
2. ***binary code***: assigns a string of 0s and 1s to each character in the alphabet.
3. each symbol in input $=$ a codeword over some other alphabet.
4. Useful for transmitting messages over a wire: only $0/1$.
5. receiver gets a binary stream of bits...
6. ... decode the message sent.
7. ***prefix code***: reading a prefix of the input binary string uniquely match it to a code word.
8. ... continuing to decipher the rest of the stream.
9. binary/prefix code is ***prefix-free*** if no code is a

Codes...

1. Σ: alphabet.
2. ***binary code***: assigns a string of 0s and 1s to each character in the alphabet.
3. each symbol in input = a codeword over some other alphabet.
4. Useful for transmitting messages over a wire: only $0/1$.
5. receiver gets a binary stream of bits...
6. ... decode the message sent.
7. ***prefix code***: reading a prefix of the input binary string uniquely match it to a code word.
8. ... continuing to decipher the rest of the stream.
9. binary/prefix code is ***prefix-free*** if no code is a

Codes...

1. Σ: alphabet.
2. ***binary code***: assigns a string of 0s and 1s to each character in the alphabet.
3. each symbol in input = a codeword over some other alphabet.
4. Useful for transmitting messages over a wire: only $0/1$.
5. receiver gets a binary stream of bits...
6. ... decode the message sent.
7. ***prefix code***: reading a prefix of the input binary string uniquely match it to a code word.
8. ... continuing to decipher the rest of the stream.
9. binary/prefix code is ***prefix-free*** if no code is a

Codes...

1. Σ: alphabet.
2. ***binary code***: assigns a string of 0s and 1s to each character in the alphabet.
3. each symbol in input = a codeword over some other alphabet.
4. Useful for transmitting messages over a wire: only $0/1$.
5. receiver gets a binary stream of bits...
6. ... decode the message sent.
7. ***prefix code***: reading a prefix of the input binary string uniquely match it to a code word.
8. ... continuing to decipher the rest of the stream.
9. binary/prefix code is ***prefix-free*** if no code is a

Codes...

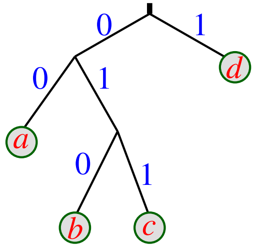
1. Morse code is binary+prefix code but **not** prefix-free.
2. ... code for `S` $(\cdot\cdot\cdot)$ includes the code for `E` $(\cdot)$ as a prefix.
3. Prefix codes are binary trees...
4. ...characters in leafs, code word is path from root.
5. prefix treestree!prefix tree or ***code trees***.
6. Decoding/encoding is easy.

Codes...

1. Morse code is binary+prefix code but **not** prefix-free.
2. ... code for S ($\cdot\cdot\cdot$) includes the code for E ($\cdot$) as a prefix.
3. Prefix codes are binary trees...
4. ...characters in leafs, code word is path from root.
5. prefix treestree!prefix tree or ***code trees***.
6. Decoding/encoding is easy.

Codes...

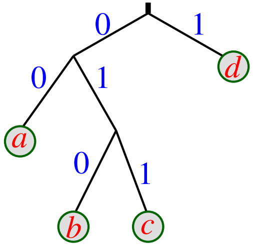
1. Morse code is binary+prefix code but **not** prefix-free.
2. ... code for `S` ($\cdot\cdot\cdot$) includes the code for `E` ($\cdot$) as a prefix.
3. Prefix codes are binary trees...



4. ...characters in leafs, code word is path from root.
5. prefix treestree!prefix tree or *code trees*

Codes...

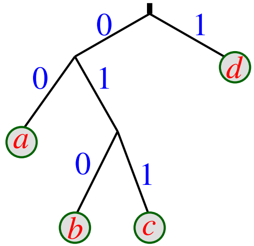
1. Morse code is binary+prefix code but **not** prefix-free.
2. ... code for `S` (· · ·) includes the code for `E` (·) as a prefix.
3. Prefix codes are binary trees...



4. ...characters in leafs, code word is path from root.
5. prefix treestree|prefix tree or ***code trees***

Codes...

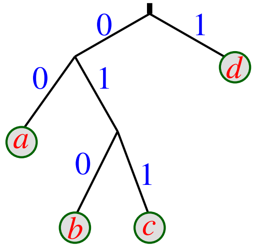
1. Morse code is binary+prefix code but **not** prefix-free.
2. ... code for `S` $(\cdot\cdot\cdot)$ includes the code for `E` $(\cdot)$ as a prefix.
3. Prefix codes are binary trees...



4. ...characters in leafs, code word is path from root.
5. prefix treestree|prefix tree or ***code trees***.

Codes...

1. Morse code is binary+prefix code but **not** prefix-free.
2. ... code for `S` ($\cdot\cdot\cdot$) includes the code for `E` ($\cdot$) as a prefix.
3. Prefix codes are binary trees...

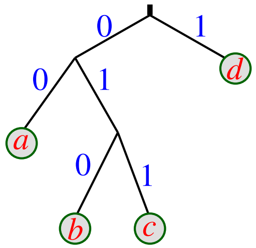


4. ...characters in leafs, code word is path from root.
5. prefix treestree|prefix tree or ***code trees***

What this string encodes?

Clicker question

Consider the string '0100001111001010101011111' using the prefix tree:



This encodes the string:

1. abcdabbbc.
2. cdbabdbaa.
3. bacbacdbdbddd.
4. [illegible]

Codes...

1. Encoding: given frequency table: $f[1 \ldots n]$.
2. $f[i]$: frequency of ith character.
3. $\mathrm{code}(i)$: binary string for ith character. $\mathrm{len}(s)$: length (in bits) of binary string s.
4. Compute tree $\mathcal{T}$ that minimizes

$$\mathrm{cost}(\mathcal{T}) = \sum_{i=1}^{n} f[i] * \mathrm{len}(\mathrm{code}(i)), \qquad (1)$$

Codes...

1. Encoding: given frequency table: $f[1 \ldots n]$.
2. $f[i]$: frequency of ith character.
3. $\mathrm{code}(i)$: binary string for ith character.
   $\mathrm{len}(s)$: length (in bits) of binary string s.
4. Compute tree $\mathcal{T}$ that minimizes

$$\mathrm{cost}(\mathcal{T}) = \sum_{i=1}^{n} f[i] * \mathrm{len}(\mathrm{code}(i)), \qquad (1)$$

Codes...

1. Encoding: given frequency table: $f[1 \dots n]$.
2. $f[i]$: frequency of ith character.
3. $\mathbf{code}(i)$: binary string for ith character. $\mathbf{len}(s)$: length (in bits) of binary string s.
4. Compute tree $\mathcal{T}$ that minimizes

$$\mathrm{cost}(\mathcal{T}) = \sum_{i=1}^{n} f[i] * \mathrm{len}(\mathrm{code}(i)), \qquad (1)$$

Codes...

1. Encoding: given frequency table: $f[1 \ldots n]$.
2. $f[i]$: frequency of ith character.
3. $\mathbf{code}(i)$: binary string for ith character. $\mathbf{len}(s)$: length (in bits) of binary string s.
4. Compute tree $\mathcal{T}$ that minimizes

$$\mathbf{cost}(\mathcal{T}) = \sum_{i=1}^{n} f[i] * \mathbf{len}(\mathbf{code}(i)), \qquad (1)$$

Frequency table for...

