

## Emission Matrix (B)

	N	M	V
I	$\frac{1}{2}$	0	0
walk	$\frac{1}{4}$	0	$\frac{1}{3}$
will	0	1	0
fish	$\frac{1}{2}$	0	$\frac{2}{3}$

## Hidden State Transition Matrix (A)

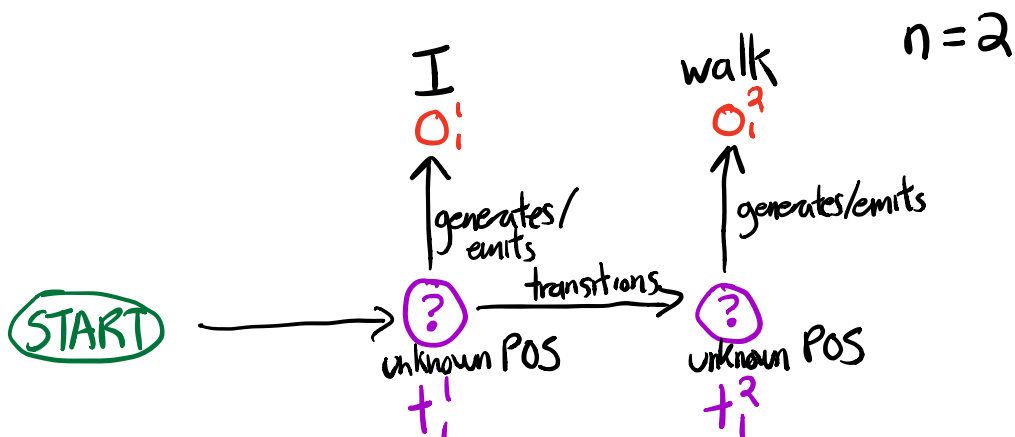
	N	M	V	END
START	$\frac{3}{4}$	$\frac{1}{4}$	0	0
N	$\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{9}$	$\frac{4}{9}$
M	$\frac{1}{4}$	0	$\frac{1}{4}$	0
V	1	0	0	0

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



①

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

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

②

$$\hat{t}_1^n = \operatorname{argmax}_{t_1^n} \frac{P(o_1^n | t_1^n) P(t_1^n)}{P(o_1^n)}$$

