Hidden Markov Model

→ Markov process is a simple ~~sto~~ stochastic process in which the distribution of future state depends only on the present state and not on how it arrived in the present state.

→ A random ~~process~~ sequence has the Markov property if its distribution is determined solely by its current state. Any random process having this property is called a Markov random process.

- For observable state sequences (state is known from data), this leads to a Markov chain Model.

- For non-observable states, this leads to a Hidden Markov Model (HMM).

Discrete Markov Model →

A discrete (finite) system :
- has N distinct states.
- Begins (at time $t=1$) in some initial state(s).
- At each time step ($t=1, 2, \dots$) the system moves from current to next state (possibly

the same as the current state) according to TRANSITION PROBABILITIES associated with current state.

- This kind of a system is called a finite or discrete Markov Model.

- Named after Andrei Andreyevich Markov (1856–1922)

Markov Property: The state of the system at time t+1 depends only on the state of the system at time t.

Stationary Assumption → In general, a process is called stationary if transition probabilities are independent of t, namely.

for all t, $P[X_{t+1} = x_j \mid X_t = x_i] = P_{ij}$

This means that if system is state i, the prob. that system will next move to state j is P_{ij}, no matter what the value of t is.

So,

Markov model → A stochastic processes holding the markov property.

Markov property → Memoryless Property

Three basic information to define a Markov model :
- Parameter space
- State space
- State transition prob.

- Markov model is associated with set of states :
  $\{S_1, S_2, \ldots S_N\}$

- Process moves from one state to another state generating a sequences of states :

  $S_{i1}, S_{i2}, S_{i3}, \ldots S_{ik}, \ldots$

& we have markov chain property : prob. of each subsequent state depends only on what was the previous state :

$$P(S_{ik} \mid S_{i1}, S_{i2}, \ldots S_{ik-1}) = P(S_{ik} \mid S_{ik-1})$$

- To define a markov model, following prob. have to be specified →

Transition Probabilities:

$$a_{ij} = P(S_i | S_j)$$ {given S_j, I can find prob. of S_i}

Initial Probabilities:

$$\pi_i = P(S_i)$$ {Prob. of start state}

- The output of process is set of states at each instant of time.

Markov Model Example:

- classify a weather into three states:
  - State 1 : Rainy
  - State 2 : cloudy
  - State 3 : Sunny

- The weather of some city found following weather change pattern -

Tomorrow

Today	Rainy	cloudy	Sunny
Rainy	0.4	0.3	0.3
cloudy	0.2	0.6	0.2
Sunny	0.1	0.1	0.8

Read it like →

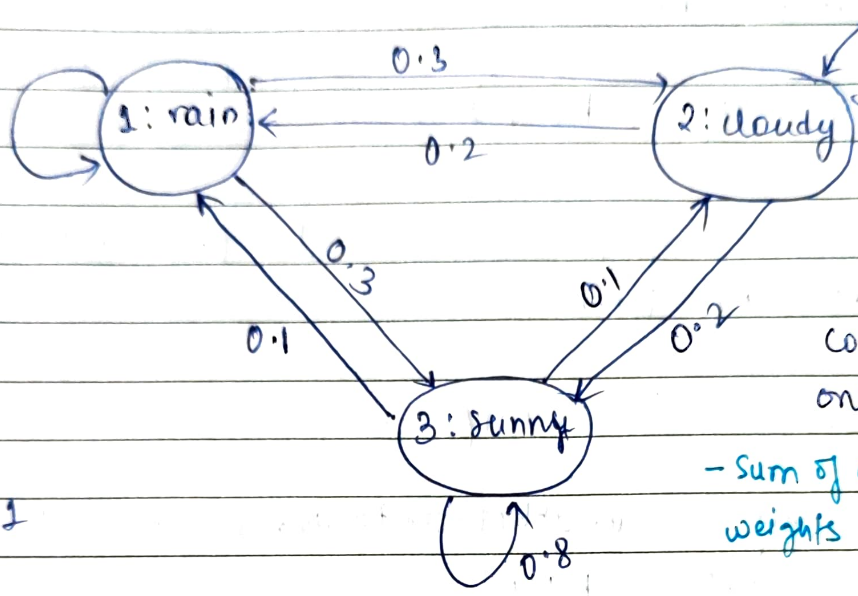
→ So, if today is Rainy → chances of being Rainy tommorow is = 0.4

→ cloudy → = 0.3

→ sunny → = 0.3

Tomorrow's weather depends only on Today's weather.

Representation using graph:



Each state corresponds to one observation.

- Sum of outgoing edge weights is 1.

let Initial Prob. = 1 of all states π_i

Que → what is the prob. that weather for the next 7 days will be "sun - sun - rain - rain - sun - cloudy - sun", when today is sunny?

