1 Prove the languages are not regular:

Prove that each of the following languages is ***not*** regular.

1. $\{\mathtt{0}^{2n}\mathtt{1}^n \mid n \geq 0\}$

Solution (verbose): Let F be the language $\mathtt{0}^*$.

Let x and y be arbitrary strings in F.

Then $x = \mathtt{0}^i$ and $y = \mathtt{0}^j$ for some non-negative integers $i \neq j$.

Let $z = \mathtt{0}^i\mathtt{1}^i$.

Then $xz = \mathtt{0}^{2i}\mathtt{1}^i \in L$.

And $yz = \mathtt{0}^{i+j}\mathtt{1}^i \notin L$, because $i + j \neq 2i$.

Thus, F is a fooling set for L.

Because F is infinite, L cannot be regular. ■

Solution (concise): For all non-negative integers $i \neq j$, the strings $\mathtt{0}^i$ and $\mathtt{0}^j$ are distinguished by the suffix $\mathtt{0}^i\mathtt{1}^i$, because $\mathtt{0}^{2i}\mathtt{1}^i \in L$ but $\mathtt{0}^{i+j}\mathtt{1}^i \notin L$. Thus, the language $\mathtt{0}^*$ is an infinite fooling set for L. ■

Solution (concise, different fooling set): For all non-negative integers $i \neq j$, the strings $\mathtt{0}^{2i}$ and $\mathtt{0}^{2j}$ are distinguished by the suffix $\mathtt{1}^i$, because $\mathtt{0}^{2i}\mathtt{1}^i \in L$ but $\mathtt{0}^{2j}\mathtt{1}^i \notin L$. Thus, the language $(\mathtt{00})^*$ is an infinite fooling set for L. ■

2. $\{\mathtt{0}^m\mathtt{1}^n \mid m \neq 2n\}$

Solution (verbose): Let F be the language $\mathtt{0}^*$.

Let x and y be arbitrary strings in F.

Then $x = \mathtt{0}^i$ and $y = \mathtt{0}^j$ for some non-negative integers $i \neq j$.

Let $z = \mathtt{0}^i\mathtt{1}^i$.

Then $xz = \mathtt{0}^{2i}\mathtt{1}^i \notin L$.

And $yz = \mathtt{0}^{i+j}\mathtt{1}^i \in L$, because $i + j \neq 2i$.

Thus, F is a fooling set for L.

Because F is infinite, L cannot be regular. ■

Solution (concise, different fooling set): For all non-negative integers $i \neq j$, the strings $\mathtt{0}^{2i}$ and $\mathtt{0}^{2j}$ are distinguished by the suffix $\mathtt{1}^i$, because $\mathtt{0}^{2i}\mathtt{1}^i \notin L$ but $\mathtt{0}^{2j}\mathtt{1}^i \in L$. Thus, the language $(\mathtt{00})^*$ is an infinite fooling set for L. ■

3. $\left\{\mathtt{0}^{2^n} \;\middle|\; n \geq 0\right\}$

Solution (verbose): Let $F = L = \left\{\mathtt{0}^{2^n} \;\middle|\; n \geq 0\right\}$.

Let x and y be arbitrary elements of F.

Then $x = \mathtt{0}^{2^i}$ and $y = \mathtt{0}^{2^j}$ for some non-negative integers x and y.

Let $z = \mathtt{0}^{2^i}$.

Then $xz = \mathtt{0}^{2^i}\mathtt{0}^{2^i} = \mathtt{0}^{2^{i+1}} \in L$.

And $yz = \mathtt{0}^{2^j}\mathtt{0}^{2^i} = \mathtt{0}^{2^i+2^j} \notin L$, because $i \neq j$

Thus, F is a fooling set for L.