“A tale of two cities” by Dickens

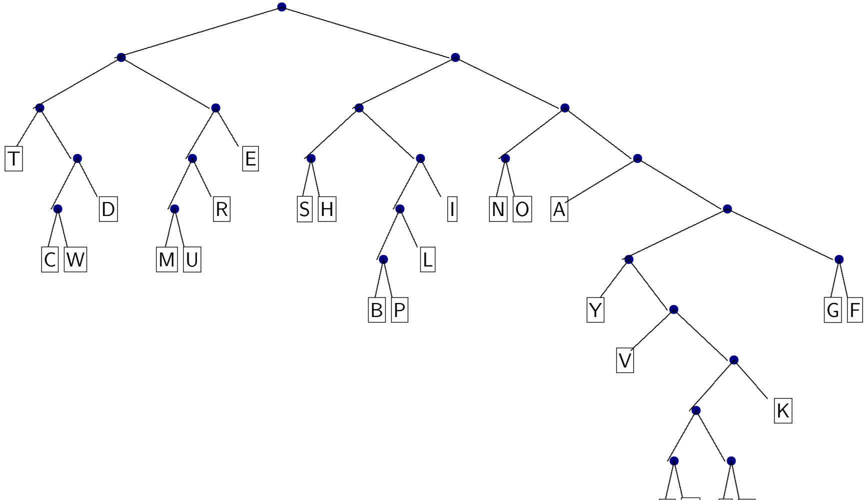
\n	16,492	‘1’	61	‘C’	13,896
’ ’	130,376	‘2’	10	‘D’	28,041
‘!’	955	‘3’	12	‘E’	74,809
‘”’	5,681	‘4’	10	‘F’	13,559
‘$’	2	‘5’	14	‘G’	12,530
‘%’	1	‘6’	11	‘H’	38,961
‘’’	1,174	‘7’	13	‘I’	41,005
‘(’	151	‘8’	13	‘J’	710
‘)’	151	‘9’	14	‘K’	4,782
‘*’	70	‘:’	267	‘L’	22,030
‘,’	13,276	‘;’	1,108	‘M’	15,298
‘–’	2,430	‘?’	913	‘N’	42,380
‘.’	6,769	‘A’	48,165	‘O’	46,499

Computed prefix codes...

char	frequency	code
‘A’	48165	1110
‘B’	8414	101000
‘C’	13896	00100
‘D’	28041	0011
‘E’	74809	011
‘F’	13559	111111
‘G’	12530	111110
‘H’	38961	1001
‘I’	41005	1011
‘J’	710	1111011010
‘K’	4782	11110111
‘L’	22030	10101
‘M’	15298	01000

The Huffman tree generating the code

Build only on A-Z for clarity.



Merging prefix codes

Clicker question

Π_1: Binary prefix code for alphabet Σ_1.
Π_2: Binary prefix code for alphabet Σ_2.
$\Sigma_1 \cap \Sigma_2 = \emptyset$.
One can get a prefix code for $\Sigma_1 \cup \Sigma_2$ such that:

1. New code word length increased by $O(\lg |\Sigma_1| + \lg |\Sigma_2|)$.
2. New code word is exactly one bit longer than before.
3. New code word can be arbitrarily longer than before.
4. New code word is same length as before.
5. There is no way to combine to prefix-free codes – have to rebuild from scratch.

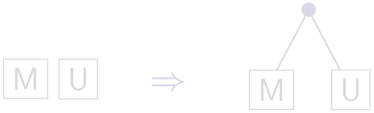
Merging code trees?

Clicker question

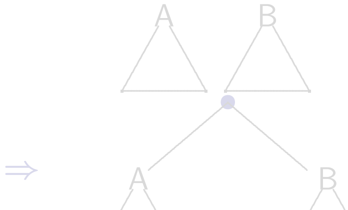
Given two code trees $\mathcal{T}_1$ and $\mathcal{T}_2$ for disjoint alphabets Σ_1 and Σ_2, one can get a prefix tree for $\Sigma_1 \cup \Sigma_2$ by:

1. Hang $\mathcal{T}_1$ from lowest leaf of $\mathcal{T}_2$.
2. Hang $\mathcal{T}_1$ from highest leaf of $\mathcal{T}_2$ (or vice versa).
3. Create a new root, and hang both trees on new root.
4. Insert all the root to leaf paths in $\mathcal{T}_1$ into $\mathcal{T}_2$ (or vice versa).
5. None of the above.

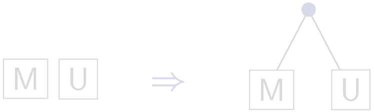
Mergeablity of code trees

1. two trees for some disjoint parts of the alphabet...
2. Merge into larger tree by creating a new node and hanging the trees from this common node.
3. 

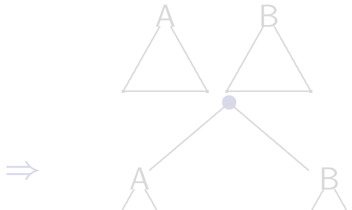
4. ...put together two subtrees.



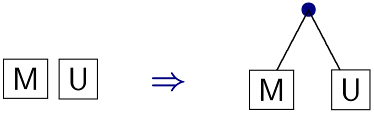
Mergeablity of code trees

1. two trees for some disjoint parts of the alphabet...
2. Merge into larger tree by creating a new node and hanging the trees from this common node.
3. 

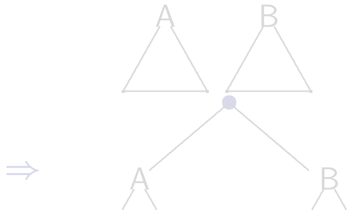
4. ...put together two subtrees.



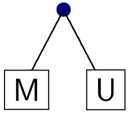
Mergeablity of code trees

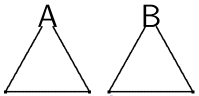
1. two trees for some disjoint parts of the alphabet...
2. Merge into larger tree by creating a new node and hanging the trees from this common node.
3. 

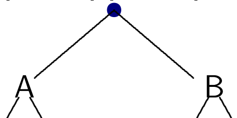
4. ...put together two subtrees.



Mergeablity of code trees

1. two trees for some disjoint parts of the alphabet...
2. Merge into larger tree by creating a new node and hanging the trees from this common node.
3. $\boxed{M}\ \boxed{U} \Rightarrow$ 
4. ...put together two subtrees.



$\Rightarrow$ 

Building optimal prefix code trees

1. take two least frequent characters in frequency table...
2. ... merge them into a tree, and put the root of merged tree back into table.
3. ...instead of the two old trees.
4. Algorithm stops when there is a single tree.
5. Intuition: infrequent characters participate in a large number of merges. Long code words.
6. Algorithm is due to David Huffman (1952).
7. Resulting code is best one can do.
8. ***Huffman coding***: building block used by numerous other compression algorithms

Building optimal prefix code trees

1. take two least frequent characters in frequency table...
2. ... merge them into a tree, and put the root of merged tree back into table.
3. ...instead of the two old trees.
4. Algorithm stops when there is a single tree.
5. Intuition: infrequent characters participate in a large number of merges. Long code words.
6. Algorithm is due to David Huffman (1952).
7. Resulting code is best one can do.
8. ***Huffman coding***: building block used by numerous other compression algorithms

Building optimal prefix code trees

1. take two least frequent characters in frequency table...
2. ... merge them into a tree, and put the root of merged tree back into table.
3. ...instead of the two old trees.
4. Algorithm stops when there is a single tree.
5. Intuition: infrequent characters participate in a large number of merges. Long code words.
6. Algorithm is due to David Huffman (1952).
7. Resulting code is best one can do.
8. ***Huffman coding***: building block used by numerous other compression algorithms

Building optimal prefix code trees

1. take two least frequent characters in frequency table...
2. ... merge them into a tree, and put the root of merged tree back into table.
3. ...instead of the two old trees.
4. Algorithm stops when there is a single tree.
5. Intuition: infrequent characters participate in a large number of merges. Long code words.
6. Algorithm is due to David Huffman (1952).
7. Resulting code is best one can do.
8. ***Huffman coding***: building block used by numerous other compression algorithms

Building optimal prefix code trees

1. take two least frequent characters in frequency table...
2. ... merge them into a tree, and put the root of merged tree back into table.
3. ...instead of the two old trees.
4. Algorithm stops when there is a single tree.
5. Intuition: infrequent characters participate in a large number of merges. Long code words.
6. Algorithm is due to David Huffman (1952).
7. Resulting code is best one can do.
8. ***Huffman coding***: building block used by numerous other compression algorithms

Building optimal prefix code trees

1. take two least frequent characters in frequency table...
2. ... merge them into a tree, and put the root of merged tree back into table.
3. ...instead of the two old trees.
4. Algorithm stops when there is a single tree.
5. Intuition: infrequent characters participate in a large number of merges. Long code words.
6. Algorithm is due to David Huffman (1952).
7. Resulting code is best one can do.
8. ***Huffman coding***: building block used by numerous other compression algorithms

Building optimal prefix code trees

1. take two least frequent characters in frequency table...
2. ... merge them into a tree, and put the root of merged tree back into table.
3. ...instead of the two old trees.
4. Algorithm stops when there is a single tree.
5. Intuition: infrequent characters participate in a large number of merges. Long code words.
6. Algorithm is due to David Huffman (1952).
7. Resulting code is best one can do.
8. ***Huffman coding***: building block used by numerous other compression algorithms

Building optimal prefix code trees

1. take two least frequent characters in frequency table...
2. ... merge them into a tree, and put the root of merged tree back into table.
3. ...instead of the two old trees.
4. Algorithm stops when there is a single tree.
5. Intuition: infrequent characters participate in a large number of merges. Long code words.
6. Algorithm is due to David Huffman (1952).
7. Resulting code is best one can do.
8. ***Huffman coding***: building block used by numerous other compression algorithms

Lemma: lowest leafs are siblings...

