Markov chains

It is a stochastic Model that depicts a sequence of possible events where predictions or probabilities of next state are based only on the previous state

stochastic means having a random probability distribution that may be analyzed stastically but may not be predicted precisely.

The Markov Property is followed which means that probability of state n depends only on $n-1$ state but not $n-2$, $n-3$ ... or any other previous states. (Memoryless property)

They are similar to Finite state Machines but have probabilities associated with them.

Markov chains are used in weather prediction genetics, finance, NLP, search engines and queue analysis

It is a very basic model. used in data science., a precursor to HMM

A state S_t is Markov iff

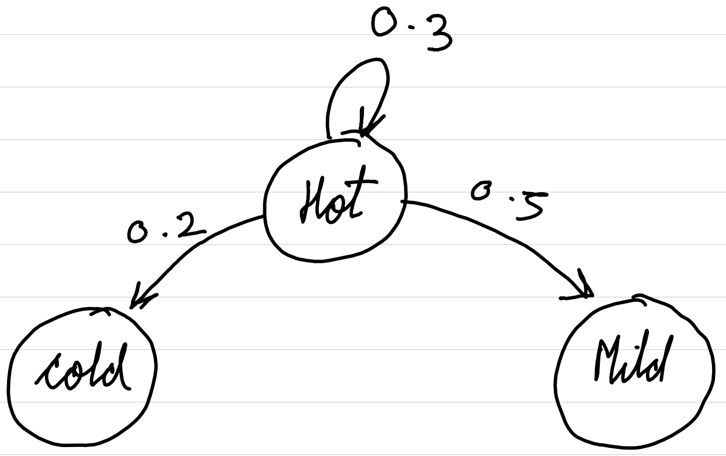
$$P\left[S_{t+1} \mid S_t\right] = P\left[S_{t+1} \mid S_1, \ldots S_t\right]$$

" The future is independant of the past given the present "

A Markov model assumes that →

Once current state is known the history is thrown away, because the current state contains all previous information example we have climate of days.

Today is a hot day. what will it be tomorrow?.



If today is hot,
The probability that it will be mild is 0.5
hot again is 0.3 and cold is 0.2

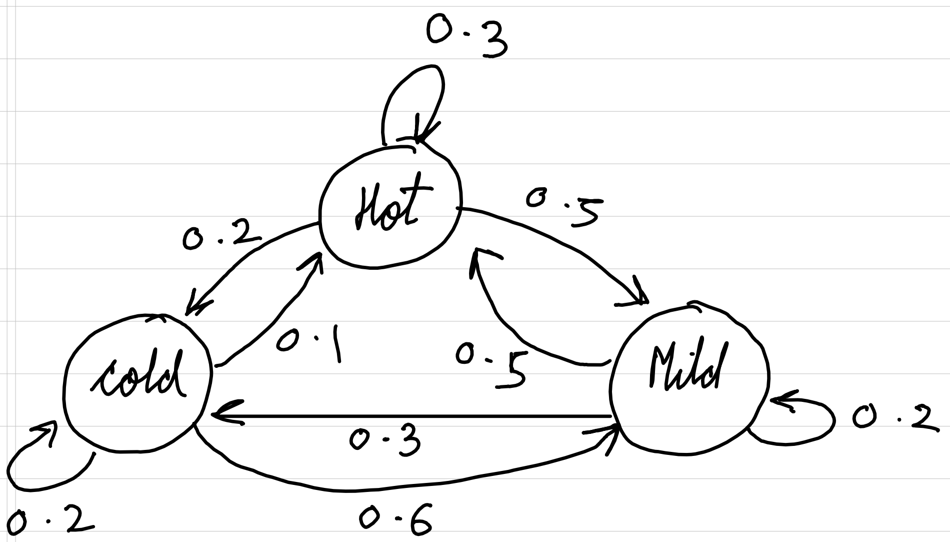
We have to assume that tomorrows temperature is dependant only on todays temperature and not yesterdays temperature

