IS 7033: Artificial Intelligence and Machine Learning

Dr. Peyman Najafirad (Paul Rad)
Associate Professor
Cyber Analytics and AI
210.872.7259

https://github.com/paulNrad/ProbabilisticGraphModels

Markov and Hidden Markov Models (HMM)

Outline

- Time series
- Markov Chain
- Hidden Markov Model

Markov Chain

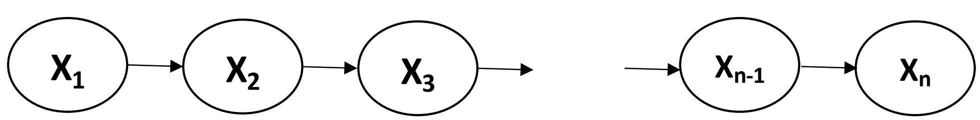
- A Markov chain describes a discrete stochastic process at successive times. The transitions from one state to any of all states including itself, are governed by a **probability distribution**.

$$\mathbf{P(X_t \mid X_1 \ldots X_{t-1}) = P(X_i \mid T_{t\text{-}m} \ldots X_{t-1})}$$

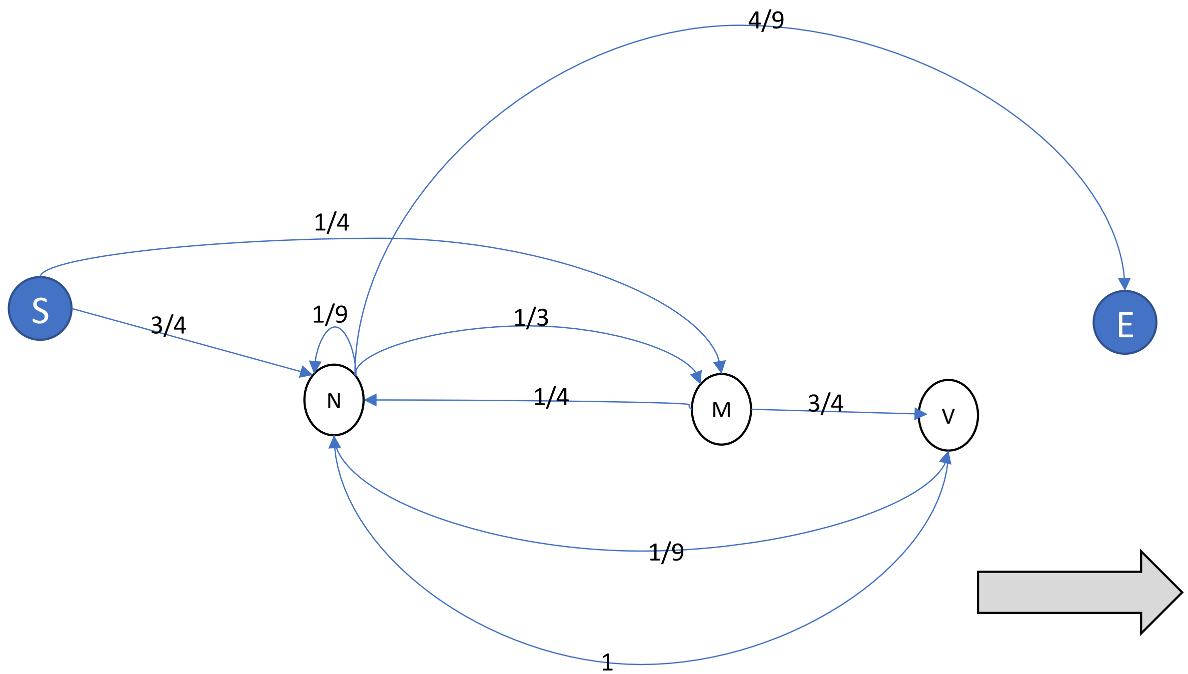
$$\mathbf{X_t = F(X_{t-1}, \ldots X_{t\text{-}m})}$$ **m-order Markov Chain**

- A chain of random variables in which the next one depends (only) on the current one