Lemma

1. $\mathcal{T}$: *optimal code tree (prefix free!).*
2. *Then $\mathcal{T}$ is a full binary tree.*
3. *... every node of $\mathcal{T}$ has either 0 or 2 children.*
4. *If height of $\mathcal{T}$ is d, then there are leafs nodes of height d that are sibling.*

Lemma: lowest leafs are siblings...

Lemma

1. $\mathcal{T}$: *optimal code tree (prefix free!).*
2. *Then* $\mathcal{T}$ *is a full binary tree.*
3. *... every node of* $\mathcal{T}$ *has either* 0 *or* 2 *children.*
4. *If height of* $\mathcal{T}$ *is* d*, then there are leafs nodes of height* d *that are sibling.*

Lemma: lowest leafs are siblings...

Lemma

1. $\mathfrak{T}$: *optimal code tree (prefix free!).*
2. *Then* $\mathfrak{T}$ *is a full binary tree.*
3. *... every node of* $\mathfrak{T}$ *has either* $\mathbf{0}$ *or* $\mathbf{2}$ *children.*
4. *If height of* $\mathfrak{T}$ *is* $\boldsymbol{d}$*, then there are leafs nodes of height* $\boldsymbol{d}$ *that are sibling.*

Lemma: lowest leafs are siblings...

Lemma

1. $\mathfrak{T}$: *optimal code tree (prefix free!).*
2. *Then* $\mathfrak{T}$ *is a full binary tree.*
3. *... every node of* $\mathfrak{T}$ *has either* $\mathbf{0}$ *or* $\mathbf{2}$ *children.*
4. *If height of* $\mathfrak{T}$ *is* $\boldsymbol{d}$*, then there are leafs nodes of height* $\boldsymbol{d}$ *that are sibling.*

Proof...

1. If $\exists$ internal node $v \in \mathbf{V}(\mathcal{T})$ with single child...
   ...remove it.
2. New code tree is better compressor:
   $\mathbf{cost}(\mathcal{T}) = \sum_{i=1}^{n} f[i] * \mathbf{len}(\mathbf{code}(i))$.
3. u: leaf u with maximum depth d in $\mathcal{T}$. Consider parent $v = \overline{\mathrm{p}}(u)$.
4. $\implies$ v: has two children, both leafs

■

Proof...

1. If $\exists$ internal node $v \in \mathbf{V}(\mathcal{T})$ with single child...
   ...remove it.
2. New code tree is better compressor:
   $\mathbf{cost}(\mathcal{T}) = \sum_{i=1}^{n} f[i] * \mathbf{len}(\mathbf{code}(i))$.
3. u: leaf u with maximum depth d in $\mathcal{T}$. Consider parent $v = \overline{\mathrm{p}}(u)$.
4. $\implies$ v: has two children, both leafs

■

Proof...

1. If $\exists$ internal node $v \in \mathsf{V}(\mathcal{T})$ with single child...
   ...remove it.
2. New code tree is better compressor:
   $\mathbf{cost}(\mathcal{T}) = \sum_{i=1}^{n} f[i] * \mathbf{len}(\mathbf{code}(i))$.
3. u: leaf u with maximum depth d in $\mathcal{T}$. Consider parent $v = \overline{\mathrm{p}}(u)$.
4. $\implies$ v: has two children, both leafs

■

Proof...

1. If $\exists$ internal node $v \in \mathbf{V}(\mathcal{T})$ with single child...
   ...remove it.
2. New code tree is better compressor:
   $\mathbf{cost}(\mathcal{T}) = \sum_{i=1}^{n} f[i] * \mathbf{len}(\mathbf{code}(i))$.
3. u: leaf u with maximum depth d in $\mathcal{T}$. Consider parent $v = \overline{\mathrm{p}}(u)$.
4. $\implies$ v: has two children, both leafs

■

Proof...

1. If $\exists$ internal node $v \in \mathsf{V}(\mathcal{T})$ with single child...
   ...remove it.
2. New code tree is better compressor:
   $\mathbf{cost}(\mathcal{T}) = \sum_{i=1}^{n} f[i] * \mathbf{len}(\mathbf{code}(i))$.
3. u: leaf u with maximum depth d in $\mathcal{T}$. Consider parent $v = \overline{\mathrm{p}}(u)$.
4. $\implies$ v: has two children, both leafs

Proof...

1. If $\exists$ internal node $v \in \mathsf{V}(\mathcal{T})$ with single child...
   ...remove it.
2. New code tree is better compressor:
   $\mathbf{cost}(\mathcal{T}) = \sum_{i=1}^{n} f[i] * \mathbf{len}(\mathbf{code}(i))$.
3. u: leaf u with maximum depth d in $\mathcal{T}$. Consider parent $v = \overline{\mathrm{p}}(u)$.
4. $\implies$ v: has two children, both leafs

■

Frequency and depth

Clicker question

Alphabet $\Sigma = \{1, \ldots, n\}$.
$f[1, \ldots, n]$: frequencies.
$\mathcal{T}$: optimal prefix-free code tree for it
Assume that $f[i] > f[j]$, and let v_i and v_j be the two corresponding nodes in $\mathcal{T}$. Then it must be that:

1. $\mathbf{depth}(v_i) > \mathbf{depth}(v_j)$
2. $\mathbf{depth}(v_i) \geq \mathbf{depth}(v_j)$
3. $\mathbf{depth}(v_i) = \mathbf{depth}(v_j)$
4. $\mathbf{depth}(v_i) < \mathbf{depth}(v_j)$
5. $\mathbf{depth}(v_i) \leq \mathbf{depth}(v_j)$

Infrequent characters are stuck together...

Lemma

x, y: two least frequent characters (breaking ties arbitrarily).
$\exists$ optimal code tree in which x and y are siblings.

Proof...

1. Claim: $\exists$ optimal code s.t. x and y are siblings + deepest.
2. $\mathcal{T}$: optimal code tree with depth d.
3. By lemma... $\mathcal{T}$ has two leafs at depth d that are siblings,
4. If not x and y, but some other characters α and β.
5. $\mathcal{T}'$: swap x and α.
6. x depth inc by Δ, and depth of α decreases by Δ.
7. $\mathbf{cost}(\mathcal{T}') = \mathbf{cost}(\mathcal{T}) - \big(f[\alpha] - f[x]\big)\Delta$.
8. x: one of the two least frequent characters. ...but α is not.
9. $\implies f[\alpha] \geq f[x]$.
10. Swapping x and α does not increase cost.

Proof...

1. Claim: $\exists$ optimal code s.t. x and y are siblings + deepest.
2. $\mathcal{T}$: optimal code tree with depth d.
3. By lemma... $\mathcal{T}$ has two leafs at depth d that are siblings,
4. If not x and y, but some other characters α and β.
5. $\mathcal{T}'$: swap x and α.
6. x depth inc by Δ, and depth of α decreases by Δ.
7. $\mathrm{cost}(\mathcal{T}') = \mathrm{cost}(\mathcal{T}) - \big(f[\alpha] - f[x]\big)\Delta$.
8. x: one of the two least frequent characters.
   ...but α is not.
9. $\implies f[\alpha] \geq f[x]$.
10. Swapping x and α does not increase cost.

Proof...