Because F is infinite, L cannot be regular. ■

Solution (concise): For any non-negative integers $i \neq j$, the strings $\mathtt{0}^{2^i}$ and $\mathtt{0}^{2^j}$ are distinguished by the suffix $\mathtt{0}^{2^i}$, because $\mathtt{0}^{2^i}\mathtt{0}^{2^i} = \mathtt{0}^{2^{i+1}} \in L$ but $\mathtt{0}^{2^j}\mathtt{0}^{2^i} = \mathtt{0}^{2^i+2^j} \notin L$. Thus L itself is an infinite fooling set for L. ■

4. Strings over $\{\mathtt{0}, \mathtt{1}\}$ where the number of $\mathtt{0}$s is exactly twice the number of $\mathtt{1}$s.

Solution (verbose): Let F be the language $\mathtt{0}^*$.

Let x and y be arbitrary strings in F.

Then $x = \mathtt{0}^i$ and $y = \mathtt{0}^j$ for some non-negative integers $i \neq j$.

Let $z = \mathtt{0}^i\mathtt{1}^i$.

Then $xz = \mathtt{0}^{2i}\mathtt{1}^i \in L$.

And $yz = \mathtt{0}^{i+j}\mathtt{1}^i \notin L$, because $i + j \neq 2i$.

Thus, F is a fooling set for L.

Because F is infinite, L cannot be regular. ■

Solution (concise, different fooling set): For all non-negative integers $i \neq j$, the strings $\mathtt{0}^{2i}$ and $\mathtt{0}^{2j}$ are distinguished by the suffix $\mathtt{1}^i$, because $\mathtt{0}^{2i}\mathtt{1}^i \in L$ but $\mathtt{0}^{2j}\mathtt{1}^i \notin L$. Thus, the language $(\mathtt{00})^*$ is an infinite fooling set for L. ■

Solution (closure properties): If L were regular, then the language

$$L \cap \mathtt{0}^*\mathtt{1}^* \;=\; \left\{\mathtt{0}^{2n}\mathtt{1}^n \;\middle|\; n \geq 0\right\}$$

would also be regular since regular languages are closed under intersection but we have seen in Problem 1 that $\left\{\mathtt{0}^{2n}\mathtt{1}^n \;\middle|\; n \geq 0\right\}$ is not regular.

Another solution based on closure properties. If L were regular, then the language

$$\big((\mathtt{0}+\mathtt{1})^* \setminus L\big) \cap \mathtt{0}^*\mathtt{1}^* \;=\; \{\mathtt{0}^m\mathtt{1}^n \mid m \neq 2n\}$$

would also be regular, because regular languages are closed under complement and intersection. But we just proved that $\{\mathtt{0}^m\mathtt{1}^n \mid m \neq 2n\}$ is not regular in problem 2. *[Yes, this proof would be worth full credit, either in homework or on an exam.]*

Note that the proofs based on closure properties relied on non-regularity of some previously known languages. One could also think of the proofs as allowing you to simplify the initial language to a more structured one which may be easier to work with. ■

5. Strings of properly nested parentheses `()`, brackets `[]`, and braces `{}`. For example, the string `([]){}` is in this language, but the string `([)]` is not, because the left and right delimiters don't match.

Solution (verbose): Let F be the language $\texttt{(}^*$.

Let x and y be arbitrary strings in F.

Then $x = \texttt{(}^i$ and $y = \texttt{(}^j$ for some non-negative integers $i \neq j$.

Let $z = \texttt{)}^i$.

Then $xz = \texttt{(}^i\texttt{)}^i \in L$.

And $yz = \texttt{(}^j\texttt{)}^i \notin L$, because $i \neq j$.

Thus, F is a fooling set for L.

Because F is infinite, L cannot be regular. ■

Solution (concise): For any non-negative integers $i \neq j$, the strings $\texttt{(}^i$ and $\texttt{(}^j$ are distinguished by the suffix $\texttt{)}^i$, because $\texttt{(}^i\texttt{)}^i \in L$ but $\texttt{(}^i\texttt{)}^j \notin L$. Thus, the language $\texttt{(}^*$ is an infinite fooling set. ■

Solution (closure properties): If L were regular, then the language $L \cap \texttt{(}^*\texttt{)}^* = \{\texttt{(}^n\texttt{)}^n \mid n \geq 0\}$ would be regular. The language $\{\texttt{(}^n\texttt{)}^n \mid n \geq 0\}$ is the same as $\{\texttt{0}^n\texttt{1}^n \mid n \geq 0\}$ modulo changing the symbol names and is not regular from lecture. Thus L is not regular. ■

6. Strings of the form $w_1\#w_2\#\cdots\#w_n$ for some $n \geq 2$, where each substring w_i is a string in $\{0,1\}^*$, and some pair of substrings w_i and w_j are equal.