0.3
0.2
0.5 } Transition Probabilities

Hot, Cold, Mild → states.

eg
$$P(x_{n+1} = \text{Cold} \mid x_n = \text{Hot}) = 0.2$$

Similarly for every state we have transitions



The sum of all transition probabilities from one state must be 1
ie. 0.3 + 0.2 + 0.5 = 1 all outgoing arrows must sum to 1

The incomming arrows need not be 1

Calculation of Probability

what is the probability that day after tomorrow is Hot if today is Mild?

$$P(X_{n+2} = \text{Hot} \mid X_n = \text{Mild})$$

$$P(X_{n+2} = \text{Hot} \mid X_{n+1} = \text{Hot}, X_n = \text{Mild}) \, P(X_{n+1} = \text{Hot} \mid X_n = \text{Mild})$$
$$+ \, P(X_{n+2} = \text{Hot} \mid X_{n+1} = \text{Mild}, X_n = \text{Mild}) \, P(X_{n+1} = \text{Mild} \mid X_n = \text{Mild})$$
$$+ \, P(X_{n+2} = \text{Hot} \mid X_{n+1} = \text{Cold}, X_n = \text{Mild}) \, P(X_{n+1} = \text{Cold} \mid X_n = \text{Mild})$$

By local Markov Property

$$P(X_{n+2} = ☺ \mid X_{n+1} = ☺, X_n = \text{Mild}) =$$

$$P(X_{n+2} = ☺ \mid X_{n+1} = ☺)$$ as todays' Mild climate doesnt affect day after tom.

$$= 0.2 \times 0.5 + 0.2 \times 0.5 + 0.3 \times 0.1$$

$$= 0.23$$

Adjacency Matrix

The transition is represented as an adjacency Matrix for solving

	Hot	Mild	Cold
Hot	0.2	0.3	0.5
Mild	0.5	0.2	0.3
Cold	0.1	0.6	0.3

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

This is used to translate the probability calculations into efficient matrix multiplications

Day 1	Day 2		
Hot	Hot	$\rightarrow$	0.2
Hot	Mild	$\rightarrow$	0.3
Cold	Cold	$\rightarrow$	0.3

This is also called as state transition matrix

$$P_{ss'} = P\left[\underset{\substack{\uparrow \\ \text{Next state}}}{S_{t+1}} = \underset{\substack{\uparrow \\ \text{Possible state}}}{s'} \;\middle|\; \underset{\substack{\uparrow \\ \text{current state}}}{S_t = s} \right]$$

Random Walk

Random walk means moving along the transition of states in a random manner respecting the probabilities of transition

eg a random walk might be

Hot - Mild - Hot - Mild - Cold - Mild - Hot

This can be used for generation of a sequence in line with the transition probabilities

Helpful to model particle diffusion, stock prices

Used to compose music, generate text

Calculation of Probabilities using Matrix

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

Consider prior probability π

Today is Mild $\therefore \pi = [0 \quad 1 \quad 0]$

When we Multiply π by A, we get the probability that what tomorrow will be

$$\pi A = [0.5 \quad 0.2 \quad 0.3] = \pi_1$$

$\therefore$ Tomorrow it will be Hot with probability 0.5, Mild with 0.2, cold with 0.3

what is the probability that Day after tomorrow is Hot given today is Mild?

Multiply the New Vector π_1 by A

$$\pi_1 A = [0.23 \quad 0.37 \quad 0.4] = \pi_2$$

Probability that it is hot is 0.23
cold is 0.4

The elements of π must add up to 1

We can use π_2 to calculate the probabilities for further days

In the end, π will converge to a value and remain the same which is the stationary state.

Stationary Distribution

If we perform random walk for a long time, the probabilities of a day climate start converging

eg $P(\text{Day} = \text{Hot})$ or $P(\text{Day} = \text{cold})$ can be calculated over ∞ random walks wont change.