1. Claim: $\exists$ optimal code s.t. x and y are siblings + deepest.
2. $\mathcal{T}$: optimal code tree with depth d.
3. By lemma... $\mathcal{T}$ has two leafs at depth d that are siblings,
4. If not x and y, but some other characters α and β.
5. $\mathcal{T}'$: swap x and α.
6. x depth inc by Δ, and depth of α decreases by Δ.
7. $\mathrm{cost}(\mathcal{T}') = \mathrm{cost}(\mathcal{T}) - \big(f[\alpha] - f[x]\big)\Delta$.
8. x: one of the two least frequent characters.
   ...but α is not.
9. $\implies f[\alpha] \geq f[x]$.
10. Swapping x and α does not increase cost.

Proof...

1. Claim: $\exists$ optimal code s.t. x and y are siblings + deepest.
2. $\mathcal{T}$: optimal code tree with depth d.
3. By lemma... $\mathcal{T}$ has two leafs at depth d that are siblings,
4. If not x and y, but some other characters α and β.
5. $\mathcal{T}'$: swap x and α.
6. x depth inc by Δ, and depth of α decreases by Δ.
7. $\mathrm{cost}(\mathcal{T}') = \mathrm{cost}(\mathcal{T}) - \big(f[\alpha] - f[x]\big)\Delta$.
8. x: one of the two least frequent characters.
   ...but α is not.
9. $\implies f[\alpha] \geq f[x]$.
10. Swapping x and α does not increase cost.
11. [illegible]

Proof...

1. Claim: $\exists$ optimal code s.t. x and y are siblings + deepest.
2. $\mathcal{T}$: optimal code tree with depth d.
3. By lemma... $\mathcal{T}$ has two leafs at depth d that are siblings,
4. If not x and y, but some other characters α and β.
5. $\mathcal{T}'$: swap x and α.
6. x depth inc by Δ, and depth of α decreases by Δ.
7. $\mathrm{cost}(\mathcal{T}') = \mathrm{cost}(\mathcal{T}) - \big(f[\alpha] - f[x]\big)\Delta$.
8. x: one of the two least frequent characters.
   ...but α is not.
9. $\implies f[\alpha] \geq f[x]$.
10. Swapping x and α does not increase cost.

Proof...

1. Claim: $\exists$ optimal code s.t. x and y are siblings $+$ deepest.
2. $\mathcal{T}$: optimal code tree with depth d.
3. By lemma... $\mathcal{T}$ has two leafs at depth d that are siblings,
4. If not x and y, but some other characters α and β.
5. $\mathcal{T}'$: swap x and α.
6. x depth inc by Δ, and depth of α decreases by Δ.
7. $\mathrm{cost}(\mathcal{T}') = \mathrm{cost}(\mathcal{T}) - \big(f[\alpha] - f[x]\big)\Delta$.
8. x: one of the two least frequent characters. ...but α is not.
9. $\implies f[\alpha] \geq f[x]$.
10. Swapping x and α does not increase cost.
11. [illegible]

Proof...

1. Claim: $\exists$ optimal code s.t. x and y are siblings $+$ deepest.
2. $\mathfrak{T}$: optimal code tree with depth d.
3. By lemma... $\mathfrak{T}$ has two leafs at depth d that are siblings,
4. If not x and y, but some other characters α and β.
5. $\mathfrak{T}'$: swap x and α.
6. x depth inc by Δ, and depth of α decreases by Δ.
7. $\mathrm{cost}(\mathfrak{T}') = \mathrm{cost}(\mathfrak{T}) - \big(f[\alpha] - f[x]\big)\Delta$.
8. x: one of the two least frequent characters. ...but α is not.
9. $\implies f[\alpha] \geq f[x]$.
10. Swapping x and α does not increase cost.
11. [illegible]

Proof...

1. Claim: $\exists$ optimal code s.t. x and y are siblings + deepest.
2. $\mathcal{T}$: optimal code tree with depth d.
3. By lemma... $\mathcal{T}$ has two leafs at depth d that are siblings,
4. If not x and y, but some other characters α and β.
5. $\mathcal{T}'$: swap x and α.
6. x depth inc by Δ, and depth of α decreases by Δ.
7. $\mathrm{cost}(\mathcal{T}') = \mathrm{cost}(\mathcal{T}) - \big(f[\alpha] - f[x]\big)\Delta$.
8. x: one of the two least frequent characters.
   ...but α is not.
9. $\implies f[\alpha] \geq f[x]$.
10. Swapping x and α does not increase cost.

Proof...

1. Claim: $\exists$ optimal code s.t. x and y are siblings + deepest.
2. $\mathcal{T}$: optimal code tree with depth d.
3. By lemma... $\mathcal{T}$ has two leafs at depth d that are siblings,
4. If not x and y, but some other characters α and β.
5. $\mathcal{T}'$: swap x and α.
6. x depth inc by Δ, and depth of α decreases by Δ.
7. $\mathbf{cost}(\mathcal{T}') = \mathbf{cost}(\mathcal{T}) - \big(f[\alpha] - f[x]\big)\Delta$.
8. x: one of the two least frequent characters.
   ...but α is not.
9. $\implies f[\alpha] \geq f[x]$.
10. Swapping x and α does not increase cost.

Proof...

1. Claim: $\exists$ optimal code s.t. x and y are siblings $+$ deepest.
2. $\mathcal{T}$: optimal code tree with depth d.
3. By lemma... $\mathcal{T}$ has two leafs at depth d that are siblings,
4. If not x and y, but some other characters α and β.
5. $\mathcal{T}'$: swap x and α.
6. x depth inc by Δ, and depth of α decreases by Δ.
7. $\mathbf{cost}(\mathcal{T}') = \mathbf{cost}(\mathcal{T}) - \big(f[\alpha] - f[x]\big)\Delta$.
8. x: one of the two least frequent characters. ...but α is not.
9. $\implies f[\alpha] \geq f[x]$.
10. Swapping x and α does not increase cost.

Proof...

1. Claim: $\exists$ optimal code s.t. x and y are siblings $+$ deepest.
2. $\mathcal{T}$: optimal code tree with depth d.
3. By lemma... $\mathcal{T}$ has two leafs at depth d that are siblings,
4. If not x and y, but some other characters α and β.
5. $\mathcal{T}'$: swap x and α.
6. x depth inc by Δ, and depth of α decreases by Δ.
7. $\mathbf{cost}(\mathcal{T}') = \mathbf{cost}(\mathcal{T}) - \big(f[\alpha] - f[x]\big)\Delta$.
8. x: one of the two least frequent characters.
   ...but α is not.
9. $\implies f[\alpha] \geq f[x]$.
10. Swapping x and α does not increase cost.
11. [illegible]

Proof...

1. Claim: $\exists$ optimal code s.t. x and y are siblings + deepest.
2. $\mathcal{T}$: optimal code tree with depth d.
3. By lemma... $\mathcal{T}$ has two leafs at depth d that are siblings,
4. If not x and y, but some other characters α and β.
5. $\mathcal{T}'$: swap x and α.
6. x depth inc by Δ, and depth of α decreases by Δ.
7. $\mathbf{cost}(\mathcal{T}') = \mathbf{cost}(\mathcal{T}) - \big(f[\alpha] - f[x]\big)\Delta$.
8. x: one of the two least frequent characters.
   ...but α is not.
9. $\implies f[\alpha] \geq f[x]$.
10. Swapping x and α does not increase cost.

Proof...

1. Claim: $\exists$ optimal code s.t. $\boldsymbol{x}$ and $\boldsymbol{y}$ are siblings $+$ deepest.
2. $\mathcal{T}$: optimal code tree with depth $\boldsymbol{d}$.
3. By lemma... $\mathcal{T}$ has two leafs at depth $\boldsymbol{d}$ that are siblings,
4. If not $\boldsymbol{x}$ and $\boldsymbol{y}$, but some other characters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.
5. $\mathcal{T}'$: swap $\boldsymbol{x}$ and $\boldsymbol{\alpha}$.
6. $\boldsymbol{x}$ depth inc by $\boldsymbol{\Delta}$, and depth of $\boldsymbol{\alpha}$ decreases by $\boldsymbol{\Delta}$.
7. $\mathbf{cost}(\mathcal{T}') = \mathbf{cost}(\mathcal{T}) - \big(\boldsymbol{f}[\boldsymbol{\alpha}] - \boldsymbol{f}[\boldsymbol{x}]\big)\boldsymbol{\Delta}$.
8. $\boldsymbol{x}$: one of the two least frequent characters.
   ...but $\boldsymbol{\alpha}$ is not.
9. $\implies \boldsymbol{f}[\boldsymbol{\alpha}] \geq \boldsymbol{f}[\boldsymbol{x}]$.
10. Swapping $\boldsymbol{x}$ and $\boldsymbol{\alpha}$ does not increase cost.

