CS181 Winter 2015 - Problem Set #2 Solutions

1. (**30 points**) Prove that the following language is not regular without using pumping lemma (but you can use the examples we showed in the class and the discussion section to be irregular).

$$L = \{0^a 1^b \ : \ \forall a, b \in \mathbb{N}, a^2 + 3b = ab + 3a\}$$

(Hint: Use the closure properties of regular languages.)

Solution. We prove this by contradiction. Suppose, L is indeed regular. We can rewrite the equation $a^2 + 3b = ab + 3a$ as $a(a-3) - b(a-3) = 0$, which can further be rewritten as $(a-b)(a-3) = 0$. The only values of a and b that satisfy this equation are when $a = b$ and $a = 3$. Using this fact, we can now rewrite L as

$$L = L_1 \cup L_2,$$

where $L_1 = \{0^n 1^n : \forall n \in \mathbb{N} \backslash \{3\}\}$ and $L_2 = \{0^3 1^*\}$. It can be verified that the language L_2 is regular. Now, $L_1 = L \backslash L_2$ is also regular, from the fact that the regular languages are closed under set difference operation. In the next step, we observe the following fact:

$$L_4 = \{0^n 1^n : \forall n \in \mathbb{N}\} = L_1 \cup \{0^3 1^3\}$$

We have already observed that L_1 is regular, from our hypothesis, and we note that the language $\{0^3 1^3\}$ is regular. From the fact that the regular languages are closed under union operation, we have that L_4. But, we have already seen in the class that $L_4 = \{0^n 1^n \ : \ \forall n \in \mathbb{N}\}$ is irregular. Thus, we have arrived at a contradiction!

2. (**40 points**) For a language L over alphabet Σ, we define

$$L_{\frac{1}{3}-\frac{1}{3}} = \{xz \in \Sigma^* \mid \exists y \in \Sigma^* \text{ with } |x| = |y| = |z| \text{ such that } xyz \in L\}.$$

Prove that if L is regular, then $L_{\frac{1}{3}-\frac{1}{3}}$ need not be regular. (*Hint: Consider* 0^*21^* *and recall that the regular languages are closed under intersection.*)

Solution. Consider $L = 0^*21^*$. Let $xyz \in L$. Consider the middle third y of w. There are three interesting cases: y contains the 2, y is completely composed of 0s, y is completely composed of 1s. In the second two cases, xz must contain a 2. However, when y contains the 2, then x is composed completely of 0s and z is composed completely of 1s. Moreover, $|x| = |z|$. Then, when y contains the 2, $xz = 0^i 1^i$ for some i!

Assume, by way of contradiction that $L_{\frac{1}{3}-\frac{1}{3}}$ were regular. Since regular languages are closed under intersection, we know that

$$L_{\frac{1}{3}-\frac{1}{3}} \cap \{0,1\}^* = \{0^n 1^n \mid n \geq 0\} = L_{eq}$$

must be regular. However, we've shown in class that L_{eq} is not regular! This contradiction tells us that $L_{\frac{1}{3}-\frac{1}{3}}$ must not be regular.

We've now given an example where L is regular, but $L_{\frac{1}{3}-\frac{1}{3}}$ is not regular.

3. (**15 points**) Prove the following more general form of the pumping lemma: For any regular language L, there exists a number $n > 0$, such that, for any string $x \in L$ with $x = x_1 x_2 \cdots x_m$ and $m > n$, the following holds. For any set of indices S that is a subset of the set $\{1, \ldots, m\}$ with the size of S being at least n, we have that

   (a) x can be written as uvw, and we will denote v to be the substring $x_i x_{i+1} ... x_j$, where $i \geq 1$ and $j \leq m$.

   (b) There exists at least one k in the set S such that k lies in the range $[i, j]$, i.e, $k \in \{i, i+1, \ldots, j\}$.

   (c) For all $i > 0$, we have $uv^i w \in L$. That is, we can "pump" the substring v arbitrary number of times and the resulting string still belongs to the language L.

   Solution. Since L is regular, consider a DFA for L having n number of states. Consider a string x with length $m \geq n$. Let $S = \{i_1, \ldots, i_\ell\}$ be a subset of indices in $\{1, \ldots, m\}$ with $|S| = \ell$ and $\ell > n$. Let Q' be a set of states $\{q_1, q'_1, \ldots, q_\ell, q'_\ell\}$ such that q_a transitions to q'_a on input x_{i_a}, for all $a \in \{1, \ldots, \ell\}$. By pigeonhole principle, we have that there exists $a, b \in \{1, \ldots, \ell\}$ and $a \neq b$ such that $q_a = q_b$. Denote by string v that transitions from q_a to q_b. Further denote by u the string that transitions from q_0 to q_a and w by the string that transitions from q_b to the accepting state. And hence, we represent uvw by x. By the way we have chosen u, v, and w we have the following properties:

   - v contains at least one element from the set S (one such element is x_{i_a}).
   - $uv^i w \in L$. This follows from the fact that there is a path from q_a to itself on input string v.

4. (**15 points**) Prove that the language $L = \{00^*1^n2^n \ : \ \forall n \in \mathbb{N}\}$ is irregular using the above general form of the pumping lemma.

   Solution. Suppose not. Then, by the generalized pumping lemma stated above, we have that there exists some $n \geq 0$ such that the statement of the pumping lemma holds. Let $x = 01^n2^n$. It can be seen that x is in L. Further, denote the set S to be $\{2, \ldots, n+1\}$. Note that S denote the set of indices where x has 1. Suppose x can be decomposed as uvw. There are five cases for v, namely,

   (a) v contains only 1s: In this case, $uv^2w \notin L$, because the number of 1s is different from the number of 2s.

   (b) v contains 1s and 2s: In this case, $uv^2w \notin L$, because the resulting string will not be of the form $00^*1^n2^n$.

   (c) v contains 0 and 1s: Even in this case, as in the previous case, the resulting string will not be of the form $00^*1^n2^n$.

   (d) v contains 0, 1s and 2s: This is same as the previous two cases.

(e) v contains only 0s or only 2s: This violates our rule that v should contain at least one index from S.

The above cases contradict the fact that $uv^iw \in L$ for all $i \geq 0$. This shows that L is irregular.