This is because the MC reaches a state where the probability of next state occuring will be same for all future states

These are called stationary distributions or equilibrium state or steady state

The stationary distribution doesnt change with time. Over time the influence of the initial state diminishes and probabilities become spread out over time, balancing the transition between states.

This means that regardless of the initial state the system eventually settles into a stable distribution.

These can be calculated by linear algebra without having to perform ∞ random walks

We want to find a value that converges over time. that is when we calculate the Probabilities that day will be cold $\rightarrow$

$$\pi_k A = \pi_{k+1}$$

if there exists a stationary state then

$$\pi_{k+1} = \pi_k$$

$$\therefore \pi A = \pi$$

Solving using eigen vectors

$A v = \lambda v$ — ① $v \rightarrow \pi$ $\lambda \rightarrow 1$	elements of π must add up to 1 $\sum \pi_i = 1$ — ②

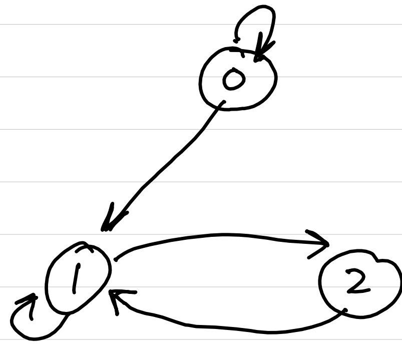
Solving ① & ②, π can be obtained

There can be more than one values of π, that is more than one stationary state

Markov chains are used in ranking of websites. A link represents transition between webpages.

Markov chain is used to determine the equilibrium condition whose probabilities represent ranking of webpages.

Reducible Markov chains



As soon as the random walk leaves state 0, there is no chance of returning to 0

Hence the Probability of revisiting state 0 on a random walk starting with 0 is less than 1

0 is the transient state. We dont know if we are comming back

1 & 2 are called recurrent states. The probability of random walk revisiting it is 1

This type of Markov chains are called reducible.

If every state is rechable from any state then it becomes irreducible Markov chain

We have calculated π for a single starting state

Similarly, we can calculate all possibilities

$$A \times A = \begin{matrix} 0.24 & 0.42 & 0.34 \\ 0.23 & 0.37 & 0.4 \\ 0.35 & 0.33 & 0.32 \end{matrix}$$

Sum of all rows must be 1

A^2	Hot	Mild	Cold
Hot	0.24	0.42	0.34
Mild	0.23	0.37	0.4
Cold	0.35	0.33	0.32

Day 0	Day 2	
$\therefore$ Hot,	Hot $\rightarrow$	0.24
Mild,	Hot $\rightarrow$	0.23
Cold,	Cold $\rightarrow$	0.32

for 3rd day, Prior probabilities become A^2 while transition probabilities remain A.

This way for 3rd day all the combinations can be calculated using A^3

In other words, A^2 can be expressed as the probability of reaching a state from a given state after 2 steps

ie. Probability of reaching Hot from starting with Hot day in 2 days is 0.24

This can be done by any of the paths

Hot - Cold - Hot

Hot - Hot - Hot

Hot - Mild - Hot

The total probability is the sum of all the 3 paths ie.

$$P = P(\text{Hot} - \text{cold}) \times P(\text{cold} - \text{Hot}) + P(\text{Hot} - \text{Hot}) \times P(\text{Hot} - \text{Hot}) + P(\text{Hot} - \text{Mild}) \times P(\text{Mild} - \text{Hot})$$

also represented as

$$P_{00}(2) = P_{00}(1) * P_{02}(1) + P_{01}(1) + P_{12}(1) + P_{02}(1) + P_{22}(1) \quad - (1)$$

for A^n we thus have the probability of reaching state i to j in exactly n steps.

$$P_{ij}(n) = A^n_{ij}$$

Thus by repeated Matrix Multiplication we can find the n step transition probability.