Proof continued...

1. y: second least frequent character.
2. β: lowest leaf in tree. Sibling to x.
3. Swapping y and β must give yet another optimal code tree.
4. Final opt code tree, x, y are max-depth siblings. ■

Proof continued...

1. y: second least frequent character.
2. β: lowest leaf in tree. Sibling to x.
3. Swapping y and β must give yet another optimal code tree.
4. Final opt code tree, x, y are max-depth siblings. ■

Proof continued...

1. y: second least frequent character.
2. β: lowest leaf in tree. Sibling to x.
3. Swapping y and β must give yet another optimal code tree.
4. Final opt code tree, x, y are max-depth siblings. ■

Proof continued...

1. $\boldsymbol{y}$: second least frequent character.
2. $\boldsymbol{\beta}$: lowest leaf in tree. Sibling to $\boldsymbol{x}$.
3. Swapping $\boldsymbol{y}$ and $\boldsymbol{\beta}$ must give yet another optimal code tree.
4. Final opt code tree, $\boldsymbol{x, y}$ are max-depth siblings. ■

Huffman's codes are optimal

Theorem

Huffman codes are optimal prefix-free binary codes.

Proof...

1. If message has $\mathbf{1}$ or $\mathbf{2}$ diff characters, then theorem easy.
2. $f[1 \dots n]$ be original input frequencies.
3. Assume $f[1]$ and $f[2]$ are the two smallest.
4. Let $f[n+1] = f[1] + f[2]$.
5. lemma $\implies$ $\exists$ opt. code tree $\mathcal{T}_{\mathrm{opt}}$ for $f[1..n]$
6. $\mathcal{T}_{\mathrm{opt}}$ has 1 and 2 as siblings.
7. Remove 1 and 2 from $\mathcal{T}_{\mathrm{opt}}$.
8. $\mathcal{T}'_{\mathrm{opt}}$: Remaining tree has $3, \dots, n$ as leafs and "special" character $n+1$ (i.e., parent $1, 2$ in $\mathcal{T}_{\mathrm{opt}}$)

Proof...

1. If message has $\mathbf{1}$ or $\mathbf{2}$ diff characters, then theorem easy.
2. $\boldsymbol{f}[\mathbf{1} \ldots \boldsymbol{n}]$ be original input frequencies.
3. Assume $\boldsymbol{f}[\mathbf{1}]$ and $\boldsymbol{f}[\mathbf{2}]$ are the two smallest.
4. Let $\boldsymbol{f}[\boldsymbol{n}+\mathbf{1}] = \boldsymbol{f}[\mathbf{1}] + \boldsymbol{f}[\mathbf{2}]$.
5. lemma $\implies$ $\exists$ opt. code tree $\mathcal{T}_{\mathbf{opt}}$ for $\boldsymbol{f}[\mathbf{1..}\boldsymbol{n}]$
6. $\mathcal{T}_{\mathbf{opt}}$ has $\mathbf{1}$ and $\mathbf{2}$ as siblings.
7. Remove $\mathbf{1}$ and $\mathbf{2}$ from $\mathcal{T}_{\mathbf{opt}}$.
8. $\mathcal{T}'_{\mathbf{opt}}$: Remaining tree has $\mathbf{3}, \ldots, \boldsymbol{n}$ as leafs and "special" character $\boldsymbol{n}+\mathbf{1}$ (i.e., parent $\mathbf{1}, \mathbf{2}$ in $\mathcal{T}_{\mathbf{opt}}$)

Proof...

1. If message has $\mathbf{1}$ or $\mathbf{2}$ diff characters, then theorem easy.
2. $f[1 \ldots n]$ be original input frequencies.
3. Assume $f[1]$ and $f[2]$ are the two smallest.
4. Let $f[n+1] = f[1] + f[2]$.
5. lemma $\implies$ $\exists$ opt. code tree $\mathcal{T}_{\mathrm{opt}}$ for $f[1..n]$
6. $\mathcal{T}_{\mathrm{opt}}$ has 1 and 2 as siblings.
7. Remove 1 and 2 from $\mathcal{T}_{\mathrm{opt}}$.
8. $\mathcal{T}'_{\mathrm{opt}}$: Remaining tree has $3, \ldots, n$ as leafs and "special" character $n+1$ (i.e., parent $1, 2$ in $\mathcal{T}_{\mathrm{opt}}$)

Proof...

1. If message has $\mathbf{1}$ or $\mathbf{2}$ diff characters, then theorem easy.
2. $f[1 \ldots n]$ be original input frequencies.
3. Assume $f[1]$ and $f[2]$ are the two smallest.
4. Let $f[n+1] = f[1] + f[2]$.
5. lemma $\implies$ $\exists$ opt. code tree $\mathcal{T}_{\mathrm{opt}}$ for $f[1..n]$
6. $\mathcal{T}_{\mathrm{opt}}$ has 1 and 2 as siblings.
7. Remove 1 and 2 from $\mathcal{T}_{\mathrm{opt}}$.
8. $\mathcal{T}'_{\mathrm{opt}}$: Remaining tree has $3, \ldots, n$ as leafs and "special" character $n+1$ (i.e., parent $1, 2$ in $\mathcal{T}_{\mathrm{opt}}$)

Proof...

1. If message has $\mathbf{1}$ or $\mathbf{2}$ diff characters, then theorem easy.
2. $f[1 \ldots n]$ be original input frequencies.
3. Assume $f[1]$ and $f[2]$ are the two smallest.
4. Let $f[n+1] = f[1] + f[2]$.
5. lemma $\implies$ $\exists$ opt. code tree $\mathcal{T}_{\mathbf{opt}}$ for $f[1..n]$
6. $\mathcal{T}_{\mathbf{opt}}$ has $\mathbf{1}$ and $\mathbf{2}$ as siblings.
7. Remove $\mathbf{1}$ and $\mathbf{2}$ from $\mathcal{T}_{\mathbf{opt}}$.
8. $\mathcal{T}'_{\mathbf{opt}}$: Remaining tree has $3, \ldots, n$ as leafs and "special" character $n+1$ (i.e., parent $1, 2$ in $\mathcal{T}_{\mathbf{opt}}$)

Proof...

1. If message has $\mathbf{1}$ or $\mathbf{2}$ diff characters, then theorem easy.
2. $f[1 \ldots n]$ be original input frequencies.
3. Assume $f[1]$ and $f[2]$ are the two smallest.
4. Let $f[n+1] = f[1] + f[2]$.
5. lemma $\implies$ $\exists$ opt. code tree $\mathcal{T}_{\text{opt}}$ for $f[1..n]$
6. $\mathcal{T}_{\text{opt}}$ has $\mathbf{1}$ and $\mathbf{2}$ as siblings.
7. Remove $\mathbf{1}$ and $\mathbf{2}$ from $\mathcal{T}_{\text{opt}}$.
8. $\mathcal{T}'_{\text{opt}}$: Remaining tree has $3, \ldots, n$ as leafs and "special" character $n+1$ (i.e., parent $1, 2$ in $\mathcal{T}_{\text{opt}}$)

Proof...

1. If message has $\mathbf{1}$ or $\mathbf{2}$ diff characters, then theorem easy.
2. $\boldsymbol{f}[\mathbf{1} \ldots \boldsymbol{n}]$ be original input frequencies.
3. Assume $\boldsymbol{f}[\mathbf{1}]$ and $\boldsymbol{f}[\mathbf{2}]$ are the two smallest.
4. Let $\boldsymbol{f}[\boldsymbol{n}+\mathbf{1}] = \boldsymbol{f}[\mathbf{1}] + \boldsymbol{f}[\mathbf{2}]$.
5. lemma $\implies$ $\exists$ opt. code tree $\mathcal{T}_{\mathbf{opt}}$ for $\boldsymbol{f}[\mathbf{1}..\boldsymbol{n}]$
6. $\mathcal{T}_{\mathbf{opt}}$ has $\mathbf{1}$ and $\mathbf{2}$ as siblings.
7. Remove $\mathbf{1}$ and $\mathbf{2}$ from $\mathcal{T}_{\mathbf{opt}}$.
8. $\mathcal{T}'_{\mathbf{opt}}$: Remaining tree has $\mathbf{3}, \ldots, \boldsymbol{n}$ as leafs and "special" character $\boldsymbol{n}+\mathbf{1}$ (i.e., parent $\mathbf{1}, \mathbf{2}$ in $\mathcal{T}_{\mathbf{opt}}$)

Proof...

