Doc 1: I eat apples.

START → N → V → N → END

Doc 2: Green bananas are nasty.

START → ADJ → N → V → ADJ → END

this is your pre-labelled dataset

Doc 3. I eat bananas.

START → N → V → N → END

Calculating Transition Matrix

Three states (POS): Noun, Adjective, Verb

	N	ADJ	V	END
START	2/3	1/3	0	0
N	0	0	$\boxed{3/5}$	2/5
ADJ	1/2	0	0	1/2
V	2/3	1/3	0	0

Look at all the N transitions:

$\boxed{N \to V}$

N → END

$\boxed{N \to V}$

$\boxed{N \to V}$

N → END

Calculating Emission Matrix

	N	ADJ	V
I	$\frac{2}{5}$	0	0
eat	0	0	$\frac{1}{3}$
are	0	0	$\frac{2}{3}$
bananas	$\frac{2}{5}$	0	0
nasty	0	$\frac{1}{2}$	0
green	0	$\frac{1}{2}$	0
apples	$\boxed{\frac{1}{5}}$	0	0

columns sum to 1

Look at all the nouns. How many are the word apple?

I bananas I

$\boxed{\text{apples}}$ bananas

$\Rightarrow \frac{1}{5}$

Emission Matrix(B)

	N	M	V
I	1/2	0	0
walk	1/4	0	1/3
will	0	1	0
fish	1/2	0	2/3

Hidden State Transition Matrix (A)

	N	M	V	END
START	3/4	1/4	0	0
N	$\boxed{1/9}$	1/3	1/9	4/9
M	1/4	0	1/4	0
V	1	0	0	0

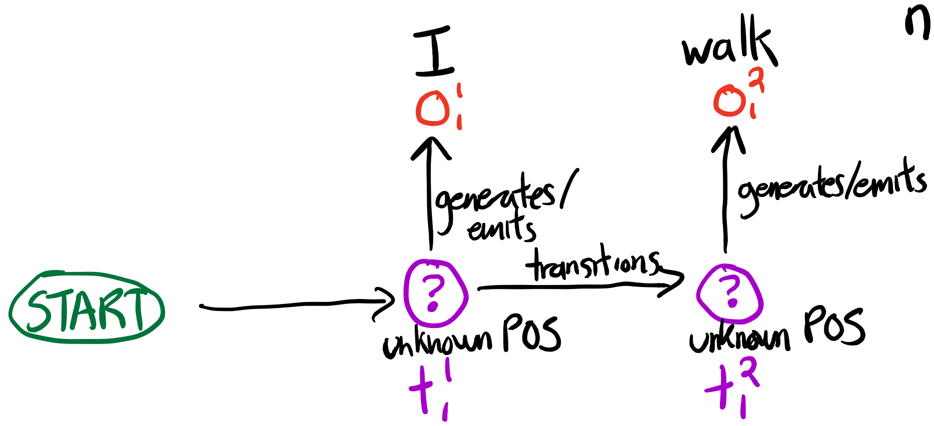
chance that a noun will be followed by another noun

Problem:

Given emission matrix B and transition matrix A, and observed sequence O_1^n (an observed sequence of words from index 1 to n), find t_1^n (the most likely POS tags (hidden states) for index position 1 to n.

Our test document is

$n=2$



1. $$\hat{t}_1^n = \underset{t_1^n}{\operatorname{argmax}} \underbrace{P(t_1^n \mid o_1^n)}_{\text{posterior!}}$$

find the combination of POS tags t_1^n that maximizes the posterior

2. $$\hat{t}_1^n = \underset{t_1^n}{\operatorname{argmax}} \frac{P(o_1^n \mid t_1^n) P(t_1^n)}{P(o_1^n)}$$

convert posterior using Bayes Rule

3. $$\hat{t}_1^n = \underset{t_1^n}{\operatorname{argmax}} P(o_1^n \mid t_1^n) P(t_1^n)$$

we are just trying to maximize, so we don't care about evidence (denominator)

4. $$\hat{t}_1^n = \underset{t_1^n}{\operatorname{argmax}} \prod_{i=1}^{n} \underbrace{P(o_i \mid t_i)}_{\text{probability from emission table}} P(t_1^n)$$

