Date

Hidden Markov ~~models~~ Process | Chain

* HMM is a stochastic Process, where future state depends only upon current state.

$$P(X_{n+1} = X_n \mid \underbrace{X_{n-1}, X_{n-2}, X_{n-3} \cdots}_{\text{can be ignored.}})$$

$$P(X_{n+1} = X_n)$$

* If it is discrete in nature, it is known as Markov chain

* However, Many real world problems do not hold this assumption.

* In Practice, there is correlation/dependence b/w the States, as we have discussed in Sequence.

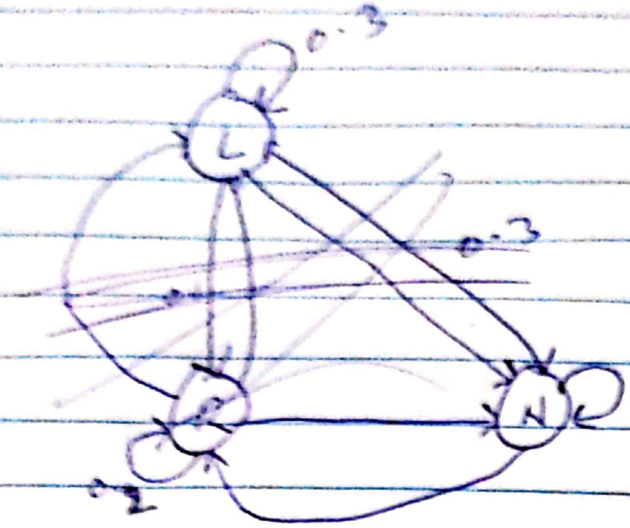
* This Assumption is also known as the Samples/states are i.i.d (Independent and Identically distributed, which may not be true always.

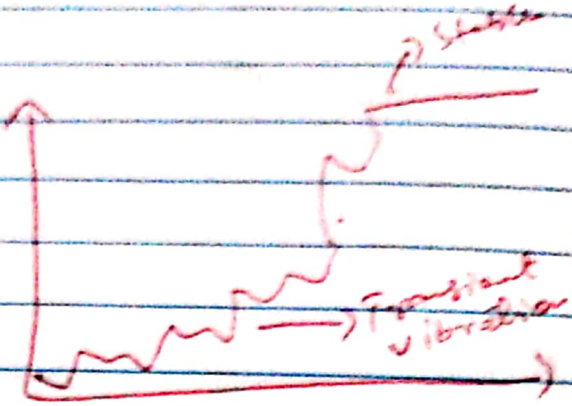
* HMM Can formulate most of the RL Problems.

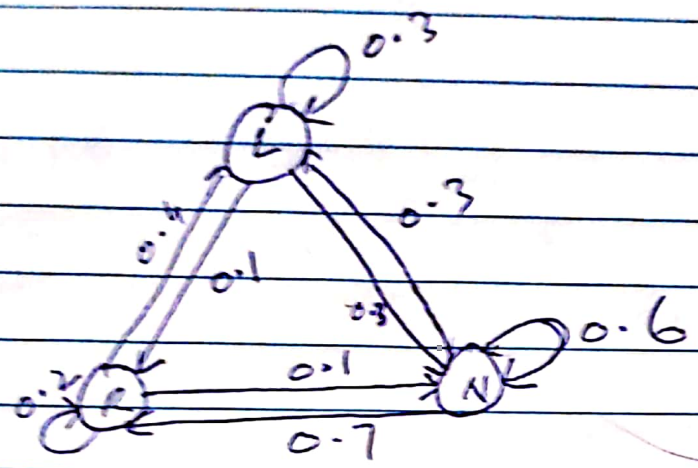
* State Transition Matrix:- Probabilities of transition from one state to another

$$A = \begin{array}{c} \\ L \\ R \\ N \end{array} \begin{array}{c} \begin{array}{ccc} L & R & N \end{array} \\ \begin{bmatrix} 0.3 & 0.1 & 0.3 \\ 0.4 & 0.2 & 0.1 \\ 0.3 & 0.7 & 0.6 \end{bmatrix} \end{array}$$

Date







visualization of the State
Transition matrix

* The columns of State Transition Matrix is one. The Probability of reaching any State from the current state is 1.

