Machine Learning & Deep Learning

Week-11

Instructor: *Engr. Najam Aziz*

Hidden Markov Model (HMM)

Stochastic/Random Processes

- Weather - Sunny, Rainy, Cloudy
- Customer next phone - Nokia, Samsung, iPhone
- Infant baby - Playing, Eating, Sleeping, Crying
- Tourist city visit - A, B, C
- Game Result - Lose, Draw, Win

Markov Model

- A Markov Model is a stochastic model used to model randomly changing systems/events/states of finite number..
- It is assumed that future state depends on the current state, not on the one occurred before it. Event/state at time t+1 depends on t, not on t-1 - ***Markov Property***
- ***Markov Chain/Process:*** Any change of events/ chain that is following the markov property is called markov chain/process.
- Let $\{x_0, x_1, x_3, ..., x_n\}$ be a sequence of distinct random variables/ events. Then $\{x_0, x_1, x_3, ..., x_n\}$ is a Markov chain/process if it satisfy the markov property.

While doing Exercise, since you pick the next exercise set based on the set you've done before, your workout routine follows the Markov assumption. It assumes the transition probability between each state only depends on the state you are in.

A Markov chain has short-term memory, it only remembers where you are now and where you want to go next.

https://towardsdatascience.com/markov-models-and-markov-chains-explained-in-real-life-probabilistic-workout-routine-65e47b5c9a73

Markov Chain/Process

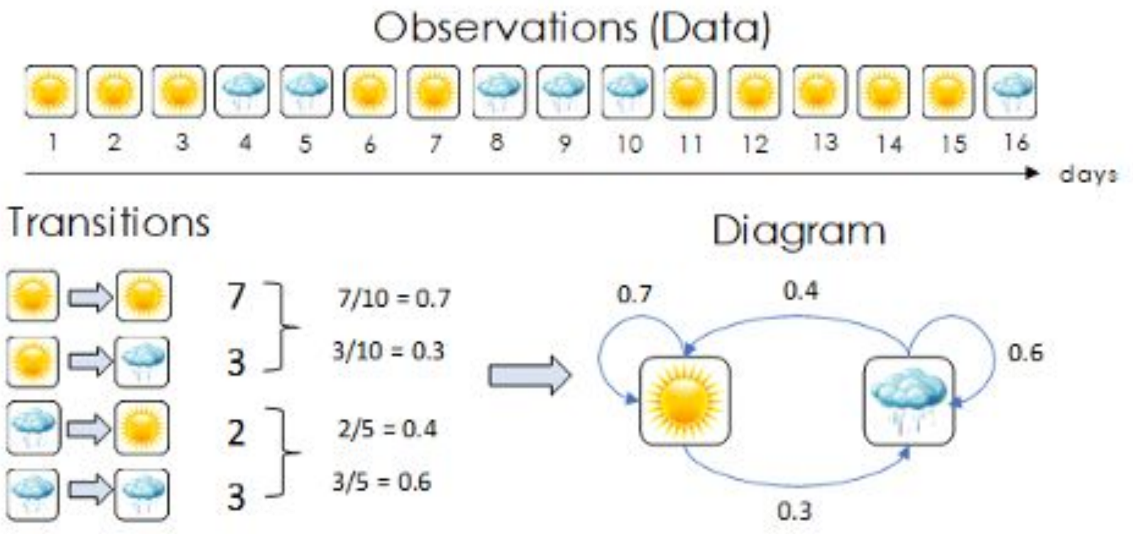
type of discrete stochastic process in which the probability of an event occurring only depends on the immediately previous event.

Assumption: The underlying assumption is that the “future is independent of the past given the present”. In other words, if we know the present state or value of a system or variable, we do not need any past information to try to predict the future states or values.
Markov chains are generally defined by a set of states and the transition probabilities between each state.These transition probabilities are usually represented in the form of a Matrix, called the Transition Matrix, also called the Markov Matrix.
How do we calculate these probabilities? - *from data.*

Modeling a Stochastic Process

Calculating the probabilities - *Transition* + *Initial*

State Space: *{Sunny , Rainy}*



	Sunny	Rainy
Sunny	0.7	0.3
Rainy	0.4	0.6

State Transition Diagram

Initial state distribution/Probabilities:

Sunny = 10 / 16 = 0.625

Rainy = 6 / 16 = 0.375

π = { S = 0.625, R = 0.375}

Example Problems

We have Weather Historical Data.