Solution (verbose): Let F be the language 0^*.

Let x and y be arbitrary strings in F.

Then $x = 0^i$ and $y = 0^j$ for some non-negative integers $i \neq j$.

Let $z = \#0^i$.

Then $xz = 0^i\#0^i \in L$.

And $yz = 0^j\#0^i \notin L$, because $i \neq j$.

Thus, F is a fooling set for L.

Because F is infinite, L cannot be regular. ■

Solution (concise): For any non-negative integers $i \neq j$, the strings 0^i and 0^j are distinguished by the suffix $\#0^i$, because $0^i\#0^i \in L$ but $0^j\#0^i \notin L$. Thus, the language 0^* is an infinite fooling set. ■

Work on these later:

7. w, such that $|w| = \lceil k\sqrt{k}\rceil$, for some natural number k.

   Hint: since this one is more difficult, we'll even give you a fooling set that works: try $F = \{\mathtt{0}^{m^6} | m \geq 1\}$. We'll also provide a bound that can help: the difference between consecutive strings in the language, $\lceil (k+1)^{1.5}\rceil - \lceil k^{1.5}\rceil$, is bounded above and below as follows
   $$1.5\sqrt{k} - 1 \leq \lceil (k+1)^{1.5}\rceil - \lceil k^{1.5}\rceil \leq 1.5\sqrt{k} + 3$$
   All that's left is you need to carefully prove that F is a fooling set for L.

Solution: Let F be the set $\{\mathtt{0}^{m^6} | m \in \mathbb{N}\}$.

We can also write this as $\{\mathtt{0}^{\lceil k\sqrt{k}\rceil} | k = m^4, m \in \mathbb{N}\}$. Note that each element in F is also an element in L.

Let $x = \mathtt{0}^{m^6}$ and $y = \mathtt{0}^{n^6}$ for some $m < n$.

Let z be the smallest string such that $xz \in L$. By the given bound, $|z| \leq 1.5m^2 + 3$.

Suppose for contradiction $yz \in L$. By the other side of the given bound, we would need $|z| \geq 1.5n^2 - 1$. We can show both of these contraints on z can't be satisfied, since $1 \leq m \leq n-1$, so

$$1.5m^2 + 3 \leq 1.5(n-1)^2 + 3 = 1.5(n^2 - 2n + 1) + 3 = 1.5n^2 - 1 + (5.5 - 3n) \leq 1.5n^2 - 1$$

.

■

Solution: From my experience in office hours, I wanted to write another solution which clarifies a few things (since this is a difficult problem).

First let's start with the fooling set $F = \{\mathtt{0}^{m^6} | m \geq 1\}$. This set is a subset of the language $L_{P5} = \{\mathtt{0}^{m^6} | m \in \mathbb{N}\}$ but that's ok for us. If we prove that F has infinite distinguishable states, then it means L_{P5} has at least infinite distinguishable states which is a problem for L_{P5} being regular.

So that's the big picture but how do we get there? Well first let's consider two strings from the fooling set:

$$x = \mathtt{0}^{i^6}$$
$$y = \mathtt{0}^{j^6}$$

for $i < j$. So both these strings are part of the original language (assuming $k = i^4 or k = j^4$). But what about the next string in their sequence? Is there another run of zeros (z) that you can add to x such that $xz \in L_{P5}$. More importantly if x and y are distinguishable then it means $yz \notin L_{P5}$? If $L_{GoforthScientificInc}$ is not regular, then we need to prove that such a z cannot exist which let's xz & $yz \in L_{P5}$.

