Hidden Markov Model :- [HMM]

Ex:- 1. HMM mainly consider two probabilities
(i) Transition probability
(ii) ~~H~~ Emission probability

Emission probabilities :-

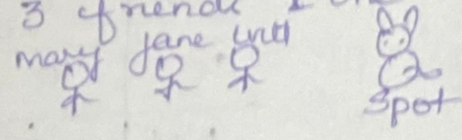
How likely is that a word ~~jane~~ will be a noun, will be a modal, ..

Transition probabilities :-

How likely is that, Noun is followed by a modal which is foll(ow)ed by a verb & than a noun.

Ex :- Jane will spot will → tag the words

3 friends & one cat



Training corpus

1. <s> Mary jane can see will <E>
N N M V N

2. <s> Spot will see mary <E>
N M V N

3. <s> Will jane spot mary? <E>
M N V N

4. <s> Mary will pat spot <E>
N M V N

	N	M	V
Mary	4	0	0
jane	2	0	0
will	1	3	0
spot	2	0	1
can	0	1	0
see	0	0	2
pat	0	0	1

Emission probabilities :-

	N	M	V
marry	4/9	0	0
jane	2/9	0	0
will	1/9	3/4	0
spot	2/9	0	1/4
can	0	1/4	0
see	0	0	1/2
pat	0	0	1/4

Noun appers 9 times in training corpus

Verb appears 4 times & ~~so~~ on.

Modal verb appear 4 Times

consider corpus

<s> M N M N

<s> - start

<E> - End.

