A question that I did not put on the exam (b/c it's impossible)

Take a language $L$ and let $\text{repeat}(L)$ be the language consisting of strings which are some string in $L$ repeated (exactly twice)

(If $abaaab$ is in $L$, $abaaababaaab$ is in $\text{repeat}(L)$.)

Tell me how, given a DFA for $L$, I can modify it to create a DFA for $\text{repeat}(L)$.

It turns out this is impossible.
Think about how this string is processed by the DFA.

By the time the DFA has "read" the first $m$ a's, it has gone through $m+1$ states. Since there are only $m$ states, it must have gone through some state twice.

Initial state + 1 state for every letter read.

So, for $w=a_1a_2a_3\ldots a_k$, the path to a final state looks like
Next 30 (?) minutes - a proof this is impossible.

In fact, it's impossible for $L = \Sigma^*$.
So $\text{repeat}(L) = \{ww \mid w \in \{a,b\}^* \}$.

set of strings that are some string repeated twice.

---

for the purpose of a proof by contradiction.

Assume there is a such a DFA $M$.

This DFA has some finite number of states $m$.

Consider the string $w = \underbrace{aa \ldots ab \ldots ab}_{m \quad m \quad m}$

$w$ is in $\text{repeat}(L)$.
first \text{ in } a's

(The there could be more loops, but I'm forced to have a loop in the first \text{ in } a's)

I can repeat (or delete) the loop and still end up at a final state.

So, if we \underline{read} go through the loop twice, we will process

\[
\frac{a \ldots a}{m \text{ times}} \frac{b \ldots q}{m \text{ times}} \ldots \frac{a \ldots b}{\text{ and end at a final state}}
\]
(\(k \geq 1\) is the length of the loop)

This is NOT some string repeated twice. So \(M\) accepts a string that isn't of the form \(ww\).

Conclusion: It is impossible to build a DFA that only accepts repeat(\(\{a, b\}\)) = \(\{ww \mid w \in \{a, b\}\}\).

Another example of a (provably) non-regular language:

\[ L = \{a^n b^n \mid n \geq 0\ \text{some integers}\} \]

\[ \text{with} \quad a^1 b^1 a^1 b^1 \cdots \]

\[ = \{\lambda, ab, aabb, aaabbb, aaaaabbbb, \cdots\} \]
Why is this not possible?

Suppose there was a DFA $M$ for $L$. $M$ has some number $m$ of states. Whatever $m$ is, we can construct the string $a^m b^m \in L$, so $a^m b^m$ is accepted. Somewhere in the first $m$ letters of the path we take for $a^m b^m$, there has to be a loop. Our DFA looks like

If we repeat the loop then we have a path reading $a^{mk} b^m$ (for some $k$, which is the length of the loop).
and ending up at a final state. So this means $M$ accepts $a^{m+k}b^m$, which is not in $L$.

So it's not possible to build $M$ that accepts only the strings in $L$.

These proofs are similar, so we should be able to abstract out the principle behind them to use in other similar problems.

**Pumping Lemma:** Let $L$ be a regular language. Then there exists a positive integer $m$ so that, for every string $w$ with $|w| \geq m$, we can break $w$ up as $w=xyz$, where $|xy| \leq m$, $y \neq \lambda$, so that $xy^iz \in L$ for all $i$. Repeat $y$ $i$ times.
We usually use the pumping lemma to prove a language $L$ is not regular by showing that it does not have the required property.

Revisit the first 2 proofs.

For $L = \text{repeat}(\{a,b\}^*)$,

Assume $L$ is regular. Given $m$, then there exists some $m$, and we can choose $w = a^m b a^m b \in L$. By the pumping lemma $w = xyz$, $|xy| \leq m$, $y \neq \lambda$, with $xy^iz \in L$. But, since $|xy| \leq m$, $xy = a^d$ for some $d$, so $y = a^k$, $k \geq 1$.

This means $xy^2z = a^{m+k} b a^m b \notin L$, a contradiction. So $L$ can't be regular.