So we can generalize (1) as

$$P_{ij}(n) = \sum_k P_{ik}(r) \times P_{kj}(n-r)$$

This is called as the Chapman-Kolmogorov Theorem

k = intermediate state given in r steps

$$\lim_{n \to \infty} A^n = \text{stable state}$$

Advantages → Simplicity

Efficiency

Interpretability

Scalability to large datasets

Disadvantages → Memoryless

Cant handle complex patterns

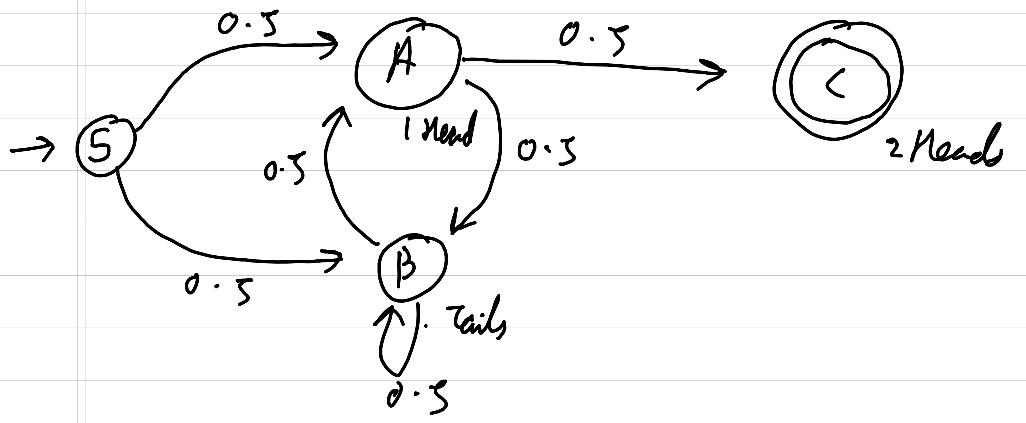
Needs large dataset to be trained accurately

Homogenity Assumption → the transition table is same over time

Markov chains can be trained through count based estimation (counting no of times the transition occurs & updating values) or through prior knowledge

① A coin is tossed repeatedly. What are the expected number of throws before two consecutive heads?

→



A : Current toss was head & last one was not head
B : Current toss was tails
C : Current toss was head & last one was head
S : starting state

Random walk → BABBABBAA, AA, BAA, ..
BABAA, ABBAA, etc

Expected no of throws from state A

$$E_A = \underset{\substack{\uparrow \\ \text{Current} \\ \text{toss}}}{1} + \underset{\substack{\uparrow \\ \text{Going to } E_B}}{0.5 \times E_B} + \underset{\substack{\uparrow \\ \text{End} \\ \text{of} \\ \text{game}}}{0.5 \times E_C}$$

Expected no of throws from state B

$$E_B = \underset{\substack{\uparrow \\ \text{Current} \\ \text{toss}}}{1} + \underset{\substack{\uparrow \\ \text{Going to} \\ E_A}}{0.5 \times E_A} + \underset{\substack{\uparrow \\ \text{Comming} \\ \text{back to } E_B}}{0.5 \times E_B}$$

$E_C = 0$ (end)

∴ We get

$$E_A = 1 + 0.5 E_B \quad ①$$
$$E_B = 1 + 0.5 E_A + 0.5 E_B \quad ②$$

from ① & ② Subs E_A

$$E_B = 1 + 0.5(1 + 0.5 E_B) + 0.5 E_B$$

$$\therefore E_B = 1 + 0.5 + 0.75 E_B$$
$$\therefore E_B = \frac{1.5}{0.25} = 6$$

$$\therefore E_A = 4$$

from starting state S

$$E_S = 1 + 0.5 E_A + 0.5 E_B$$
$$= 1 + 0.5 \times 4 + 0.5 \times 6$$
$$= 6$$

Answer → 6