1. If message has $\mathbf{1}$ or $\mathbf{2}$ diff characters, then theorem easy.
2. $f[1 \ldots n]$ be original input frequencies.
3. Assume $f[1]$ and $f[2]$ are the two smallest.
4. Let $f[n+1] = f[1] + f[2]$.
5. lemma $\implies$ $\exists$ opt. code tree $\mathcal{T}_{\mathrm{opt}}$ for $f[1..n]$
6. $\mathcal{T}_{\mathrm{opt}}$ has $\mathbf{1}$ and $\mathbf{2}$ as siblings.
7. Remove $\mathbf{1}$ and $\mathbf{2}$ from $\mathcal{T}_{\mathrm{opt}}$.
8. $\mathcal{T}'_{\mathrm{opt}}$: Remaining tree has $3, \ldots, n$ as leafs and "special" character $n+1$ (i.e., parent $1, 2$ in $\mathcal{T}_{\mathrm{opt}}$)

La proof continued...

1. character $n+1$: has frequency $f[n+1]$.
   Now, $f[n+1] = f[1] + f[2]$, we have

$$\mathbf{cost}(\mathcal{T}_{\mathrm{opt}}) = \sum_{i=1}^{n} f[i]\mathbf{depth}_{\mathcal{T}_{\mathrm{opt}}}(i)$$

$$= \sum_{i=3}^{n+1} f[i]\mathbf{depth}_{\mathcal{T}_{\mathrm{opt}}}(i) + f[1]\mathbf{depth}_{\mathcal{T}_{\mathrm{opt}}}(1)$$

$$+ f[2]\mathbf{depth}_{\mathcal{T}_{\mathrm{opt}}}(2) - f[n+1]\mathbf{depth}_{\mathcal{T}_{\mathrm{opt}}}(n+1)$$

$$= \mathbf{cost}\left(\mathcal{T}'_{\mathrm{opt}}\right) + \left(f[1]+f[2]\right)\mathbf{depth}(\mathcal{T}_{\mathrm{opt}})$$

$$- \left(f[1]+f[2]\right)\left(\mathbf{depth}(\mathcal{T}_{\mathrm{opt}}) - 1\right)$$

$$= \mathbf{cost}\left(\mathcal{T}'_{\mathrm{opt}}\right) + f[1] + f[2].$$

La proof continued...

1. character $n+1$: has frequency $f[n+1]$.
   Now, $f[n+1] = f[1] + f[2]$, we have

$$
\begin{aligned}
\mathbf{cost}(\mathcal{T}_{\mathbf{opt}}) &= \sum_{i=1}^{n} f[i]\mathbf{depth}_{\mathcal{T}_{\mathbf{opt}}}(i) \\
&= \sum_{i=3}^{n+1} f[i]\mathbf{depth}_{\mathcal{T}_{\mathbf{opt}}}(i) + f[1]\mathbf{depth}_{\mathcal{T}_{\mathbf{opt}}}(1) \\
&\quad + f[2]\mathbf{depth}_{\mathcal{T}_{\mathbf{opt}}}(2) - f[n+1]\mathbf{depth}_{\mathcal{T}_{\mathbf{opt}}}(n+1) \\
&= \mathbf{cost}\left(\mathcal{T}'_{\mathbf{opt}}\right) + \big(f[1] + f[2]\big)\mathbf{depth}(\mathcal{T}_{\mathbf{opt}}) \\
&\quad - \big(f[1] + f[2]\big)(\mathbf{depth}(\mathcal{T}_{\mathbf{opt}}) - 1) \\
&= \mathbf{cost}\left(\mathcal{T}'_{\mathbf{opt}}\right) + f[1] + f[2].
\end{aligned}
$$

La proof continued...

1. character $n+1$: has frequency $f[n+1]$.
   Now, $f[n+1] = f[1] + f[2]$, we have

$$\mathrm{cost}(\mathcal{T}_{\mathrm{opt}}) = \sum_{i=1}^{n} f[i]\mathrm{depth}_{\mathcal{T}_{\mathrm{opt}}}(i)$$

$$= \sum_{i=3}^{n+1} f[i]\mathrm{depth}_{\mathcal{T}_{\mathrm{opt}}}(i) + f[1]\mathrm{depth}_{\mathcal{T}_{\mathrm{opt}}}(1)$$
$$+ f[2]\mathrm{depth}_{\mathcal{T}_{\mathrm{opt}}}(2) - f[n+1]\mathrm{depth}_{\mathcal{T}_{\mathrm{opt}}}(n+1)$$

$$= \mathrm{cost}\left(\mathcal{T}'_{\mathrm{opt}}\right) + \big(f[1] + f[2]\big)\mathrm{depth}(\mathcal{T}_{\mathrm{opt}})$$
$$- \big(f[1] + f[2]\big)(\mathrm{depth}(\mathcal{T}_{\mathrm{opt}}) - 1)$$

$$= \mathrm{cost}\left(\mathcal{T}'_{\mathrm{opt}}\right) + f[1] + f[2].$$

La proof continued...

1. character $n+1$: has frequency $f[n+1]$.
   Now, $f[n+1] = f[1] + f[2]$, we have

$$\begin{aligned}
\mathrm{cost}(\mathcal{T}_{\mathrm{opt}}) &= \sum_{i=1}^{n} f[i]\mathrm{depth}_{\mathcal{T}_{\mathrm{opt}}}(i) \\
&= \sum_{i=3}^{n+1} f[i]\mathrm{depth}_{\mathcal{T}_{\mathrm{opt}}}(i) + f[1]\mathrm{depth}_{\mathcal{T}_{\mathrm{opt}}}(1) \\
&\quad + f[2]\mathrm{depth}_{\mathcal{T}_{\mathrm{opt}}}(2) - f[n+1]\mathrm{depth}_{\mathcal{T}_{\mathrm{opt}}}(n+1) \\
&= \mathrm{cost}\left(\mathcal{T}'_{\mathrm{opt}}\right) + \big(f[1] + f[2]\big)\mathrm{depth}(\mathcal{T}_{\mathrm{opt}}) \\
&\quad - \big(f[1] + f[2]\big)(\mathrm{depth}(\mathcal{T}_{\mathrm{opt}}) - 1) \\
&= \mathrm{cost}\left(\mathcal{T}'_{\mathrm{opt}}\right) + f[1] + f[2].
\end{aligned}$$

La proof continued...

1. implies $\mathbf{min}$ cost of $\mathcal{T}_{\mathbf{opt}} \equiv \mathbf{min}$ cost $\mathcal{T}'_{\mathbf{opt}}$.
2. $\mathcal{T}'_{\mathbf{opt}}$: must be optimal coding tree for $f[3 \ldots n+1]$.
3. $\mathcal{T}'_{H}$: Huffman tree for $f[3, \ldots, n+1]$
   $\mathcal{T}_{H}$: overall Huffman tree constructed for $f[1, \ldots, n]$.
4. By construction:
   $\mathcal{T}'_{H}$ formed by removing leafs 1 and 2 from $\mathcal{T}_{H}$.
5. By induction:
   Huffman tree generated for $f[3, \ldots, n+1]$ is optimal.
6. $\mathbf{cost}\left(\mathcal{T}'_{\mathbf{opt}}\right) = \mathbf{cost}\left(\mathcal{T}'_{H}\right)$.
7. $\implies \mathbf{cost}(\mathcal{T}_{H}) = \mathbf{cost}(\mathcal{T}'_{H}) + f[1] + f[2] =$

La proof continued...

