Theory of Computation

Pumping Lemma

Mayank Yadav

February 12, 2025

1 Introduction

Pumping Lemma describes an essential characteristic of regular languages. It says that a string with sufficiently long length belonging to a regular language can be divided into three sections, then the string obtained by pumping(or repeating) the middle section also belongs to the same language.

> Pumping Lemma states that for given a regular language L, there exists an integer $p \geq 1$ such that every string w having $\mid w \mid \geq p$ can be expressed as $w = xyz$ and the following conditions hold:
>
> $$\mid y \mid \geq 1$$
> $$\mid xy \mid \leq p$$
> $$(\forall n \geq 0)(xy^n z \in L)$$
>
> Mathematically,
>
> $$\forall L \subseteq \Sigma^*, \text{ regular}(L)$$
> $$\implies \exists p \geq 1, \forall w \in L, \mid w \mid \geq p$$
> $$\implies \exists x, y, z \in \Sigma^*, (w = xyz) \land (\mid y \mid \geq 1) \land (\mid xy \mid \leq p) \land (\forall n \geq 0, xy^n z \in L)$$

We can get intuitions behind this lemma by considering a DFA that accepts a regular language L (as the language is regular we can construct a DFA), let w be a string in L having length at least p say n. Let A,B and C be some of the states through which w goes through. As w is in the language, it will be accepted by DFA so C is a final state. As each character corresponds to a transition, the number of characters is equal to the number of transitions, hence there are n transitions so $n+1$ states from initial to the final state. As $n \geq p \implies n+1 > p$, hence according to pigeonhole principle there is a repetition of one state and say it is B.

We can divide the string w into x, y, z; x being the part before reaching B, y being the part between two occurrences of B and z being the part after B.

We can now check the conditions given in the lemma, $\mid y \mid \geq 1$ as that is the part that occurs between two repetition of B, $\mid xy \mid \leq p$ as a repetition occurs in $p+1$ states and third condition holds as y is being handled in the loop, so for example in $xyyz$, the both y get handled by the loop "read y".

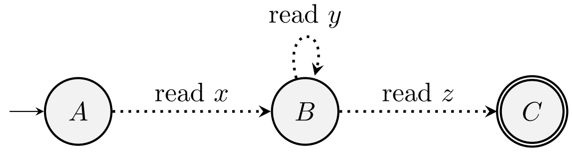

Figure 1: Illustration of lemma with DFA

2 Proof:

Proof. Let $D = (Q, \Sigma, \delta, q_0, F)$ be a DFA accepting L and p be the number of states of D Let $w = w_1 w_2 \cdots w_n \in L$ and $\mid w \mid = n$ and $n \geq p$. Let the sequence of states w go through be $r_1, r_2 \cdots r_{n+1}$

$$r_{i+1} = \delta(r_i, w_i) \forall i \in [1, n]$$

According to pigeonhole principle($p+1$ pigeons and p pigeonholes) , the first $p+1$ states must have a repetition of a state, let the first occurrence of the state r_a and the occurrence be r_b and also $b \leq p+1$.

Now let

$$x = w_1 \cdots w_{a-1}$$

$$y = w_a \cdots w_{b-1}$$

$$z = w_b \cdots w_n$$

For $i \geq 0$, $w' = xy^i z$, x takes D from r_1 to r_a, y takes D from r_a to r_a and z takes D from r_a to $r_n + 1$, hence D must accept w' . Furthermore,

$$a \neq b \implies \mid y \mid > 0$$

$$b \leq p+1 \implies b - 1 \leq p \implies \mid xy \mid \leq p$$

□

3 Application:

Pumping lemma can be used to prove that a language is not regular by ”Proof by Contradiction”. We assume that the language is regular and choose a string whose length is greater than the pumping constant and that can't be pumped and we reach a contradiction by failing to divide the string into any x, y, z and then put forward that the language is not regular.

It is important to note that the converse of pumping lemma is not true i.e. a language that satisfies these conditions is not necessarily a regular language.

3.1 Example 1

Prove that the language $L = \{0^n 1^n \mid n \in \mathbb{N}\}$ is not regular.