S_1 = Rain $\quad S_2$ = Cloudy $\quad S_3$ = Sunny

sequence of observation

Today

$P(O \mid model) = P(S_3, S_3, S_3, S_1, S_1, S_3, S_2, S_3 \mid model)$

$= P(S_3) \cdot P(S_3 \mid S_3) \cdot P(S_3 \mid S_3) \cdot P(S_1 \mid S_3) \cdot P(S_1 \mid S_1) \cdot P(S_3 \mid S_1) \cdot P(S_2 \mid S_3) \cdot P(S_3 \mid S_2)$

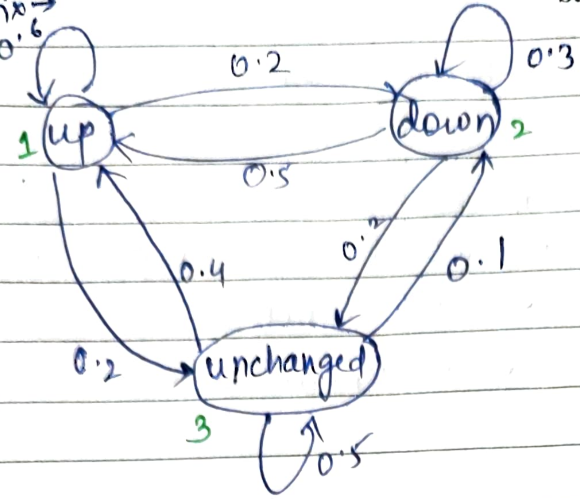
$= \pi_3 \cdot a_{33} \cdot a_{33} \cdot a_{31} \cdot a_{11} \cdot a_{13} \cdot a_{32} \cdot a_{23}$

$= 1 \cdot (0.8) \cdot (0.8) \cdot (0.1) \cdot (0.4) \cdot (0.3) \cdot (0.1) \cdot (0.2)$

$= 1.536 \times 10^{-4}$

Markov model Matrix →

Eg:



Initial state probabity matrix

$$\pi = \begin{bmatrix} 0.5 \\ 0.2 \\ 0.3 \end{bmatrix}$$

State transition prob. matrix =

$$A = \{a_{ij}\} = \begin{bmatrix} 0.6 & 0.2 & 0.2 \\ 0.5 & 0.3 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix}$$

→ what is the prob. of 5 consecutive up days?

sequence ⇒ up - up - up - up - up

$P(O|\text{model}) = P(S_1, S_1, S_1, S_1, S_1)$

{Here, we are not given with "what is today", so first "up" will be considered as initial prob.)

$= P(S_1) \cdot P(S_1|S_1) \cdot P(S_1|S_1) \cdot P(S_1|S_1) \cdot P(S_1, S_1)$

$= \pi_1 \cdot a_{11} \cdot a_{11} \cdot a_{11} \cdot a_{11}$

$= 0.5 * (0.6)^4$

$= 0.0648$

Hidden Markov Model →

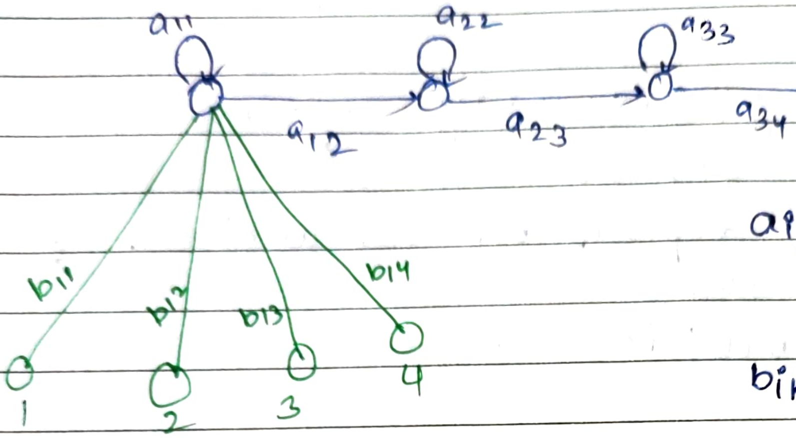
- A hidden markov model is an extension of a markov model in which the input symbols are not the same as the states
- This means we don't know which states we are in.

Eg: In HMM POS-tagging.
→ Input symbols → words
→ states : part of speech tags.

Input symbol	words	The	man	went	to	the	park
	tags	Det	Noun	Verb	prep	Det	Noun

- Hidden Markov Models are probabilistic finite state automata.

- often we face scenarios where states cannot be directly observed. so we need an extension which is HMM.



a_{ij} = State transition prob.

b_{ik} = observation (output) prob. or emmission prob.