Transition probabilities :-

	N	M	V	<E>
<S>	3/4	1/4	0	0
N	1/9	3/9 = 1/3	1/9	4/9
M	1/4	0	3/4	0
V	4/4 = 1	0	0	0

<S> appear = 4

⇒ How many times (prob) Noun appear ~~affa~~ after <S> = 3/4

prob. of M(Modal) appear after <S> = 1/4

" " V & <E> appear after <S> = 0

⇒ Noun = 9

prob. of Noun appear after Noun = 1/9

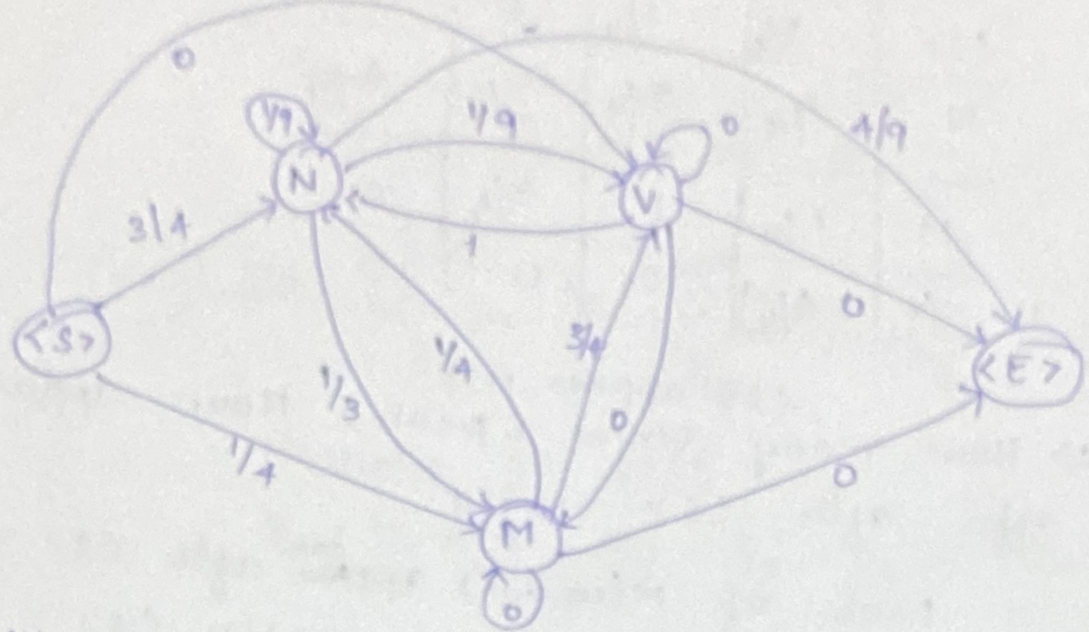
" " Modal " " " = 3/9 = 1/3

" " verb " " " = 1/9

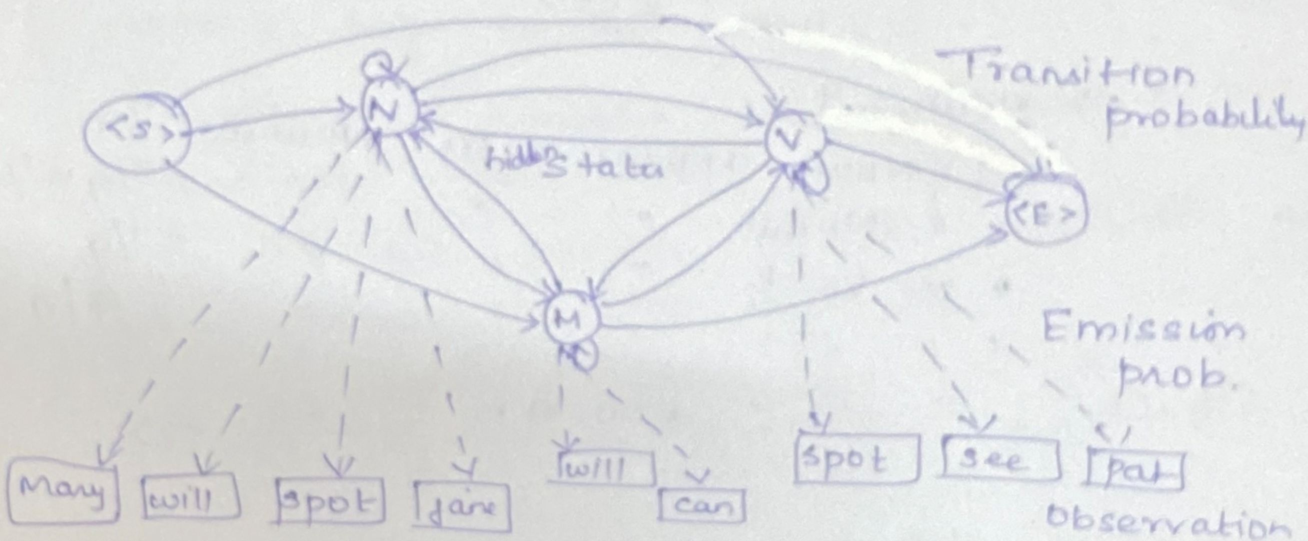
" " <E> " " " = 4/9

& so on

Transition probabilities state dgm



Hidden Markov Model



~~we will go from~~

Between each hidden states we have transition prob. & observation & state we have emission prob.

So to find correct POS to a word we start from <S> & then we go to N

From Noun we have to check emission prob of jane then after from Noun we go to Modal verb & also check modal verb (will) emission prob & so on

Jane will spot Will
N M V N

$$\langle E\rangle \xrightarrow{3/4} \underset{2/9}{\overset{jane}{N}} \xrightarrow{1/3} \underset{3/4}{\overset{will}{M}} \xrightarrow{3/4} \underset{1/4}{\overset{spot}{V}} \xrightarrow{1} \underset{1/9}{\overset{will}{N}} \xrightarrow{4/9} \langle E\rangle$$

Multiply all prob

$$3/4 \times 2/9 \times 1/3 \times 3/4 \times 3/4 \times 1/4 \times 1 \times 1/9 = 0.0003858$$

How can we say that above is the correct tag to the word.

So we consider the prob relative to all the other states ~~the~~

If we consider all the observation as noun. then the prob will be

$$\langle S\rangle \xrightarrow{3/4} \underset{2/9}{\overset{jane}{N}} \xrightarrow{1/9} \underset{1/9}{\overset{will}{N}} \xrightarrow{1/9} \underset{2/9}{\overset{spot}{N}} \xrightarrow{1/9} \overset{will}{N} \xrightarrow{4/9} E$$

$$3/4 \times 2/9 \times 1/9 \times 1/9 \times 1/9 \times 2/9 \times 1/9 \times 4/9 = 0.000000278g$$

This prob is so small ~~when~~ compared to the above ~~as~~ prob.

So basically we can find the prob of each observation with diff POS. & The one with greatest ~~one~~ ~~with~~ ~~is~~ prob chosen.

81 possibilities are there

jane	will	spot	will
N	N	N	N
M	M	M	M
V	V	V	V

- N M V N
- N N N N
- N V N V
- N M M V

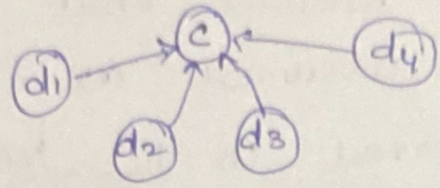
& so on.

It HMM is not so efficient. As we can see that they are 81 possibilities & Need to find prob of all & Need to pick up the largest one. It is very computationally expensive

And also as a data grows & hidden state grows & as observation number grows, the no of prob for comparing will also grows. so to avoid this we can go with viterbi algorithm, etc..

Maximum Entropy Model MEM | Max Ent

Discriminative model :-



Data is there and we assume that hidden state is a generated from the data.

Here flow is upward. (that data identifies class).

For. eg. statements are given (data) and in that data is either +ve or -ve class is hidden.

→ The goal is to determine the POS tag of a observed word.

→ various properties linked with the observed word are used.

→ It includes both previous & future words, as well as their tags.

→ we determine the POS tag for a observed word based on its properties

The Maximum Entropy Model (Max Ent) is a probabilistic model that assigns the most likely tag to a word on a variety of features & ensures that no prior knowledge is assumed ~~as~~ except what is explicitly observed in the training data.

– Entropy states that when predicting t the pos tag for a word, you should make the most uniform, non-biased assumption about it unless the data suggests otherwise.

Feature Representation: Max Ent uses a rich set of features such as:

- The word itself
- The previous word.
- The next word
- Suffixes & prefixes of the word
- Contextual information, such as whether the word is capitalized or not.

Ex :- Training Data: "The dog runs"

- Features: The word "runs" has the feature "ends with -s", which might indicate that it is a verb in the present tense
- other features: 'The' is a determiner (DT) & "dog" " " noun (NN)

Model o/p :-

- "The" ⟶ DT
- "dog" ⟶ NN
- "runs" ⟶ VBZ (verb present tense)