Proof. Let us assume L to be a regular language. Now according to pumping lemma $\exists(p \geq 1)$. Let the string $w \in L$ and

$$w = 0^p 1^p$$

$$\mid w \mid = 2p > p$$

$$w = xyz$$

As $\mid xy \mid \leq p$, $xy = 0^q$ where $q = \mid xy \mid \leq p$. Let $y = 0^k$ and $x = 0^{q-k}$ where $k \in [1, q]$ and $z = 0^{p-q} 1^p$. Now by statement 3 of lemma,

$$(\forall n \geq 0)(xy^n z) \in L$$

$$\implies 0^{q-k} 0^{kn} 0^{p-q} 1^p \in L$$

$$\implies 0^{k(n-1)+p} 1^p \in L$$

Since $k \in [1, p]$, $k(n-1) + p \neq p$ hence $0^{k(n-1)+p} 1^p \notin L$. A contradiction!

Hence, the initial assumption was wrong and the language L is not regular. □

3.2 Example 2

Prove that the language $L = \{ss \mid s \in \{0,1\}^*\}$ is not regular.

Proof. Let us assume L to be a regular language. Now according to pumping lemma $\exists(p \geq 1)$. Let the string $w \in L$ and

$$w = 0^p10^p1$$

$$\mid w \mid = 2p + 2 > p$$

$$w = xyz$$

As $\mid xy \mid \leq p$, $xy = 0^q$ where $q = \mid xy \mid \leq p$. Let $y = 0^k$ and $x = 0^{q-k}$ where $k \in [1, q]$ and $z = 0^{p-q}10^p1$. Now by statement 3 of lemma,

$$(\forall n \geq 0)(xy^nz) \in L$$

$$\implies 0^{q-k}0^{kn}0^{p-q}10^p1 \in L$$

$$\implies 0^{k(n-1)+p}10^p1 \in L$$

Since $k \in [1, p]$, $k(n-1) + p \neq p$ hence $0^{k(n-1)+p}1^p \notin L$. A contradiction!

Hence, the initial assumption was wrong and the language L is not regular. □

3.3 Example 3

Prove that the language $L = \{1^{n^2} \mid n \geq 0\}$ is not regular.

Proof. Let us assume L to be a regular language. Now according to pumping lemma $\exists(p \geq 1)$. Let the string $w \in L$ and

$$w = 1^{p^2}$$

$$\mid w \mid = p^2 > p$$

$$w = xyz$$

As $\mid xy \mid \leq p$, $xy = 1^q$ where $q = \mid xy \mid \leq p$. Let $y = 1^k$ and $x = 1^{q-k}$ where $k \in [1, q]$ and $z = 1^{p^2-q}$. Now by statement 3 of lemma, for $n = 2$

$$(xy^2z) \in L$$

$$\implies 1^{q-k}1^{2k}1^{p^2-q} \in L$$

$$\implies 1^{p^2+k} \in L$$

$p^2 + k$ must be a perfect square. $1 \leq k \leq q$ and $q \leq p$ implies $1 \leq k \leq p$

$$(1 \leq k \leq q) \wedge (q \leq p) \implies 1 \leq k \leq p$$

The next perfect square after p^2 is $(p+1)^2$,so there are no perfect square in $(p^2, (p+1)^2)$, Also, $(p+1)^2 - p^2 = 2p + 1$, But

$$k \leq p$$

$$k < 2p$$

$$k < 2p + 1$$

So, $1^{p^2+k} \notin L$. A contradiction!

Hence, the initial assumption was wrong and the language L is not regular. □

3.4 Example 4

Prove that the language $L = \{0^i 1^j \mid i > j\}$ is not regular.

Proof. Let us assume L to be a regular language. Now according to pumping lemma $\exists(p \geq 1)$. Let the string $w \in L$ and

$$w = 0^p 1^{p-1}$$

$$\mid w \mid = 2p - 1 \geq p$$

$$w = xyz$$

As $\mid xy \mid \leq p$, $xy = 0^q$ where $q = \mid xy \mid \leq p$. Let $y = 0^k$ and $x = 0^{q-k}$ where $k \in [1, q]$ and $z = 0^{p-q} 1^{p-1}$. Now by statement 3 of lemma, for $n = 0$

$$(xy^0 z) \in L$$

$$\implies 0^{q-k} 0^0 0^{p-q} 1^{p-1} \in L$$

$$\implies 0^{p-k} 1^{p-1} \in L$$

Since $k \in [1, p]$, $p - k \leq p - 1$ hence $0^{p-k} 1^{p-1} \notin L$. A contradiction!

