Solutions to Midterm 1

1. Every day, a weather station records whether the day was sunny (S), cloudy (C) or rainy (R). A sequence of records over several days is a string in $\{S, C, R\}^*$. We call a sequence of records *Berkeley-like* if:

   - There are never more than three consecutive rainy days, and
   - There are never more than five consecutive non-sunny days.

   Show that the language of Berkeley-like sequences is regular.
   [20 points]

   Solution Outline: The language of sequences with no more than three consecutive rainy days is regular since it can be expressed by the regular expression $((\epsilon + R)(\epsilon + R)(\epsilon + R)(S + C))^*$. Similarly, the the language of sequences with no more than five non-sunny days can be captured by the expression $((\epsilon + R + C)(\epsilon + R + C)(\epsilon + R + C)(\epsilon + R + C)(\epsilon + R + C)S)^*$. Since the intersection of two regular languages is regular, the given language must be regular.

2. Consider the language

   $$A_{\mathsf{DFAinf}} = \{D \mid D \text{ is a DFA that recognizes a language containing infinitely many strings}\}$$

   Prove that A_{DFAinf} is decidable.
   [40 points]

   Solution Outline: One method to decide the language is to convert the DFA into an equivalent regular expression. It is easy to see that the language is infinite if and only if the expression contains a sub-expression of the form A^* for some non-empty expression A (expression not equivalent to the empty string). Also, since union, concatenation and repetition of non-empty expressions is non-empty, we can check this simply by recursing on parts of A.

   Alternatively, we can see using the pumping lemma that a DFA D with n states recognizes an infinite language iff it accepts some string of length between n and $2n$. The "if" part follows directly from the statement of the pumping lemma. To prove the "only if" part, note that any DFA accepting an infinite language must have a path from the start state to an accepting state, with a cycle involving at least one of the vertices in the path. Since both the path and the cycle are of length at most n, finitely many repetitions of the cycle will give a string of length between n and $2n$ accepted by the DFA. Hence, to decide if D recognizes an infinite language, it suffices to enumerate over all strings w of length between n and $2n$, and accept if D accepts (at least) one of these strings, and reject otherwise.

3. One of the following languages is Turing-recognizable; the other is not. Which is which?

   - $A = \{\langle M \rangle \mid M \text{ accepts at most 172 distinct inputs}\}$
   - $B = \{\langle M \rangle \mid M \text{ accepts more than 172 distinct inputs}\}$

[40 points]

SOLUTION OUTLINE: By Rice's Theorem, both A and B are undecidable. In addition, it is easy to see that B is Turing-recognizable, either by nondeterminisically guessing 173 distinct inputs (noting that nondeterministic Turing machines are no more powerful than deterministic ones) and simulating M on these inputs, or by using the "time-slicing" trick from constructing an enumerator for Turing-recognizable languages and accepting whenever M accepts 173 distinct inputs. Since A and B are complement languages, it follows that A is not Turing-recognizable.