So let's do a **Proof by Contradiction** as we do with most fooling set problems.

- First let's look at xz which is the next largest run of zeros after x that belongs to

L_{P5}.

- Looking at the definition for L_{P5}, in order for $x \in L_{P5}$, $k = i^4$ which give us the string $x = \mathtt{0}^{i^6} = \mathtt{0}^{(i^4)^{1.5}}$.
- So the next largest run of $\mathtt{0}$'s in L_{P5} occurs when $k = i^4 + 1$ which would give us the string $xz = \mathtt{0}^{(i^4+1)^{1.5}}$.
- This means that we can finding the length of z by

$$|xz| - |x| = |\mathtt{0}^{(i^4+1)^{1.5}}| - |\mathtt{0}^{(i^4)^{1.5}}| = (i^4+1)^{1.5} - (i^4+1)^{1.5} = |z|$$

- According to boundaries given in the problem this means that

$$1.5\sqrt{i^4} - 1 = 1.5i^2 - 1 \leq |z| \leq 1.5i^2 + 3 = 1.5\sqrt{i^4} + 3 \tag{1}$$

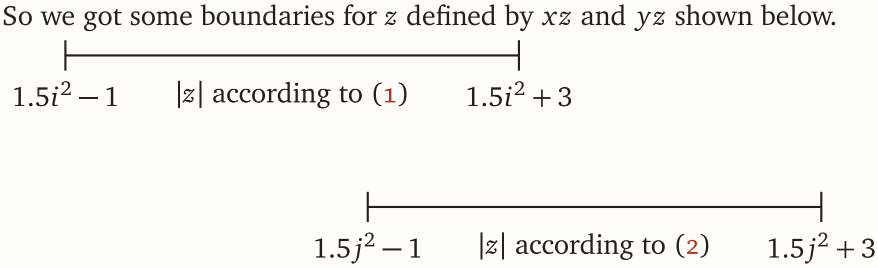
- Next, because of the proof by contradiction we're assuming $yz \in L_{P5}$ as well. This is the next largest run of zeros after y that is in L_{P5}. Here we follow the exact steps as above but with j instead of i.
  - Looking at the definition for L_{P5}, in order for $y \in L_{P5}$, $k = j^4$ which give us the string $y = \mathtt{0}^{j^6} = \mathtt{0}^{(j^4)^{1.5}}$.
  - The next largest run of $\mathtt{0}$'s in L_{P5} occurs when $k = j^4 + 1$ which would give us the string $yz = \mathtt{0}^{(j^4+1)^{1.5}}$.
  - This means that we can finding the length of z by

$$|yz| - |y| = |\mathtt{0}^{(j^4+1)^{1.5}}| - |\mathtt{0}^{(j^4)^{1.5}}| = (j^4+1)^{1.5} - (j^4+1)^{1.5} = |z|$$

  - According to boundaries given in the problem this means that

$$1.5j^2 - 1 \leq |z| \leq 1.5j^2 + 3 \tag{2}$$

- So we got some boundaries for z defined by xz and yz shown below.



Now if the states of x and y are not distinguishable (i.e. both xz and yz can be in $L_{P5)}$), then there should be some value of z that both prefixes can follow to an accept state. Namely,

$$1.5j^2 - 1 \leq |z| \leq 1.5i^2 + 3 \tag{3}$$

- But wait! Didn't we say $i < j$? If $i > 0$ then (3) is impossible!
- Therefore, there is run of zeroes for z where both xz and yz would be in L_{P5}.

- x and y denote distinguishable states states of the language L_{P5}.
- Because F is infinite, the DFA representing L_{P5} would require infinite states which violates the definition of regular language and hence, L_{P5} can't be regular.

■

8. $\left\{\mathtt{0}^{n^2} \;\middle|\; n \geq 0\right\}$

Solution: Let x and y be distinct arbitrary strings in L.

Without loss of generality, $x = \mathtt{0}^{2i+1}$ and $y = \mathtt{0}^{2j+1}$ for some $i > j \geq 0$.

Let $z = \mathtt{0}^{i^2}$.