$$\mathbf{P(X_t \mid X_1 \ldots X_{t-1}) = P(X_i \mid X_{t-1})}$$



Transition Probability

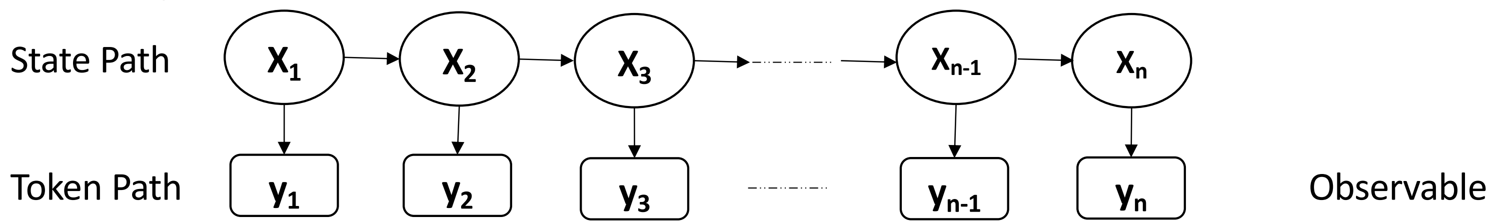


Transition Probabilities

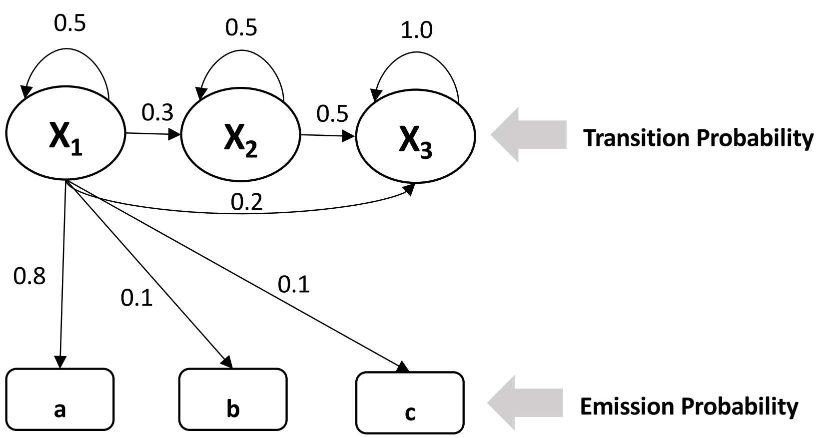
	N	M	V	End
Start	3/4	1/4	0	0
N	1/9	1/3	1/9	4/9
M	1/4	0	3/4	0
V	1	0	0	0

Hidden Markov Model (HMM)

- In addition to State **Transition Probability**, each state of HMM has a probability distribution over the possible output tokens **(Emission Probability)**.
- Thus, a HMM is consist of two strings of information.
  - **The state path** is not directly visible
  - **The token path** (emitted sequence). Infer state path based on the observable token path.



Hidden Markov Model Example



X_1

Out	Trans
a	0.8
b	0.1
c	0.1

X_2

Out	Trans
a	0.2
b	0.6
c	0.2

X_3

Out	Trans
a	0.7
b	0.3
c	0.1

What us probability of HMM producing “a,a,b,c”?

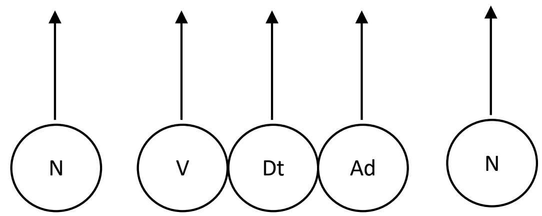
Given a HMM, a sequence of tokens could be generated as following:

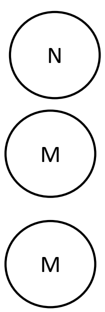
- When we “visit” a state, we emit a token from the state’s emission probability distribution.
- Then we choose which state to visit next, according to the state’s transition probability distribution.

P(a,a,b,c) via 1,1,2,3 = 0.8 x 0.5 x 0.8 x 0.3 x 0.6 x 0.5 x 0.1 = 0.004068
P(a,a,b,c) via 1,2,3,3 = 0.8 x 0.3 x 0.2 x 0.5 x 0.3 x 1 x 0.1 = 0.00072
P(a,a,b,c) via 1,3,3,3 = 0.8 x 0.2 x 0.7 x 1.0 x 0.3 x 1.0 x 0.1 = 0.00336

Example: Part of Speech Tagging

Mary had a little lamb.





Noun

Modal verb

Verb

Jan can run

Look up Table

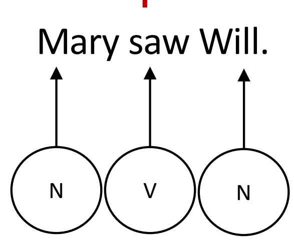
Given Data:

Mary saw Jane.
Jane saw Will.

