CS 170 DIS 06

Released on 2019-2-25

1 Horn Formula Practice

Find the variable assignment that solves the following horn formulas:

1. $(w \wedge y \wedge z) \Rightarrow x, (x \wedge z) \Rightarrow w, x \Rightarrow y, \Rightarrow x, (x \wedge y) \Rightarrow w, (\bar{w} \vee \bar{x} \vee \bar{y}), (\bar{z})$

2. $(x \wedge z) \Rightarrow y, z \Rightarrow w, (y \wedge z) \Rightarrow x, \Rightarrow z, (\bar{z} \vee \bar{x}), (\bar{w} \vee \bar{y} \vee \bar{z})$

2 Longest Huffman Tree

Under a Huffman encoding of n symbols with frequencies $f_1, f_2, \ldots, f_n$, what is the longest a codeword could possibly be? Give an example set of frequencies that would produce this case, and argue that it is the longest possible.

3 Proof of Huffman Coding

In this question, we will prove that Huffman coding indeed produces the best prefix-free code for a given set of characters and associated frequencies. Recall that we are given as input a set of characters $c_1, \ldots, c_n$ and frequencies $f_1, \ldots, f_n$ and the goal is produce a binary tree T where the leaves of the tree correspond to the characters c_i which is as efficient as possible. That is, the tree produced should minimize $\sum_{i=1}^{n} f_i d_T(c_i)$ where $d_T(c_i)$ denotes the depth of c_i in the tree, T. For this question, we will view Huffman coding as a recursive algorithm which proceeds along the following lines:

1. Merge the two characters with the lowest frequencies, say c_1 and c_2, to produce a "meta-character", (c_1, c_2).
2. Run the Huffman tree procedure on the set of characters $(c_1, c_2), c_3, \ldots, c_n$ with frequencies $(f_1 + f_2), f_3, \ldots, f_n$.
3. Let the tree obtained in the previous step be $T^\dagger$. Replace the node corresponding to (c_1, c_2) with an internal node with two children c_1 and c_2 to produce the final tree T.

(a) For the first part of the question, we will prove that every internal node of the optimal tree, T^*, has two children. (*Hint: Does a violation of this property create a contradiction?*)

(b) Now, let c_1 and c_2 be the two characters with the lowest frequencies. Prove that the cost of the optimal tree, T^*, can only reduce if c_1 and c_2 are made siblings in the lowest leaves of the tree.

(c) Conclude via induction that Huffman coding indeed produces the optimal tree. (*Hint: Can you relate the cost of the tree, T, produced by Huffman coding to the cost of $T^\dagger$?*)