1. implies $\mathbf{min}$ cost of $\mathcal{T}_{\mathbf{opt}} \equiv \mathbf{min}$ cost $\mathcal{T}'_{\mathbf{opt}}$.
2. $\mathcal{T}'_{\mathbf{opt}}$: must be optimal coding tree for $f[3 \ldots n+1]$.
3. $\mathcal{T}'_H$: Huffman tree for $f[3, \ldots, n+1]$
   $\mathcal{T}_H$: overall Huffman tree constructed for $f[1, \ldots, n]$.
4. By construction:
   $\mathcal{T}'_H$ formed by removing leafs 1 and 2 from $\mathcal{T}_H$.
5. By induction:
   Huffman tree generated for $f[3, \ldots, n+1]$ is optimal.
6. $\mathbf{cost}\left(\mathcal{T}'_{\mathbf{opt}}\right) = \mathbf{cost}\left(\mathcal{T}'_H\right)$.
7. $\implies \mathbf{cost}(\mathcal{T}_H) = \mathbf{cost}(\mathcal{T}'_H) + f[1] + f[2] =$

La proof continued...

1. implies $\mathbf{min}$ cost of $\mathcal{T}_{\mathbf{opt}} \equiv \mathbf{min}$ cost $\mathcal{T}'_{\mathbf{opt}}$.
2. $\mathcal{T}'_{\mathbf{opt}}$: must be optimal coding tree for $f[3 \ldots n+1]$.
3. $\mathcal{T}'_H$: Huffman tree for $f[3, \ldots, n+1]$
   $\mathcal{T}_H$: overall Huffman tree constructed for $f[1, \ldots, n]$.
4. By construction:
   $\mathcal{T}'_H$ formed by removing leafs 1 and 2 from $\mathcal{T}_H$.
5. By induction:
   Huffman tree generated for $f[3, \ldots, n+1]$ is optimal.
6. $\mathbf{cost}\left(\mathcal{T}'_{\mathbf{opt}}\right) = \mathbf{cost}\left(\mathcal{T}'_H\right)$.
7. $\implies \mathbf{cost}(\mathcal{T}_H) = \mathbf{cost}(\mathcal{T}'_H) + f[1] + f[2] =$

La proof continued...

1. implies $\mathbf{min}$ cost of $\mathcal{T}_{\mathbf{opt}} \equiv \mathbf{min}$ cost $\mathcal{T}'_{\mathbf{opt}}$.
2. $\mathcal{T}'_{\mathbf{opt}}$: must be optimal coding tree for $f[3 \ldots n+1]$.
3. $\mathcal{T}'_{H}$: Huffman tree for $f[3, \ldots, n+1]$
   $\mathcal{T}_{H}$: overall Huffman tree constructed for $f[1, \ldots, n]$.
4. By construction:
   $\mathcal{T}'_{H}$ formed by removing leafs 1 and 2 from $\mathcal{T}_{H}$.
5. By induction:
   Huffman tree generated for $f[3, \ldots, n+1]$ is optimal.
6. $\mathrm{cost}\left(\mathcal{T}'_{\mathrm{opt}}\right) = \mathrm{cost}\left(\mathcal{T}'_{H}\right)$.
7. $\implies \mathrm{cost}(\mathcal{T}_{H}) = \mathrm{cost}\left(\mathcal{T}'_{H}\right) + f[1] + f[2] =$

La proof continued...

1. implies $\mathbf{min}$ cost of $\mathcal{T}_{\mathbf{opt}} \equiv \mathbf{min}$ cost $\mathcal{T}'_{\mathbf{opt}}$.
2. $\mathcal{T}'_{\mathbf{opt}}$: must be optimal coding tree for $f[3 \ldots n+1]$.
3. $\mathcal{T}'_H$: Huffman tree for $f[3, \ldots, n+1]$
   $\mathcal{T}_H$: overall Huffman tree constructed for $f[1, \ldots, n]$.
4. By construction:
   $\mathcal{T}'_H$ formed by removing leafs 1 and 2 from $\mathcal{T}_H$.
5. By induction:
   Huffman tree generated for $f[3, \ldots, n+1]$ is optimal.
6. $\mathrm{cost}\left(\mathcal{T}'_{\mathrm{opt}}\right) = \mathrm{cost}\left(\mathcal{T}'_H\right)$.
7. $\implies \mathrm{cost}(\mathcal{T}_H) = \mathrm{cost}\left(\mathcal{T}'_H\right) + f[1] + f[2] =$

La proof continued...

1. implies $\mathbf{min}$ cost of $\mathcal{T}_{\mathbf{opt}} \equiv \mathbf{min}$ cost $\mathcal{T}'_{\mathbf{opt}}$.
2. $\mathcal{T}'_{\mathbf{opt}}$: must be optimal coding tree for $f[3 \ldots n+1]$.
3. $\mathcal{T}'_{H}$: Huffman tree for $f[3, \ldots, n+1]$
   $\mathcal{T}_{H}$: overall Huffman tree constructed for $f[1, \ldots, n]$.
4. By construction:
   $\mathcal{T}'_{H}$ formed by removing leafs 1 and 2 from $\mathcal{T}_{H}$.
5. By induction:
   Huffman tree generated for $f[3, \ldots, n+1]$ is optimal.
6. $\mathbf{cost}\left(\mathcal{T}'_{\mathbf{opt}}\right) = \mathbf{cost}\left(\mathcal{T}'_{H}\right)$.
7. $\implies \mathbf{cost}(\mathcal{T}_{H}) = \mathbf{cost}(\mathcal{T}'_{H}) + f[1] + f[2] =$

La proof continued...

1. implies $\mathbf{min}$ cost of $\mathcal{T}_{\mathbf{opt}} \equiv \mathbf{min}$ cost $\mathcal{T}'_{\mathbf{opt}}$.
2. $\mathcal{T}'_{\mathbf{opt}}$: must be optimal coding tree for $f[3 \ldots n+1]$.
3. $\mathcal{T}'_H$: Huffman tree for $f[3, \ldots, n+1]$
   $\mathcal{T}_H$: overall Huffman tree constructed for $f[1, \ldots, n]$.
4. By construction:
   $\mathcal{T}'_H$ formed by removing leafs 1 and 2 from $\mathcal{T}_H$.
5. By induction:
   Huffman tree generated for $f[3, \ldots, n+1]$ is optimal.
6. $\mathbf{cost}\left(\mathcal{T}'_{\mathbf{opt}}\right) = \mathbf{cost}\left(\mathcal{T}'_H\right)$.
7. $\implies \mathbf{cost}(\mathcal{T}_H) = \mathbf{cost}\left(\mathcal{T}'_H\right) + f[1] + f[2] =$

La proof continued...

1. implies $\mathbf{min}$ cost of $\mathcal{T}_{\mathbf{opt}} \equiv \mathbf{min}$ cost $\mathcal{T}'_{\mathbf{opt}}$.
2. $\mathcal{T}'_{\mathbf{opt}}$: must be optimal coding tree for $f[3 \ldots n+1]$.
3. $\mathcal{T}'_{H}$: Huffman tree for $f[3, \ldots, n+1]$
   $\mathcal{T}_{H}$: overall Huffman tree constructed for $f[1, \ldots, n]$.
4. By construction:
   $\mathcal{T}'_{H}$ formed by removing leafs 1 and 2 from $\mathcal{T}_{H}$.
5. By induction:
   Huffman tree generated for $f[3, \ldots, n+1]$ is optimal.
6. $\mathbf{cost}\left(\mathcal{T}'_{\mathbf{opt}}\right) = \mathbf{cost}\left(\mathcal{T}'_{H}\right)$.
7. $\implies \mathbf{cost}(\mathcal{T}_{H}) = \mathbf{cost}\left(\mathcal{T}'_{H}\right) + f[1] + f[2] =$

What we get...

1. A tale of two cities: 779,940 bytes.
2. using above Huffman compression results in a compression to a file of size 439,688 bytes.
3. Ignoring space to store tree.
4. `gzip`: 301,295 bytes
   `bzip2`: 220,156 bytes!
5. Huffman encoder can be easily written in a few hours of work!
6. All later compressors use it as a black box...

What we get...

1. A tale of two cities: 779,940 bytes.
2. using above Huffman compression results in a compression to a file of size 439,688 bytes.
3. Ignoring space to store tree.
4. `gzip`: 301,295 bytes
   `bzip2`: 220,156 bytes!
5. Huffman encoder can be easily written in a few hours of work!
6. All later compressors use it as a black box...

What we get...