State Space:
{ Sunny = S, Rainy = R, Cloudy = C}

States initial Probabilities:
{ 0.7, 0.25, 0.05 }

States Transition Probabilities:

	S	**R**	**C**
S	0.8	0.15	0.05
R	0.38	0.6	0.02
C	0.75	0.05	0.2

Q1: Given that today is Sunny, what is the Probability that tomorrow is Cloudy?

Sol: P(Cloudy / Sunny) = 0.05 - From transition probability matrix

Q2: Given that today is Sunny, what is the probability that tomorrow is Sunny, and day after is Raining?

Sol: $P(R_3/S_2) * P(S_2/S_1) = 0.15 * 0.8 = 0.12 = 12\%$

General Rule: $P(t_1, t_2, \ldots, t_n) = \Pi_{i=1}^{n} P(t_i / t_{i-1})$

Q3: Given that today is Cloudy and yesterday was Raining. What is the probability that tomorrow would be Sunny?

Sol: $P(S_3/C_2) * P(C_2/R_1) = 0.75*0.02=0.015=1.5\%$

Q4: What is the Probability of given series?

S - R - R - R - C - C

Sol: $P(S_1)*P(R_2/S_1)*P(R_3/R_2)*P(R_4/R_3)*P(C_5/R_4)*P(C_6/C_5)$
$P(S_1)$ - initial state doesn't depend on anyone. So its P(S)

Hidden Markov Model (HMM)

In many cases, the events we are interested in might be not directly observable, but hidden. Instead, we have some evidence/observation for them.

E.g. Sitting in room and seeing a person entered with umbrella.

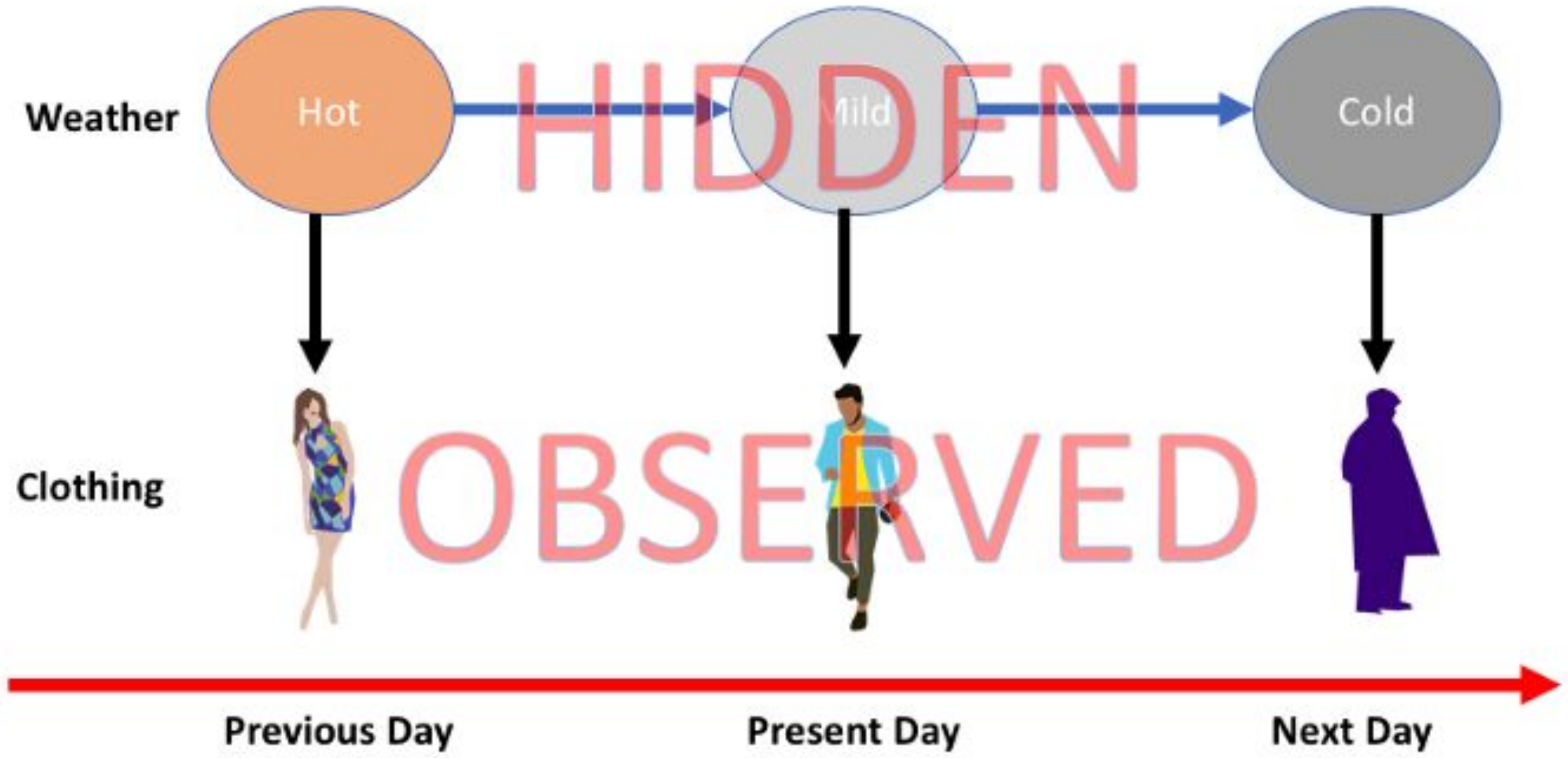
HMM model such events where states itself are hidden, but have some evidence as visible states.