Then $xz = \mathtt{0}^{i^2+2i+1} = \mathtt{0}^{(i+1)^2} \in L$

On the other hand, $yz = \mathtt{0}^{i^2+2j+1} \notin L$, because $i^2 < i^2 + 2j + 1 < (i+1)^2$.

Thus, z distinguishes x and y.

We conclude that L is an infinite fooling set for L, so L cannot be regular. ■

Solution: Let x and y be distinct arbitrary strings in $\mathtt{0}^*$.

Without loss of generality, $x = \mathtt{0}^i$ and $y = \mathtt{0}^j$ for some $i > j \geq 0$.

Let $z = \mathtt{0}^{i^2+i+1}$.

Then $xz = \mathtt{0}^{i^2+2i+1} = \mathtt{0}^{(i+1)^2} \in L$.

On the other hand, $yz = \mathtt{0}^{i^2+i+j+1} \notin L$, because $i^2 < i^2 + i + j + 1 < (i+1)^2$.

Thus, z distinguishes x and y.

We conclude that $\mathtt{0}^*$ is an infinite fooling set for L, so L cannot be regular. ■

Solution: Let x and y be distinct arbitrary strings in $\mathtt{0000}^*$.

Without loss of generality, $x = \mathtt{0}^i$ and $y = \mathtt{0}^j$ for some $i > j \geq 3$.

Let $z = \mathtt{0}^{i^2-i}$.

Then $xz = \mathtt{0}^{i^2} \in L$.

On the other hand, $yz = \mathtt{0}^{i^2-i+j} \notin L$, because

$$(i-1)^2 \;=\; i^2 - 2i + 1 \;<\; i^2 - i \;<\; i^2 - i + j \;<\; i^2.$$

(The first inequalities requires $i \geq 2$, and the second $j \geq 1$.)

Thus, z distinguishes x and y.

We conclude that $\mathtt{0000}^*$ is an infinite fooling set for L, so L cannot be regular. ■

9. $\{w \in (\mathtt{0} + \mathtt{1})^* \mid w \text{ is the binary representation of a perfect square}\}$

Solution: We design our fooling set around numbers of the form $(2^k+1)^2 = 2^{2k} + 2^{k+1} + 1 = \mathtt{1}\mathtt{0}^{k-2}\mathtt{1}\mathtt{0}^k\mathtt{1} \in L$, for any integer $k \geq 2$. The argument is somewhat simpler if we further restrict k to be even.

Let $F = \mathtt{1}(\mathtt{00})^*\mathtt{1}$, and let x and y be arbitrary strings in F.

Then $x = \mathtt{1}\mathtt{0}^{2i-2}\mathtt{1}$ and $y = \mathtt{1}\mathtt{0}^{2j-2}\mathtt{1}$, for some positive integers $i \neq j$.

Without loss of generality, assume $i < j$. (Otherwise, swap x and y.)

Let $z = \mathtt{0}^{2i}\mathtt{1}$.

Then $xz = \mathtt{1}\mathtt{0}^{2i-2}\mathtt{1}\mathtt{0}^{2i}\mathtt{1}$ is the binary representation of $2^{4i} + 2^{2i+1} + 1 = (2^{2i}+1)^2$, and therefore $xz \in L$.

On the other hand, $yz = \mathtt{1}\mathtt{0}^{2j-2}\mathtt{1}\mathtt{0}^{2i}\mathtt{1}$ is the binary representation of $2^{2i+2j} + 2^{2i+1} + 1$. Simple algebra gives us the inequalities

$$\begin{aligned}
(2^{i+j})^2 &= 2^{2i+2j} \\
&< \boldsymbol{2^{2i+2j} + 2^{2i+1} + 1} \\
&< 2^{2(i+j)} + 2^{i+j+1} + 1 \\
&= (2^{i+j}+1)^2.
\end{aligned}$$

So $2^{2i+2j} + 2^{2i+1} + 1$ lies between two consecutive perfect squares, and thus is not a perfect square, which implies that $yz \notin L$.