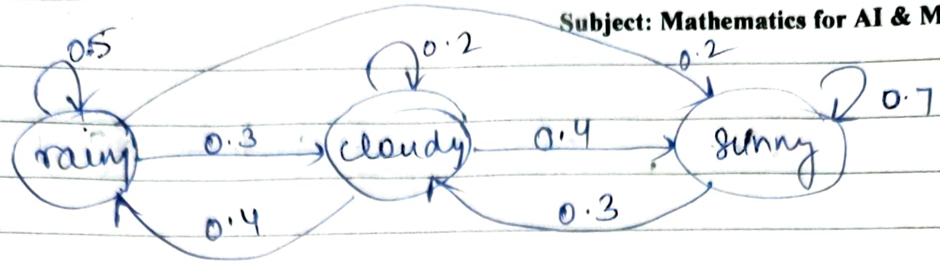
Here –

$b_{11} + b_{12} + b_{13} + b_{14} = 1$

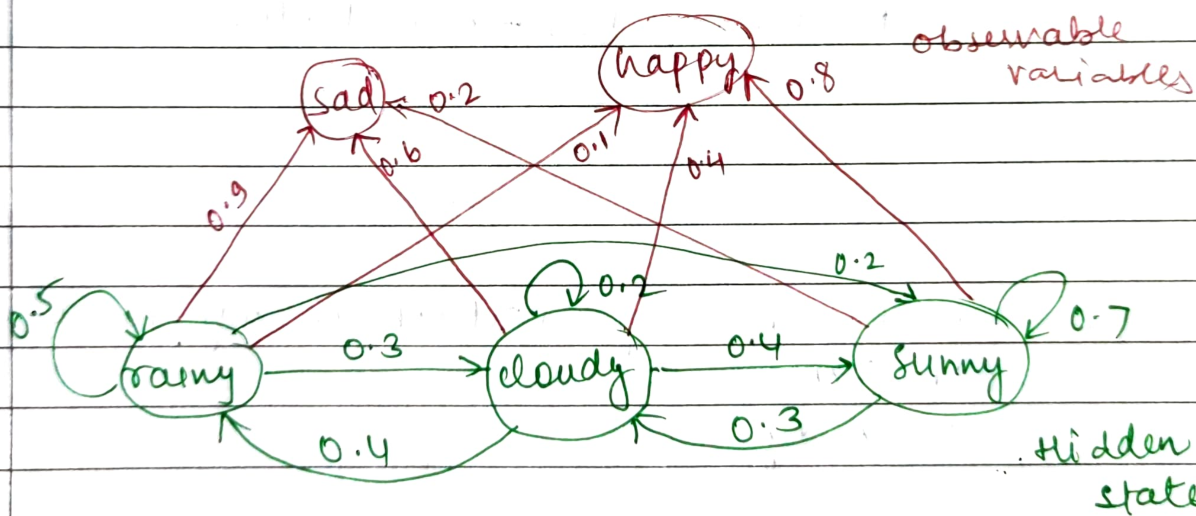
$b_{21} + b_{22} + b_{23} + b_{24} = 1$ etc.

<u>Example</u>:

- Assume the kind of weather at my place are:
  1. Rainy
  2. Cloudy
  3. Sunny
- On any given day only one of them will occur.
- The weather tomorrow depends only upon the weather today.
- We can model this with a simple markov chain.
- Adding the state transition & tranision probabilities
  →



Now, at any given day my mood can be
— Happy or Sad.
This mood depends on the weather of that particular day.
The probabilities are —



The states of the markov chain is hidden or unknown.
We can observe some variables that are dependent on these hidden states. Such a model is called as Hidden Markov Model.

HMM = Hidden Markov chain + Observed variables

Transition Prob.

$$\begin{array}{c|ccc} & \text{Rainy} & \text{Cloudy} & \text{Sunny} \\ \hline \text{Rainy} & 0.5 & 0.3 & 0.2 \\ \text{Cloudy} & 0.4 & 0.2 & 0.4 \\ \text{Sunny} & 0.0 & 0.3 & 0.7 \end{array}$$

~~Rainy cloudy sunny~~

Observable variables prob →

$$\begin{array}{c|cc} & \text{Sad} & \text{Happy} \\ \hline \text{Rainy} & 0.9 & 0.1 \\ \text{Cloudy} & 0.6 & 0.4 \\ \text{Sunny} & 0.2 & 0.8 \end{array}$$

You are given with 3 consecutive days –

1. sunny + Happy
2. cloudy + Happy
3. sunny + sad

We cant observe the hidden state (lets assume this scenario), so what is the prob. of this scenario?

→ It is same as joint probability:

P(happy – happy – sad, sunny – cloudy – sunny) = ?

By using the Markov Property →

$P(Y = \text{happy-happy-sad};\ X = \text{sunny-cloudy-sunny})$

$= P(X_1 = \text{sunny}) * P(Y_1 = \text{happy} \mid X_1 = \text{sunny}) * P(X_2 = \text{cloudy} \mid X_1 = \text{sunny}) * P(Y_2 = \text{happy} \mid X_2 = \text{cloudy}) * P(X_3 = \text{sunny} \mid X_2 = \text{cloudy}) * P(Y_3 = \text{sad} \mid X_3 = \text{sunny})$

$= 0.509 * 0.8 * 0.3 * 0.4 * 0.4 * 0.2$

$= 0.00391$

Initial prob. (given)

This is how, we calculate prob. of (Y, X).

If we hide these weather states from us, we only have a sequence of observed variables.

What is the most likely weather sequence for this observed sequence?

- There are many possible permutations (eg - rainy-cloud-sunny, rain-rain-sunny, sunny-sunny-rain etc)

→ Find the weather sequence which maximizes the joint probability.

Max. Joint Prob: P(happy-happy-sad, sunny-sunny-cloudy)

$= 0.04105$