Hidden Markov Process: A markov process where the states we are interested in are hidden themselves and have observation as evidence, are called Hidden markov process.

The hidden states form a Markov chain, and the probability distribution of the observed symbol depends on the underlying state.

- HMM models such processes.
- Generally used in predicting states (hidden) using sequential data like weather, text, speech etc.

Hidden Markov Models in general (both supervised and unsupervised) are heavily applied to model sequences of data.


Weather
Hot
HIDDEN
Mild
Cold
Clothing
OBSERVED
Previous Day
Present Day
Next Day

Modeling an Hidden Markov Process

To model a hidden markov process, we need to have

1. **States:** Hidden, Observable
2. **Probabilities:** Initial/Prior, Transition, Emission

Initial: *Hidden states prior probabilities - From data/history*
Transition: *State to State*
Emission: *State to Observation/Evidence*

Priors

Hot	0.6
Mild	0.3
Cold	0.1

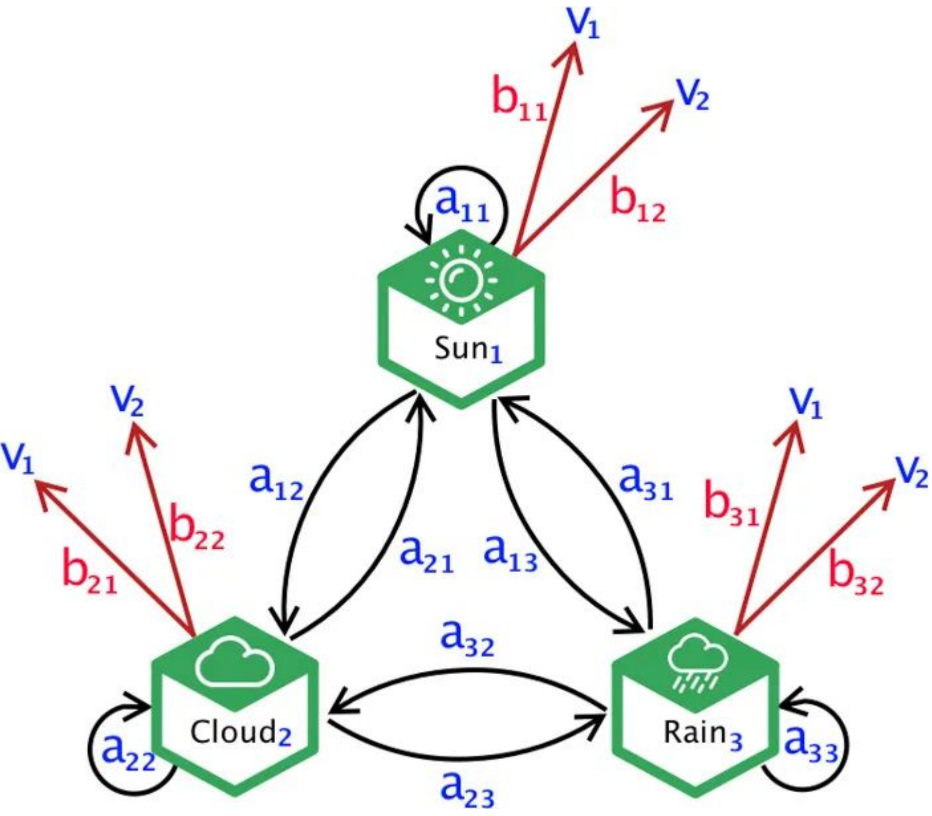
Transitions

	Hot	**Mild**	**Cold**
Hot	0.6	0.3	0.1
Mild	0.4	0.3	0.2
Cold	0.1	0.4	0.5

Emissions

	Hot	**Mild**	**Cold**
Casual Wear	0.8	0.19	0.01
Semi Casual Wear	0.5	0.4	0.1
Winter apparel	0.01	0.2	0.79

State diagram with emission probabilities

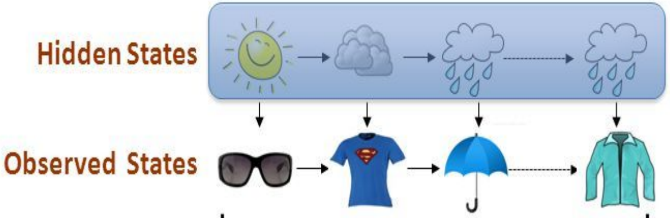


The Hidden Markov Model

	sunny	rainy	cloudy
sunny	0.8	0.05	0.15
rainy	0.2	0.6	0.2