Hence, the initial assumption was wrong and the language L is not regular. □

3.5 Example 5

Prove that the language $L = \{0^n 1^n 2^n \mid n \geq 0\}$ is not regular.

Proof. Let us assume L to be a regular language. Now according to pumping lemma $\exists(p \geq 1)$. Let the string $w \in L$ and

$$w = 0^p 1^p 2^p$$

$$\mid w \mid = 3p > p$$

$$w = xyz$$

As $\mid xy \mid \leq p$, $xy = 0^q$ where $q = \mid xy \mid \leq p$. Let $y = 0^k$ and $x = 0^{q-k}$ where $k \in [1, q]$ and $z = 0^{p-q} 1^p 2^p$. Now by statement 3 of lemma,

$$(\forall n \geq 0)(xy^n z) \in L$$

$$\implies 0^{q-k} 0^{kn} 0^{p-q} 1^p 2^p \in L$$

$$\implies 0^{k(n-1)+p} 1^p 2^p \in L$$

Since $k \in [1, p]$, $k(n-1) + p \neq p$ hence $0^{k(n-1)+p} 1^p 2^p \notin L$. A contradiction!

Hence, the initial assumption was wrong and the language L is not regular. □

3.6 Example 6

Prove that the language $L = \{sss \mid s \in \{0, 1\}^*\}$ is not regular.

Proof. Let us assume L to be a regular language. Now according to pumping lemma $\exists(p \geq 1)$. Let the string $w \in L$ and

$$w = 0^p 1 0^p 1 0^p 1$$

$$\mid w \mid = 3p + 3 > p$$

$$w = xyz$$

As $\mid xy \mid \leq p$, $xy = 0^q$ where $q = \mid xy \mid \leq p$. Let $y = 0^k$ and $x = 0^{q-k}$ where $k \in [1, q]$ and $z = 0^{p-q} 1^p 0^p 1$. Now by statement 3 of lemma,

$$(\forall n \geq 0)(xy^n z) \in L$$

$$\implies 0^{q-k} 0^{kn} 0^{p-q} 1 0^p 1 0^p 1 \in L$$

$$\implies 0^{k(n-1)+p} 1 0^p 1 0^p 1 \in L$$

Since $k \in [1, p]$, $k(n-1) + p \neq p$ hence $0^{k(n-1)+p} 1 0^p 1 0^p 1 \notin L$. A contradiction!

Hence, the initial assumption was wrong and the language L is not regular. □

3.7 Example 7

Prove that the language $L = \{1^{2^n} \mid n \geq 0\}$ is not regular.

Proof. Let us assume L to be a regular language. Now according to pumping lemma $\exists(p \geq 1)$. Let the string $w \in L$ and

$$w = 1^{2^p}$$

$$\mid w \mid = 2^p > p$$

$$w = xyz$$

As $\mid xy \mid \leq p$, $xy = 1^q$ where $q = \mid xy \mid \leq p$. Let $y = 1^k$ and $x = 1^{q-k}$ where $k \in [1, q]$ and $z = 1^{2^p - q}$. Now by statement 3 of lemma, for $n = 2$

$$(xy^2z) \in L$$

$$\implies 1^{q-k}1^{2k}1^{2^p-q} \in L$$

$$\implies 1^{2^p+k} \in L$$

$2^p + k$ must be a power of 2. But, as $k \in [1, p]$

$$2^p > k$$

$$2^p + 2^p > k + 2^p$$

$$2^{p+1} > 2^p + k$$

And no power of 2 lies in $(2^p, 2^{p+1})$. So, $1^{2^p+k} \notin L$. A contradiction!

Hence, the initial assumption was wrong and the language L is not regular. □

References

[1] Michael Sipser, *Introduction to the Theory of Computation*, Cengage Learning, 2012.

[2] Wikipedia contributors, *Pumping lemma for regular languages — Wikipedia, The Free Encyclopedia*, 2024, https://en.wikipedia.org/w/index.php?title=Pumping_lemma_for_regular_languages&oldid=1264121825,

[3] Stanford University, *CS103: Mathematical Foundations of Computing*, https://web.stanford.edu/class/cs103/