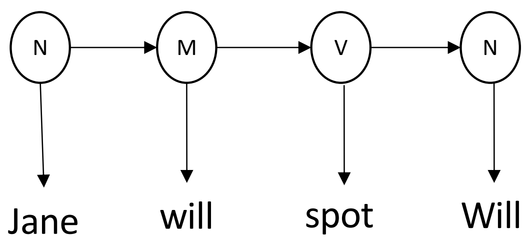
	N	V
Mary	1	0
saw	0	2
Jane	2	0
Will	1	0

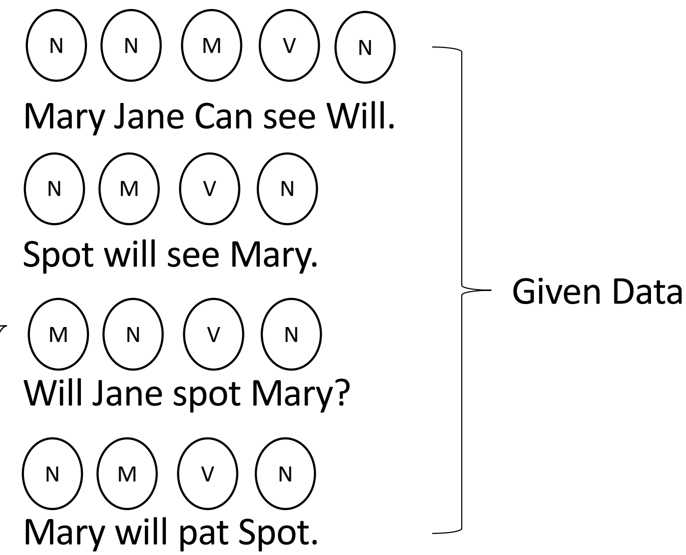
For
Mary saw Will.

Calculate:
Part of Speech Tagging?



Example:
Part of Speech Tagging





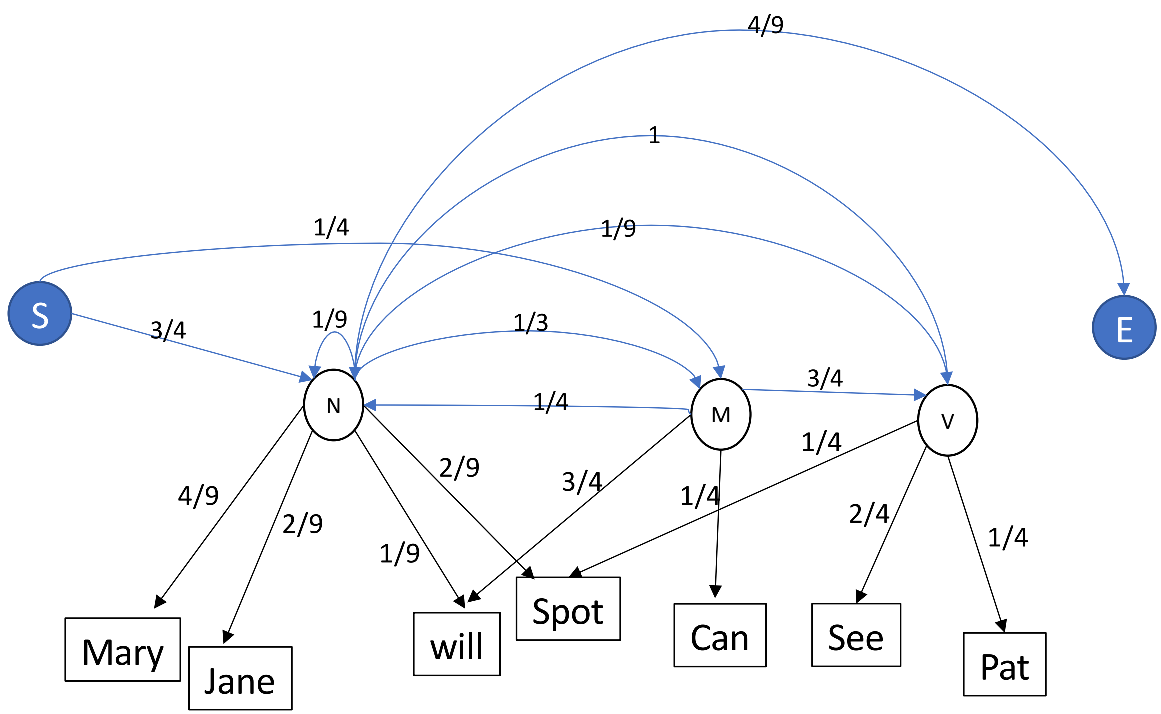
Transition Probabilities

	N	M	V	End
Start	3/4	1/4	0	0
N	1/9	1/3	1/9	4/9
M	1/4	0	3/4	0
V	1	0	0	0

Emission Probabilities

	N	M	V
Mary	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	2/4
Pat	0	0	1/4

Example: Hidden Markov Model (HMM)



Transition Probabilities

	N	M	V	End
Start	3/4	1/4	0	0
N	1/9	1/3	1/9	4/9
M	1/4	0	3/4	0
V	1	0	0	0

Emission Probabilities

	N	M	V
Mary	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	2/4
Pat	0	0	1/4

Assignment

use the **Pomegranate** library or PyMC3 to build a hidden Markov model for part of speech tagging with a universal tagset library:

thttp://www.petrovi.de/data/universal.pdf