1. A tale of two cities: 779,940 bytes.
2. using above Huffman compression results in a compression to a file of size 439,688 bytes.
3. Ignoring space to store tree.
4. `gzip`: 301,295 bytes
   `bzip2`: 220,156 bytes!
5. Huffman encoder can be easily written in a few hours of work!
6. All later compressors use it as a black box...

What we get...

1. A tale of two cities: 779,940 bytes.
2. using above Huffman compression results in a compression to a file of size 439,688 bytes.
3. Ignoring space to store tree.
4. `gzip`: 301,295 bytes
   `bzip2`: 220,156 bytes!
5. Huffman encoder can be easily written in a few hours of work!
6. All later compressors use it as a black box...

What we get...

1. A tale of two cities: 779,940 bytes.
2. using above Huffman compression results in a compression to a file of size 439,688 bytes.
3. Ignoring space to store tree.
4. `gzip`: 301,295 bytes
   `bzip2`: 220,156 bytes!
5. Huffman encoder can be easily written in a few hours of work!
6. All later compressors use it as a black box...

What we get...

1. A tale of two cities: 779,940 bytes.
2. using above Huffman compression results in a compression to a file of size 439,688 bytes.
3. Ignoring space to store tree.
4. `gzip`: 301,295 bytes
   `bzip2`: 220,156 bytes!
5. Huffman encoder can be easily written in a few hours of work!
6. All later compressors use it as a black box...

Huffman trees are...

Clicker question

$\Sigma = \{1, \ldots, n\}$: alphabet
$f[1 \ldots n]$: Frequencies.
Assume all subset sums of frequencies are unique.
Then the optimal code tree for Σ is:

1. Unique
2. There are 2^{n-1} different optimal code trees.
3. There are $n!$ different optimal code trees.
4. There are infinite number of different optimal code trees.
5. None of the above.

Average size of code word

1. input is made out of n characters.
2. p_i: fraction of input that is ith char (probability).
3. use probabilities to build Huffman tree.
4. Q: What is the length of the codewords assigned to characters as function of probabilities?
5. special case...

Average size of code word

1. input is made out of n characters.
2. p_i: fraction of input that is ith char (probability).
3. use probabilities to build Huffman tree.
4. Q: What is the length of the codewords assigned to characters as function of probabilities?
5. special case...

Average size of code word

1. input is made out of n characters.
2. p_i: fraction of input that is ith char (probability).
3. use probabilities to build Huffman tree.
4. Q: What is the length of the codewords assigned to characters as function of probabilities?
5. special case...

Average size of code word

1. input is made out of n characters.
2. p_i: fraction of input that is ith char (probability).
3. use probabilities to build Huffman tree.
4. Q: What is the length of the codewords assigned to characters as function of probabilities?
5. special case...

Average size of code word

1. input is made out of n characters.
2. p_i: fraction of input that is ith char (probability).
3. use probabilities to build Huffman tree.
4. Q: What is the length of the codewords assigned to characters as function of probabilities?
5. special case...

Average length of codewords...

Special case

Lemma

$1, \ldots, n$: symbols.
Assume, for $i = 1, \ldots, n$:

1. *$p_i = 1/2^{l_i}$: probability for the ith symbol*
2. *$l_i \geq 0$: integer.*

Then, in Huffman coding for this input, the code for i is of length l_i.

Proof

1. induction of the Huffman algorithm.
2. $n = 2$: claim holds since there are only two characters with probability $1/2$.
3. Let i and j be the two characters with lowest probability.
4. Must be $p_i = p_j$ (otherwise, $\sum_k p_k \neq 1$).
5. Huffman's tree merges this two letters, into a single "character" that have probability $2p_i$.
6. New "character" has encoding of length $l_i - 1$, by induction
   (on remaining $n - 1$ symbols).
7. resulting tree encodes i and j by code words of

Proof

1. induction of the Huffman algorithm.
2. $n = 2$: claim holds since there are only two characters with probability $1/2$.
3. Let i and j be the two characters with lowest probability.
4. Must be $p_i = p_j$ (otherwise, $\sum_k p_k \neq 1$).
5. Huffman's tree merges this two letters, into a single "character" that have probability $2p_i$.
6. New "character" has encoding of length $l_i - 1$, by induction
   (on remaining $n - 1$ symbols).
7. resulting tree encodes i and j by code words of

Proof

1. induction of the Huffman algorithm.
2. $n = 2$: claim holds since there are only two characters with probability $1/2$.
3. Let i and j be the two characters with lowest probability.
4. Must be $p_i = p_j$ (otherwise, $\sum_k p_k \neq 1$).
5. Huffman's tree merges this two letters, into a single "character" that have probability $2p_i$.
6. New "character" has encoding of length $l_i - 1$, by induction
   (on remaining $n - 1$ symbols).
7. resulting tree encodes i and j by code words of

Proof

1. induction of the Huffman algorithm.
2. $n = 2$: claim holds since there are only two characters with probability $1/2$.
3. Let i and j be the two characters with lowest probability.
4. Must be $p_i = p_j$ (otherwise, $\sum_k p_k \neq 1$).
5. Huffman's tree merges this two letters, into a single "character" that have probability $2p_i$.
6. New "character" has encoding of length $l_i - 1$, by induction
   (on remaining $n - 1$ symbols).
7. resulting tree encodes i and j by code words of

Proof

1. induction of the Huffman algorithm.
2. $n = 2$: claim holds since there are only two characters with probability $1/2$.
3. Let i and j be the two characters with lowest probability.
4. Must be $p_i = p_j$ (otherwise, $\sum_k p_k \neq 1$).
5. Huffman's tree merges this two letters, into a single "character" that have probability $2p_i$.
6. New "character" has encoding of length $l_i - 1$, by induction
   (on remaining $n - 1$ symbols).
7. resulting tree encodes i and j by code words of

Proof

1. induction of the Huffman algorithm.
2. $n = 2$: claim holds since there are only two characters with probability $1/2$.
3. Let i and j be the two characters with lowest probability.
4. Must be $p_i = p_j$ (otherwise, $\sum_k p_k \neq 1$).
5. Huffman's tree merges this two letters, into a single "character" that have probability $2p_i$.
6. New "character" has encoding of length $l_i - 1$, by induction
   (on remaining $n - 1$ symbols).
7. resulting tree encodes i and j by code words of

Proof

1. induction of the Huffman algorithm.
2. $n = 2$: claim holds since there are only two characters with probability $1/2$.
3. Let i and j be the two characters with lowest probability.
4. Must be $p_i = p_j$ (otherwise, $\sum_k p_k \neq 1$).
5. Huffman's tree merges this two letters, into a single "character" that have probability $2p_i$.
6. New "character" has encoding of length $l_i - 1$, by induction
   (on remaining $n - 1$ symbols).
7. resulting tree encodes i and j by code words of

Translating lemma...

1. $p_i = 1/2^{l_i}$
2. $l_i = \lg 1/p_i$.
3. Average length of a code word is

$$\sum_i p_i \lg \frac{1}{p_i}.$$

4. X is a random variable that takes a value i with probability p_i, then this formula is

$$\mathbb{H}(X) = \sum_i \Pr[X = i] \lg \frac{1}{\Pr[X = i]},$$

Translating lemma...

1. $p_i = 1/2^{l_i}$
2. $l_i = \lg 1/p_i$.
3. Average length of a code word is
$$\sum_i p_i \lg \frac{1}{p_i}.$$
4. X is a random variable that takes a value i with probability p_i, then this formula is
$$\mathbb{H}(X) = \sum_i \Pr[X = i] \lg \frac{1}{\Pr[X = i]},$$

Translating lemma...

1. $p_i = 1/2^{l_i}$
2. $l_i = \lg 1/p_i$.
3. Average length of a code word is

$$\sum_i p_i \lg \frac{1}{p_i}.$$

4. X is a random variable that takes a value i with probability p_i, then this formula is

$$\mathbb{H}(X) = \sum_i \Pr[X = i] \lg \frac{1}{\Pr[X = i]},$$

Translating lemma...

1. $p_i = 1/2^{l_i}$
2. $l_i = \lg 1/p_i$.
3. Average length of a code word is

$$\sum_i p_i \lg \frac{1}{p_i}.$$

4. X is a random variable that takes a value i with probability p_i, then this formula is

$$\mathbb{H}(X) = \sum_i \Pr[X = i] \lg \frac{1}{\Pr[X = i]},$$

Notes

Notes

Notes

Notes