* Once it is defined as a matrix, we can use Linear Algebra and Eigen Values to determine the Stable State if it exists, i.e a State for a long time, what is the Probability of being at a Particular State.

* To Solve this, we use Eigen vectors and Eigen values.

$(A-\lambda I)X=0$ (computed form) $\quad . \; AV = \lambda V \qquad \begin{bmatrix} \vdots & \vdots & \vdots \end{bmatrix}\begin{bmatrix} \vdots \end{bmatrix} = \begin{bmatrix} \vdots \end{bmatrix}$

A is the probability transition matrix, V is the Eigenvector and λ is the eigen value.

* Our next State Can be modeled by using the transition matrix and the current State.

$$\underbrace{u_{k+1}}_{\text{next State}} = \underbrace{A}_{\text{Matrix}}\,\underbrace{u_k}_{\text{current State}}$$

→ If this is Eigen vector of A

$$u_{k+1} = \lambda u_k = \lambda A u_{k-1} = \lambda \cdot \lambda \cdot u_{k-1} = \lambda^2 u_{k-1}$$

$$\vdots$$

$$\lambda^k u_1$$

Consider a new vector V in n. we Can represent it using above equation.

$$V_{k+1} = \lambda_1^k u_1 + \underbrace{c_2 \lambda_2^k u_2 + \dots c_n \lambda_n^k u_n}_{\to \text{Approaches to zero}}$$

$\lambda_1^k = 1$ (normalized eigenvector)

$$V_{k+1} = u_1 \text{ (Stable State)}.$$

Date

* A Markov Matrix always has an eigenvalue 1.
* All other Eigenvalues will have a magnitude smaller or equal to 1. Let's say the eigen vector $u [0.3, 0.4, 0.3]$ has an Eigen value 1 which is a stable state, means the probabilities of left, right and no movement are 0.3, 0.4, and 0.3 respectively, and these probabilities will not change for a long time.

$Ax = \lambda x$

cA

———— X ———— X ————

$A^{-1}x = \lambda^{-1}x$

Random process / Random walk:-

* An exact solution will be computationally intensive
* The problem can be solved by random walk.
* 100 Persons in a conference, we can determine clusters, Highly communicating persons (Social networks). This will become a sparse matrix, this reduces the computational complexity.

* we don't know any information about the data like ages, gender, education, still we can cluster them; which in my opinion is the best solution.

* These approaches are useful in recommendation systems.

HMM (Hidden markov Model)

* MLE (maximum likelihood estimator) assumes samples are i.i.d, but for time-series data, this may be not be true

For example: If I am happy now, there is a 90% chance that, I will also be happy tomorrow

* States are not completly observable.

* We estimate the state by observations, for example, If I am happy, I am likely to be found on a party or at cineplex. (Movie)

* The observation matrix B is the emission probability, the complexity arises when some observations happen for different states.

~~Example~~

$P(X_i = \text{happy}) = 0.8$

$P(X_i = \text{Sad}) = 0.2$

State Transition matrix.

$$A = \begin{array}{c} \\ H \\ S \end{array} \begin{array}{c} \begin{array}{cc} H & S \end{array} \\ \begin{bmatrix} 0.99 & 0.01 \\ 0.1 & 0.9 \end{bmatrix} \end{array}$$

Initial probability distribution

Movie Book party dine

$$B = P(y_i / X_i) = \begin{array}{c} H \\ S \end{array} \begin{bmatrix} 0.2 & 0.2 & 0.4 & 0.2 \\ 0.4 & 0.3 & 0.1 & 0.2 \end{bmatrix}$$

$$P(x_t | y_{1:t}) = \frac{P(y_t | x_t)\, P(x_t | y_{1:t-1})}{\sum_{x_t} P(y_t | x_t)\, P(x_t | y_{1:t-1})}$$