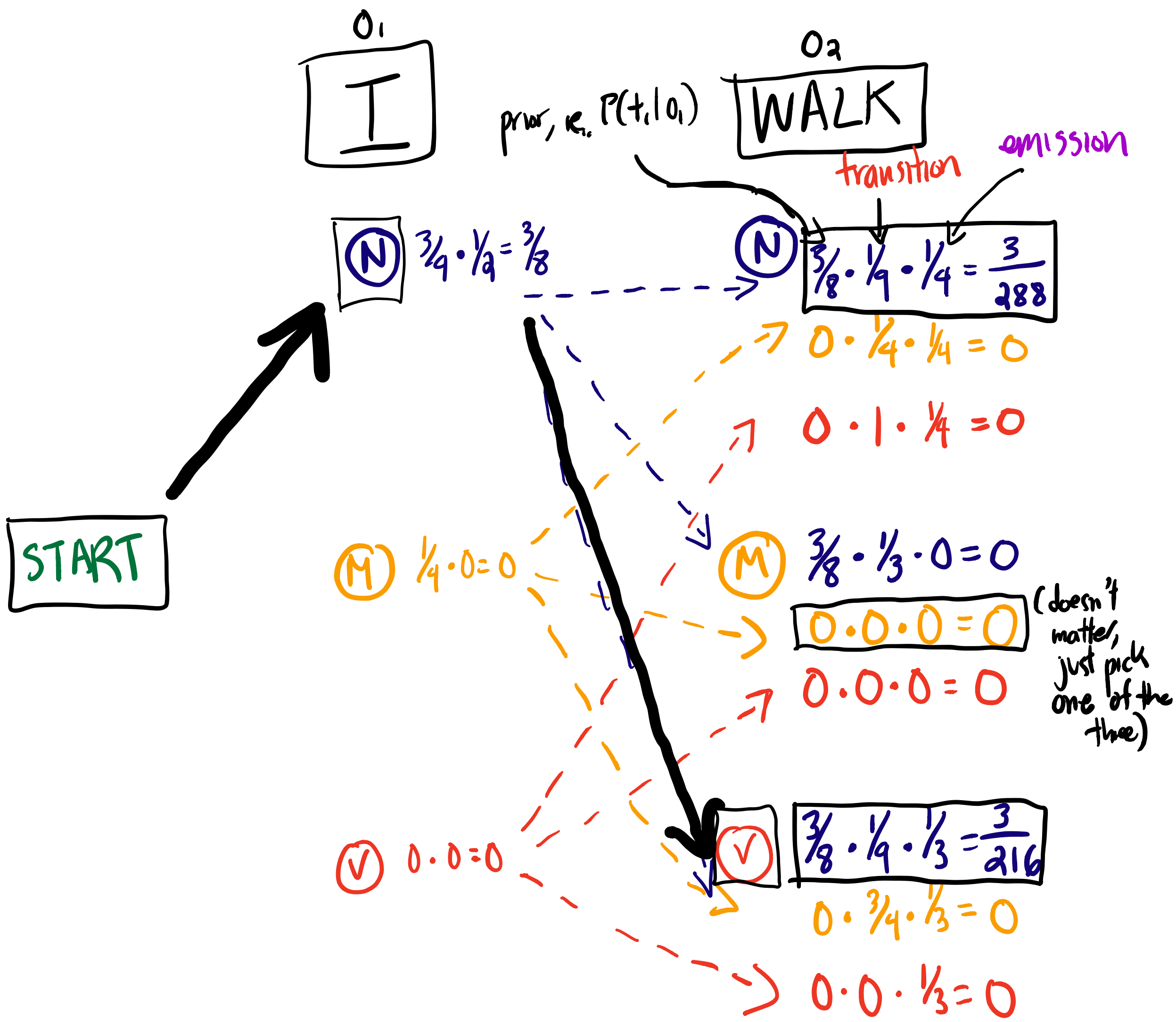
make an assumption that the probability of a word appearing (its emission) is independent of other words and depends only on its hidden state.

5. $$\hat{t}_1^n = \underset{t_1^n}{\operatorname{argmax}} \prod_{i=1}^{n} \underbrace{P(o_i \mid t_i)}_{\text{probability from emission table}} \prod_{i=1}^{n} \underbrace{P(t_i \mid t_{i-1})}_{\text{transition probability}}$$

Make an assumption that the probability of the tag is only dependent on the previous tag (this is the bigram assumption).

6. $$\hat{t}_1^n = \underset{t_1^n}{\operatorname{argmax}} \prod_{i=1}^{n} \text{emission probability} \cdot \text{transition probability}$$

factor out the product notation!



$$\underset{t_1^2}{\operatorname{argmax}} = [N, V]$$

$$\frac{3}{8} > 0 \geq 0 \qquad \frac{3}{216} > \frac{3}{288} > 0$$

Final result: