Some languages are *not* regular

When is a language is regular?

if we are able to construct one of the following: DFA *or* NFA *or* ε -NFA *or* regular expression

When is it not?

If we can show that no FA can be built for a language

How to prove languages are ***not*** regular?

What if we cannot come up with any FA?

A) Can it be language that is not regular?

B) Or is it that we tried wrong approaches?

How do we *decisively* prove that a language is not regular?

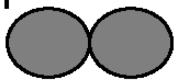
> *"The hardest thing of all is to find a black cat in a dark room, especially if there is no cat!" -Confucius*

Example of a non-regular language

Let L = {w | w is of the form 0^n1^n , for all $n \geq 0$}

- *<u>Hypothesis:</u> L is not regular*
- <u>Intuitive rationale:</u>   How do you keep track of a running count in an FA?
- <u>A more formal rationale:</u>
  - By contradition, if L is regular then there should exist a DFA for L.
  - Let k = number of states in that DFA.
  - Consider the special word w= 0^k1^k => w $\in$ L
  - DFA is in some state p_i, after consuming the first i symbols in w

Rationale…

- Let $\{p_0, p_1, \ldots\ p_k\}$ be the sequence of states that the DFA should have visited after consuming the first k symbols in w which is 0^k
- But there are only k states in the DFA!
- ==> at least one state should repeat somewhere along the path (by  +  Principle)
- ==> Let the repeating state be $p_i = p_J$ for $i < j$
- ==> We can fool the DFA by inputing $0^{(k-(j-i))}1^k$ and still get it to accept (note: k-(j-i) is at most k-1).
- ==> DFA accepts strings w/ unequal number of 0s and 1s, implying that the DFA is wrong!

The Pumping Lemma for Regular Languages

A technique that is used to show that a given language is not regular

Pumping Lemma for Regular Languages

Let L be a regular language

Then *there exists* some constant $\boldsymbol{N}$ such that *for every* string $w \in L$ s.t. $|w| \geq N$, *there exists* a way to break w into three parts, $w=xyz$, such that:

1. $y \neq \varepsilon$
2. $|xy| \leq N$
3. For all $k \geq 0$, all strings of the form $xy^kz \in L$

This clause should hold for all regular languages.

Definition: N is called the "Pumping Lemma Constant"