D7 - Compression and Satisfiability

Reading material on Satisfiability (relating to questions 1-4) is in special notes, see *Modules* in Canvas). Material on NP-completeness (questions 6-7) is pp.903–921 in Sedgewick-Wayne.

1. Consider the CNF formula

$$x_1x_2, \overline{x_1x_2}, \overline{x_1x_3}, \overline{x_2x_3}, x_2x_3$$

   a) Find a satisfying assignment for the formula, and
   b) Give a falsifying 3-clause (i.e., a clause with 3 literals that, if added to the formula, makes the formula false)

2. (8%) Perform unit propagation on the CNF formula

$$x_1x_2, \overline{x_2}, x_2x_4, \overline{x_1}x_2\overline{x_4},$$

3. Perform pure literal elimination on CNF formula

$$x_1x_2\overline{x_3}, \overline{x_2x_3}, x_2\overline{x_3}x_4, \overline{x_1}x_2\overline{x_4}, x_1x_3x_5\ .$$

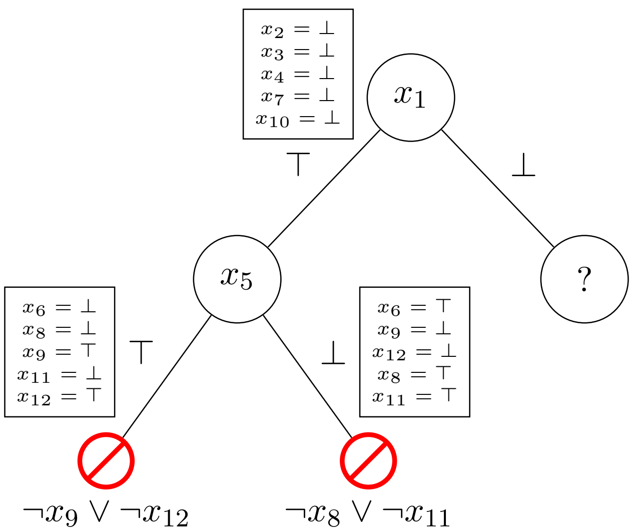

Figure 1: Partial run of the DPLL heuristic

4. (20%) The following CNF formula was treated in class (Tue 30 Oct, slide 21):

$$\begin{aligned}\mathcal{F}' :\quad & x_1x_2x_3,\ \overline{x_1}\,\overline{x_2},\ \overline{x_1}\,\overline{x_3}, \overline{x_2}\,\overline{x_3},\ x_4x_5x_6,\ \overline{x_4}\,\overline{x_5},\ \overline{x_4}\,\overline{x_6},\ \overline{x_5}\,\overline{x_6},\\ & x_7x_8x_9,\ \overline{x_7}\,\overline{x_8},\ \overline{x_7}\,\overline{x_9},\ \overline{x_8}\,\overline{x_9},\ x_{10}x_{11}x_{12},\ \overline{x_{10}}\,\overline{x_{11}},\ \overline{x_{10}}\,\overline{x_{12}},\ \overline{x_{11}}\,\overline{x_{12}},\\ & \overline{x_1}\,\overline{x_4},\ \overline{x_2}\,\overline{x_5},\ \overline{x_3}\,\overline{x_6},\ \overline{x_1}\,\overline{x_7},\ \overline{x_2}\,\overline{x_8},\ \overline{x_3}\,\overline{x_9},\ \overline{x_1}\,\overline{x_{10}},\ \overline{x_2}\,\overline{x_{11}},\ \overline{x_3}\,\overline{x_{12}},\\ & \overline{x_4}\,\overline{x_7},\ \overline{x_5}\,\overline{x_8},\ \overline{x_6}\,\overline{x_9},\ \overline{x_4}\,\overline{x_{10}},\ \overline{x_5}\,\overline{x_{11}},\ \overline{x_6}\,\overline{x_{12}},\ \overline{x_7}\,\overline{x_{10}},\ \overline{x_8}\,\overline{x_{11}},\ \overline{x_9}\,\overline{x_{12}}\ .\end{aligned}$$

The result of the initial trace of $\hat{\mathcal{F}}$ that was done in class (and lecture slides) is shown in Fig. 1. Continue the trace for the subtree given by the partial assignment $x_1 = F$. Label the edges with the unit clauses that were propagated. On "dead-end" states (shown on slide with a red X), identify the clause that is responsible. Assume that next variable explored is always the lowest unassigned one.

5. Give the LZW encoding the string a^N consisting of N repeats of the character `a`. What is the compression ratio as a function of N?

   The LZW encoding will store and use the codewords `a`, `aa`, `aaa`, etc. In round i, it will code the substring $\mathtt{a}^i$ with the i-th code value (for all but the last round), and record in the symbol table the $i+1$-st code value as representing $\mathtt{a}^{i+1}$. If k is the number of codewords used, then k is the smallest value such that $\sum_{i=1}^{k} i \geq n$, or $\binom{k+1}{2} \geq n$. Then, $k \sim \sqrt{2n}$. Thus, the compression ratio is $\Theta(1/\sqrt{n})$.]

6. (Problem 5.5.18) Let F_k be the k-th Fibonacci number. Consider N symbols, where the k-th symbol has frequency F_k. Note that $F_1 + F_2 + \ldots F_N = F_{N+2} - 1$. Describe the Huffman code. (Hint: The longest codeword has length $N-1$).

   The tree will be a chain, where the length of the i-th codeword is i, for $i = 1, 2, \ldots, N-1$; the i-th codeword will represent the $N-k+1$-th symbol.]

7. (10%) (Problem 6.62.) Suppose that $P \neq NP$. Which of the following can we infer:

   e. If X is NP-complete, then X cannot be solved in polynomial time.

   f. If X is in NP, then X cannot be solved in polynomial time.

   g. If X is in NP but not NP-complete, then X can be solved in polynomial time.

   h. If X is in NP, then X is not NP-complete.

8. (16%) Classify the following problems as in P, as NP-complete, or neither. Identify the usual name for the (underlying abstract) problem.

   (a) Given a graph, is it possible to mark each node by "X", "Y" or "Z", such that neighbors always get a different mark?

   (b) Given a set X of numbers $x_1, x_2, \ldots, x_n$, are there two sets X_1, X_2 such that $X_1 \cup X_2 = X$ and $X_1 \cap X_2 = \emptyset$ and such that $\sum_{x_i \in X_1} x_i = \sum_{x_i \in X_2} x_i$?

   (c) Given an integer k and an virtual environment consisting of n avatars, where any two are either *friends* or *enemies*, are there k avatars that are all friends of each other?

   (d) Given a computer network and two nodes a and b, and an integer k, is it possible to forward a message from a to b using at most k intermediate nodes?

Class Exercises

I. (Final Exam 2017) Decode the following LZW-compressed message:

```
41 81 42 82 83 80
```

II. Encode the string "ABABCABCD" using LZW.

III. Suppose we are given the following letter frequencies: A:60, D:6, E:75, F:3, G:9, H:4, S:17. Give the optimal Huffman coding tree. Also, how is "SEEAHEAD" compressed?

IV. Implement the pure-literal elimination rule, either in a Java function or in words. State the time complexity.