COMP 597: Assignment #1

Instructor: Xujie Si, TA: Breandan Considine

Deadline: Oct. 13th, 23:59 in Montreal

This is the first assignment of COMP 597: Automated Reasoning with ML.

1 Cryptography with SAT

The following solutions can be obtained using a solver of your choice. Please show all work. Only solutions obtained using a SAT solver will receive credit.

Emirpimes (20 points): An *emirp* is a prime whose digits, when reversed, produce a different prime. An *emirpimes* is a semiprime whose reverse is a different semiprime, e.g., 11659567_{10}. Confirm its prime factors are emirps in bases 2 and 10. Find another such number. What is the largest emirpimes you can find, whose prime factors are twin emirps in at least two bases?

Bonus (10 points): A *cryptarithm* is a cipher, $\varphi : \{A, \ldots, Z\}^* \leftrightarrow \{0, \ldots, 9\}^*$, alongside a meaningful string, whose ciphertext satisfies some equation, e.g.:

```
NINETEEN + THIRTEEN + THREE + TWO + TWO + ONE + ONE + ONE = FORTYTWO
42415114 + 56275114 + 56711 + 538 + 538 + 841 + 841 + 841 = 98750538
```

Construct a 20+-character cryptarithm parseable by the following grammar,

$$
\begin{aligned}
&E \to A \mid \ldots \mid Z \mid EE \mid EOE \mid (E)\\
&O \to + \mid \times \mid \div \mid -^1\\
&S \to E = E
\end{aligned}
$$

where $\texttt{eval}\big(\varphi(E)\big) = \texttt{eval}\big(\varphi(E')\big)$ and $\texttt{charset}(E) \neq \texttt{charset}(E')$. Every plaintext word should be defined in the English 10k dictionary.[2] In order to receive credit, it must not be possible to find your cryptarithm (or algebraic rewritings thereof) on the internet or in other classmates' assignments.

[1] Interpreted in the usual way, but additive and multiplicative identity are forbidden.

[2] https://github.com/first20hours/google-10000-english/blob/master/google-10000-english.txt

2 Build or improve a SAT solver

Programming exercise (40 points): Please select one of the following two options, write a short report, and submit your source code. Please provide instructions for how reproduce your findings and a few test cases.

1. Write a SAT solver from scratch by implementing an existing algorithm such as DPLL, unit propagation or two-watched literals, describe your implementation and evaluate it on a few toy SAT problems.

2. Make a substantive improvement to a competitive SAT solver (e.g. Kissat or MiniSat) which measurably increases performance on a standard benchmark, and document your approach and findings.

3 Uninterpreted function equivalence

SMT exercise (20 points): Show all work to receive full credit. Where required, typeset a proof sketch using LaTeX, then translate the proof into your favorite SMT solver to construct a specific example or counterexample.

1. A polynomial equation whose coefficients and solutions are integers is called *diophantine*. Let $w, x, y, z \in \mathbb{Z}$ and report your solver's largest nontrivial solutions to each of the following diophantine equations: (a) $x^2 + y^2 + z = wxy$ (b) $w^3 + x^3 = y^3 + z^3$ (c) $w^z + x^z = y^z + z$.

2. Prove that $\mathbb{Z}^{n\times n}$ is associative over $\otimes$, and $\otimes$ is distributive over $\oplus$ for some large n. **Bonus** (5 points): Give an example of a nontrivial finite commutative semiring whose elements are matrices and prove it.

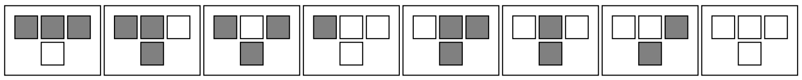
3. A nonnegative matrix whose rows and columns all sum to the same number is called *bistochastic*. Find distinct examples $M_1, M_2 : \mathbb{Z}^{n\times n}$ for some large n such that both are nontrivial bistochastic matrices. **Bonus** (5 points): Is $M_i M_j$ is bistochastic for all bistochastic M_i, M_j?

4. Prove that 1D discrete convolution, $* : (f, g)[x] \mapsto \sum_{s\in S} f[x - s]g[s]$, over $S = [-j, j]$ for some large value $j \in \mathbb{N}$ is translation equivariant. **Bonus** (10 points): Prove the 2D case for MNIST, i.e., $[0, 255]^{28\times 28}$.

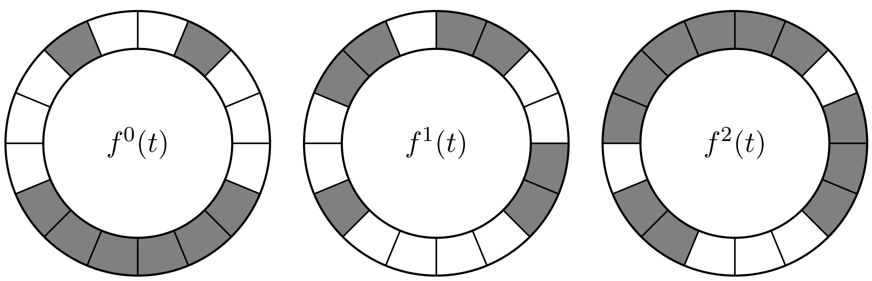
Only evidence in the form of (1) an example or counterexample, or (2) an interpretable proof will be accepted. Proofs should be *constructive* where necessary. Encoding the problem and reporting UNSAT is not constructive.

4 Exploring the multiverse

Creative coding (20 points): Consider a multiverse $\mathcal{M}$, with some strange physical laws, whose dynamics are governed by the following eight rules.



These rules describe a substitution system, in which the first row denotes a pattern match, and the second row denotes the subsequent state of the innermost cell. In this multiverse, the fabric of space folds in upon itself: travel far enough in any direction and you shall always return from whence you came. For example, we can visualize three steps in the time evolution of the universe $\mathcal{U}' : \{\square, \blacksquare\}^{16} \in \mathcal{M}$, starting from the initial state t as follows:



Now suppose you are a xenobiologist exploring $\mathcal{U} : \{\square, \blacksquare\}^{128}$, assigned with the task of discovering alien life forms inhabiting in this strange universe.

(a) Find an *uroboros*, a creature which is its own ancestor: $f^k(\sigma) = \sigma \neq f(\sigma)$.
(b) Find an *orphan*, a creature which has no parent: $\sigma \in \mathcal{U} \mid \nexists \sigma'.f(\sigma') = \sigma$.
(c) Find an *endling*, the last living descendent of its kind: $t \neq f^1(t) = f^2(t)$.
(d) Find a *chimera*, a creature with three parents: $r, s, t \mid f(r) = f(s) = f(t)$.

For identification purposes, you may label your specimens for submission using a string of 0s and 1s. Rare and fantastic specimens with additional structure will receive special handling and may be elligible for public viewing.

Bonus (10 points): Find or design intelligent life in $\mathcal{M}$. This creature should encode a learning algorithm, e.g., a neural network. You must describe the encoding and decoding scheme and demonstrate the creature can learn.

Please submit your answers as a PDF and supplemental work as a ZIP file.