We conclude that F is a fooling set for L. Because F is infinite, L cannot be regular. ∎

2 Differentiate between regular and not regular

For each of the following languages over the alphabet $\Sigma = \{\mathtt{0}, \mathtt{1}\}$, either prove that the language is regular (by constructing a DFA or regular expression) or prove that the language is not regular (using fooling sets). Recall that Σ^+ denotes the set of all nonempty strings over Σ.

1. $L_a = \{\mathtt{0}^n\mathtt{1}^n w \mid w \in \Sigma^* \text{ and } n \geq 0\}$

Solution: The language is regular.When $n = 0$ the whole string is just represented by w.In this case all the strings over $\{0, 1\}$ can be just represented by w.So,we can ignore the $0^n 1^n$ portion as every string is covered by w.So every string in L_{2a} can be represented by w .

The regular expression for L_{2a} will be $(0 + 1)^*$.

Hence,the language L_a is a regular language. ■

2. $L_b = \{w\mathtt{0}^n w \mid w \in \Sigma^* \text{and } n > 0\}$

Solution: Let us consider the fooling set $F = \{1^n 0^n | n > 0\}$

Let x and y be arbitrary strings in F.

Then $x = \mathtt{1}^i\mathtt{0}^i$ and $y = \mathtt{1}^j\mathtt{0}^j$ for some positive integers $i \neq j$.

Let $z = \mathtt{1}^i$.

Then $xz = \mathtt{1}^i\mathtt{0}^i\mathtt{1}^i \in L_{2b}$.

And $yz = \mathtt{1}^j\mathtt{0}^j\mathtt{1}^i \notin L_{2b}$, because $i \neq j$.

Thus, F is a fooling set for L_{2b}.

Because F is infinite, L_b cannot be regular. ■

3. $L_c = \{xwwy | w, x, y \in \Sigma^+\}$

Solution: L_c is a regular language. The language only contains a limited number of strings that are not part of it. By the fact that any language of finite size is regular and regularity is preserved under complement. We can prove L_c is regular.

We can say that any string of length at least 4 is in the language. For an arbitrary string z of length at least 4, Let us define z as $xz'y$ where both a and b are a single symbol in Σ. So, z' has a of length at least 2.

x and y are two non-empty strings which are covered by $((\mathtt{0} + \mathtt{1})^+)$. z' will be ww. Since, our alphabet consists of just $\mathtt{0}$s and $\mathtt{1}$s, there are two cases. First, z' must be $\mathtt{00}$ or $\mathtt{11}$. This satisfies the constraint that there must be a repeating string in the middle with length at least one. Now, let us consider the case where we can have alternating $\mathtt{0}$'s and $\mathtt{1}$'s i.e, $w=$ $\mathtt{10}$ or $\mathtt{01}$. To include this case we add $\mathtt{1010}$ and $\mathtt{0101}$ to the regular expression.

Putting it all together you get the regular expression for L_{2c}: $(\mathtt{0} + \mathtt{1})^+(\mathtt{00} + \mathtt{11} + \mathtt{1010} + \mathtt{0101})(\mathtt{0} + \mathtt{1})^+$.

Hence, the language L_c is a regular language. ■

4. $L_d = \{xwwx | w, x \in \Sigma^+\}$

Solution: Let us consider the fooling set $F = \{1^n 0^n | n > 0\}$

Let x and y be arbitrary strings in F.

Then $x = \mathtt{1}^i\mathtt{0}^i$ and $y = \mathtt{1}^j\mathtt{0}^j$ for some positive integers $i \neq j$.

Let $z = \mathtt{0}^i\mathtt{1}^i$.

Then $xz = \mathtt{1}^i\mathtt{0}^i\mathtt{0}^i\mathtt{1}^i \in L_{2d}$.

And $yz = \mathtt{1}^j\mathtt{0}^j\mathtt{0}^i\mathtt{1}^i \notin L_{2d}$, because $i \neq j$.

Thus, F is a fooling set for L_{2d}.

Because F is infinite, L_d cannot be regular. ■