convert posterior using Bayes Rule

③

$$\hat{t}_1^n = \operatorname{argmax}_{t_1^n} P(o_1^n | t_1^n) P(t_1^n)$$

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

④

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

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

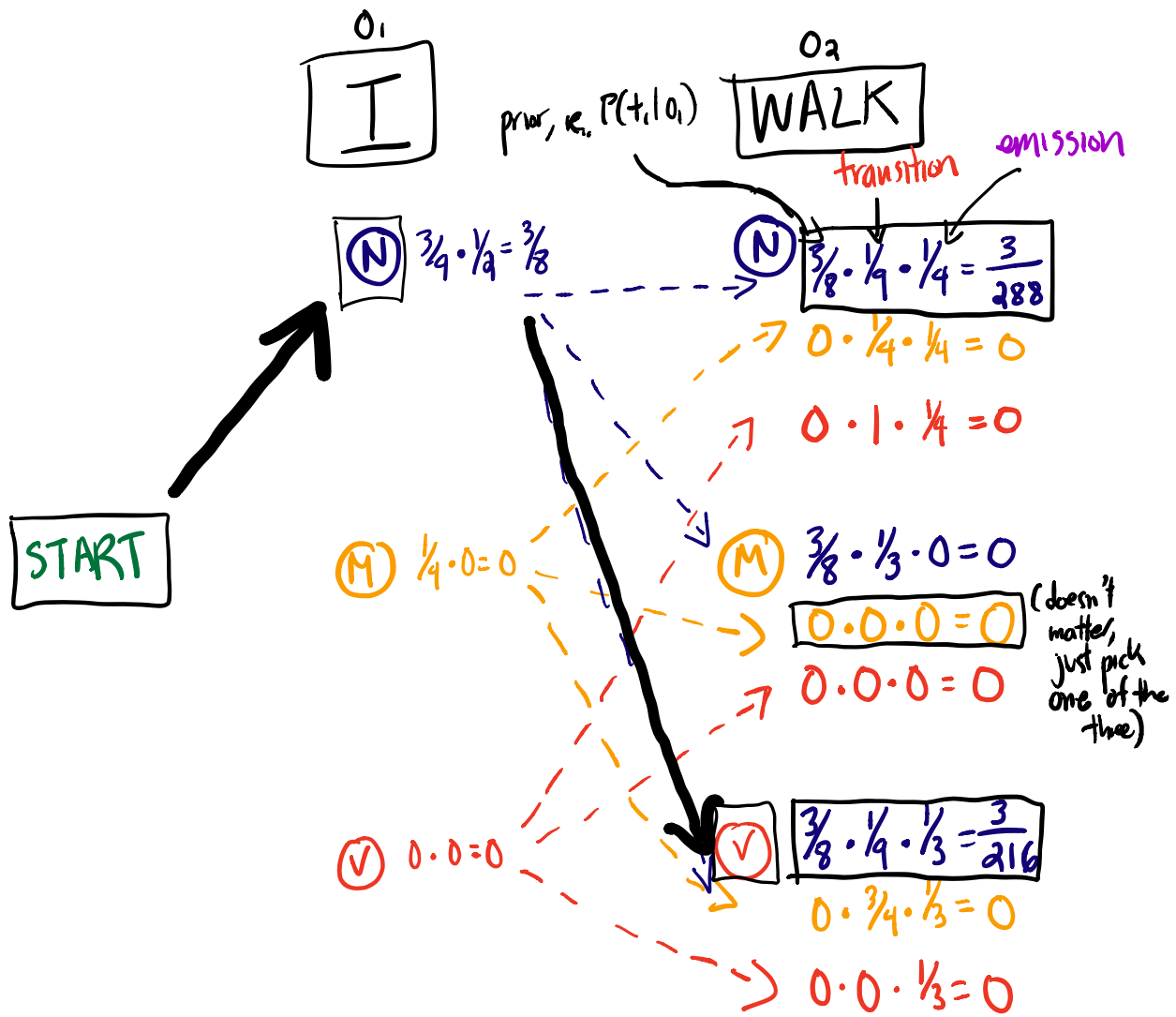
$$\hat{t}_1^n = \operatorname{argmax}_{t_1^n} \prod_{i=1}^n \underbrace{P(o_i | t_i)}_{\text{probability from emission table}} \prod_{i=1}^n \underbrace{P(t_i | t_{i-1})}_{\text{transition probability}}$$

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

⑥

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

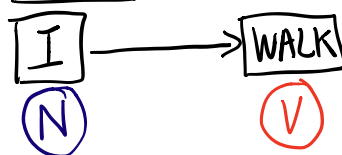
factor out the product notation!



$$\text{argmax}_{t_1} = [N, V]$$

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

Final result:



Doc1: I eat apples.  
 START  $\rightarrow$  N  $\rightarrow$  V  $\rightarrow$  N  $\rightarrow$  END

Doc2: Green bananas are nasty.  
 START  $\rightarrow$  ADJ  $\rightarrow$  N  $\rightarrow$  V  $\rightarrow$  ADJ  $\rightarrow$  END

Doc3: I eat bananas.  
 START  $\rightarrow$  N  $\rightarrow$  V  $\rightarrow$  N  $\rightarrow$  END

this is your  
pre-labelled  
dataset

## Calculating Transition Matrix

Three states (POS): Noun, Adjective, Verb

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

Look at all the N:

$N \rightarrow V$   
 $N \rightarrow END$   
 $N \rightarrow V$   
 $N \rightarrow V$   
 $N \rightarrow END$

# 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	$\frac{1}{5}$	0	0

columns sum to 1

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

