Midterm I

Instructions

1. Do not turn over this page until instructed to do so.
2. This exam contains 4 questions. Answer all of them. You have 75 minutes to do so.
3. Turn off your cellphone, smartwatch and any other electronics, and place them in your bag. Then, place your bag at the front of the classroom for the duration of the exam.
4. During the exam, you should have your Cal1 card or government-issued ID with you. It will be required when submitting the exam.
5. Before you start writing your answers, write you student ID number at the top-right corner of each page of the exam.
6. Write your answers to each question in the space provided directly after the question, using a black or blue pen with non-erasable ink. If you run out of space you may continue your answer on the scratch paper provided at the end. In that case, clearly indicate where each answer starts and ends within the scratch paper. Clearly cross out any drafts or scribbles that should not be graded.
7. In your answers, you may rely on any claim proved in the lectures or discussions if you provide a full and clear statement of the cited claim. For all other claims you make, you must provide a full proof, including for claims we saw in the homework.
8. You must work alone, without consulting any people or other sources.

Write your full legal name and student ID here to indicate that you understand the instructions:

First name: _____

Last name: _____

Student ID: _____

Problem 1

Define the following two notions:

(a) Non-deterministic finite automaton (NFA). (5 pts)

An NFA N consists of:

- A finite set of states Q.
- A finite alphabet Σ such that $\epsilon \in \Sigma$.
- A transition function $\delta : Q \times \Sigma \to 2^Q$.
- An initial state $q_0 \in Q$.
- A set of accepting states $A \subseteq Q$.

(b) The NFA N accepts the input $x \in \{0,1\}^*$. (5 pts)

Solution 1:

An NFA N *accepts* an input $x \in \{0,1\}^*$ if there exists a sequence $q_0, q_1, q_2, \ldots, q_n \in Q$ such that:

- q_0 is the initial state of N.
- For all $1 \leq t \leq n$, we have $q_t \in E(\delta(q_{t-1}, x_t))$, where for $S \subseteq Q$, the set $E(S)$ is the set of states that can be reached from S using ϵ-transitions.
- $q_n \in A$.

Solution 2:

Let $N = (Q, \Sigma, \delta, q_0, A)$ be a NFA and let $x \in \{0,1\}^*$. We say that N *accepts* x if there exist a sequence $s_1, s_2, \ldots, s_n \in \{0,1,\varepsilon\}^*$ where $s = s_1 s_2 \ldots s_n$ and a sequence of states $r_0, \ldots r_n \in Q$ such that:

- $r_0 = q_0$.
- $\forall i \in [n] : \; r_i \in \delta(r_{i-1}, s_i)$.
- $r_n \in A$.

Problem 2

Let $L \subseteq \{0,1\}^*$ be an infinite regular language.

(a) Prove that there are two disjoint infinite regular languages $L_1, L_2 \subseteq \{0,1\}^*$ such that $L = L_1 \cup L_2$. (20 pts)

Since L is regular it has a pumping constant p. Since L is infinite, it contains some string s of length at least p. By the pumping lemma, we may partition that string such that $s = xyz$ with $|y| > 0$, and $xy^iz \in L$ for all integers $i \geq 0$. The language $L_1 = \{xy^{2i}z : i \geq 0\}$ is in L. The language L_1 is regular, seeing as it can be expresses using the regular expression $x(yy)^*z$. The language $\overline{L_1} = \{0,1\}^* \setminus L_1$ is regular, from the closure of regular language under complementation. The language $L_2 = L \cap \overline{L_1} = L \setminus L_1$ is also regular, from closure under intersection. It is also infinite, seeing as it contains all strings of the form $xy^{2i+1}z$ for $i \geq 0$.

(b) Prove that there is a language $L' \subseteq L$ such that L' is not regular. (10 pts)

Suggestion: Show that there exists a sequence of strings $s_i \in L$ such that $|s_{i+1}| \geq |s_i|^2$ for all i. Now recall the Polynomials question from Problem Set 2.

Solution 1:

Define $i_1, i_2, \ldots$ inductively as follows. Let i_1 be the smallest integers such that $L \cap \Sigma^{i_1}$ is non empty and $i_1 \geq 2$; such an integer exists since L is non empty. Given i_j, let i_{j+1} be the smallest integer that is larger than i_j^2 such that $L \cap \Sigma^{i_{j+1}}$ is non empty; such an integer exists since L is infinite.

For each j, choose $s_j \in \Sigma^{i_j} \cap L$; such s_j exists by construction of i_j.

Let $L' = \{s_1, s_2, \ldots\}$. This is an infinite set such that $|s_{j+1}| \geq |s_j|^2$ for all integers $j \geq 1$. Assume towards a contradiction that L' is regular. By the pumping lemma, we can choose a string $s_j = xyz \in L'$, $|s_j| > 3$ such that such that $xy^jz \in L'$ for all integers $j \geq 0$, and $|y| > 0$. Observe that

$$|s_j| = |xyz| < |xy^2z| = |s_j| + |y| \leq 2|s_j| < |s_j|^2 \leq |s_{j+1}|,$$

by the construction of L' it holds that $xy^2z \notin L'$. This is a contradiction.

Solution 2:

We already saw that there is a string $xyz \in L$ such that $xy^iz \in L$ for all integers $i \geq 0$. The language $L' = \{xy^{i^2}z : i \geq 0\}$ is contained in L. Assume towards a contradiction that L' is regular. By the pumping lemma, there is a string $x'y'z' \in L'$ such that $x'y'^jz' \in L'$ for all integer $j \geq 0$. The length of the strings of the form $x'y'^jz'$ grows linearly in j, so it can not be in L' for all j', because the length of the strings in L' grows quadratically.

Remark: the language L cannot be assumed to be all of $\{0,1\}^*$.

Problem 3

Let $n > 0$ be an integer, and recall that $\circ$ denotes concatenation. For a string $x = x_1 \circ x_2 \in \{0,1\}^{2n}$ where $x_1, x_2 \in \{0,1\}^n$, let

$$\mathsf{swap}(x) = x_2 \circ x_1.$$

Example: For $n = 3$, $\mathsf{swap}(011001) = 001011$.

Consider the language

$$\mathsf{SWAP} = \left\{x \in \{0,1\}^{2n} : \ x = \mathsf{swap}(x)\right\}.$$

What is the minimal size for a DFA that decides SWAP? Prove your answer. (30 pts)

We use the Myhill-Nerode theorem. Let $\approx$ denote the equivalence relation on string in $\{0,1\}^*$ that SWAP defines. That is, $x \approx y$ if for all $z \in \{0,1\}^*$, the string $x \circ z$ is in SWAP if and only if the string $y \circ z$ is in SWAP.

We say that $x \in \{0,1\}^*$ *can be extended to be in* SWAP if there exists a $z \in \{0,1\}^*$ such that $x \circ z \in$ SWAP. First, we claim that if $x, y \in \{0,1\}^*$ can be extended to be in SWAP and $|x| \neq |y|$ then $x \not\approx y$. Indeed, there exists a z such that $x \circ z \in$ SWAP, but $y \circ z$ does not have length $2n$, so it is not in SWAP.

We see that the equivalence classes $\{0,1\}^*/\approx$ that contain inputs that can be extended to be in SWAP can be partitioned according to length. Let E_m be the collection of equivalence classes with accepting inputs of length m.

For $m \leq n$, we claim that $|E_m| = 2^m$. Indeed, if $x, y \in \{0,1\}^m$ are distinct then $x0^{n-m}x0^{n-m} \in$ SWAP but $y0^{n-m}x0^{n-m} \notin$ SWAP.

For $n < m \leq 2n$, we claim that $|E_m| = 2^{2n-m}$. First, for every $x \in \{0,1\}^m$ that can be extended to be in SWAP, there is exactly one z_x such that $x \circ z_x \in$ SWAP. Thus, $x, y \in \{0,1\}^m$ that can be extended to be in SWAP are in the same equivalence class if and only if $z_x = z_y$. Furthermore, for each suffix z of length $2n - m$ there exists a unique equivalence class for which z is the accepting suffix. The number of different equivalence classes is thus 2^{2n-m}, which is the number of options to choose such a suffix z of length $2n - m$.

There is one additional equivalence class that contains all strings that cannot be extended to be in SWAP. They are all equivalent because no suffix can complete any of them to a string in the language.

The number of different equivalence classes is therefore

$$\sum_{m=0}^{n} 2^m + \sum_{m=n+1}^{2n} 2^{2n-m} + 1 = 2^{n+1} - 1 + 2^n - 1 + 1 = 3 \cdot 2^n - 1.$$

By the theorem, this is the size of the language, i.e. the size of a minimal DFA that decides it. □

Note: In the above proof we showed that the number of equivalence classes is precisely $3 \cdot 2^n - 1$. We did this by presenting a set of equivalence classes and showing that all strings in $\{0,1\}^*$ belong to one of the classes we presented. Many students proved that some set contains pairwise distinguishable strings, but this only establishes a lower bound. Such a lower bound may not be tight because it is possible that some larger set exists that would also be pairwise distinguishable, entailing a larger lower bound.

The lower bound was worth 20 points, and the upper bound was worth 10.

One common answer, which only proved a (not-tight) lower bound of 2^n without establishing an upper bound, generally received 15 points.

Problem 4

A bipartite graph $G = (V, E)$ is a graph in which V can be partitioned into two disjoint sets V_1 and V_2 such that all edges $e \in E$ are of the form $e = \{v_1, v_2\}$ with $v_1 \in V_1$ and $v_2 \in V_2$.

Fix an even integer $n > 2$. Let $V = \{1, 2, \ldots, n\}$. Let Σ be the family of all sets $e \subset V$ of size $|e| = 2$. Every string $S = (e_1, e_2, \ldots, e_m) \in \Sigma^m$ defines a graph (V, E) with vertex-set V and edge-set $E = \{e_1, e_2, \ldots, e_m\}$. Let Bipartite be the streaming problem of deciding whether the graph defined by a string $S \in \Sigma^m$ is bipartite.

(a) Define: Two strings $S_1, S_2 \in \Sigma^*$ are distinguishable with respect to Bipartite. (5 pts)

Two strings $S_1, S_2 \in \Sigma^*$ are distinguishable with respect to Bipartite if there is a sting $z \in \Sigma^*$ such that $S_1 \circ z \in$ Bipartite and $S_2 \circ z \notin$ Bipartite, or vice versa. (The additional requirement that the strings be of the same length is good, but no points were reduced if it was omitted)

(b) Prove that there are strings $S_1, \ldots, S_t \in \Sigma^*$ for $t \geq \binom{n}{n/2}/2$ that are pairwise distinguishable with respect to Bipartite. (25 pts)

For every $U \subset V$ of size $|U| = n/2$ such that $1 \in U$, define S_U as a string of length $n^2/4$ that encodes the complete bipartite graph G_U with sides U and $V \setminus U$. The number of such strings is exactly $\binom{n}{n/2}/2$. First, every two such strings encode different graphs (we can reconstruct the set U from the graph G_U). Second, every graph G_U is maximal in the sense that adding a single edge to G_U makes it non-bipartite. It follows that if $U \neq U'$ are two sets as above, then there is an edge e that is in G_U but not in $G_{U'}$. It follows that $S_U \circ S_U$ is in Bipartite but $S_U \circ S_{U'}$ is not in Bipartite.