cloudy	0.2	0.3	0.5

sum to 1

State transition probability table



State emission probability table

The probability of observing a particular observable state given a particular hidden state

	sunglasses	T-shirt	umbrella	Jacket
sunny	0.4	0.4	0.1	0.1
rainy	0.1	0.1	0.5	0.3
cloudy	0.2	0.3	0.1	0.4

sum to 1

Weisstein et al. A Hands-on Introduction to Hidden Markov Models

Three Fundamental Problems of HMM

Hidden Markov Models should be characterized by three fundamental problems:

Problem 1 (Likelihood/Evaluation): Given an HMM $\lambda = (\pi, A, B)$ and an observation sequence O, determine the likelihood $P(O|\lambda)$.

Problem 2 (Decoding): Given an observation sequence O and an HMM $\lambda = (\pi, A, B)$, discover the best hidden state sequence Q.

Problem 3 (Learning): Given an observation sequence O and the set of states in the HMM, learn the HMM parameters π, A and B.

Example Problem - (Decoding)

We are observing a person entering a room while wearing ***{Coat, Coat, Umbrella}*** for three consecutive days, Find the hidden weather sequence for these three days? Given that:

State Space: {Sunny, Rainy, Cloudy}

Observation/Evidence Space: {Short, Coat, Umbrella}

π = { S = 0.75, R = 0.2, C = 0.05}

Complexity : No of Possible state sequence= M = N^T
Where N = No of Hidden States
T = No of Observations given in question = 3
(Coat, Coat, Umbrella) are 3 observation over 3 Days

A =

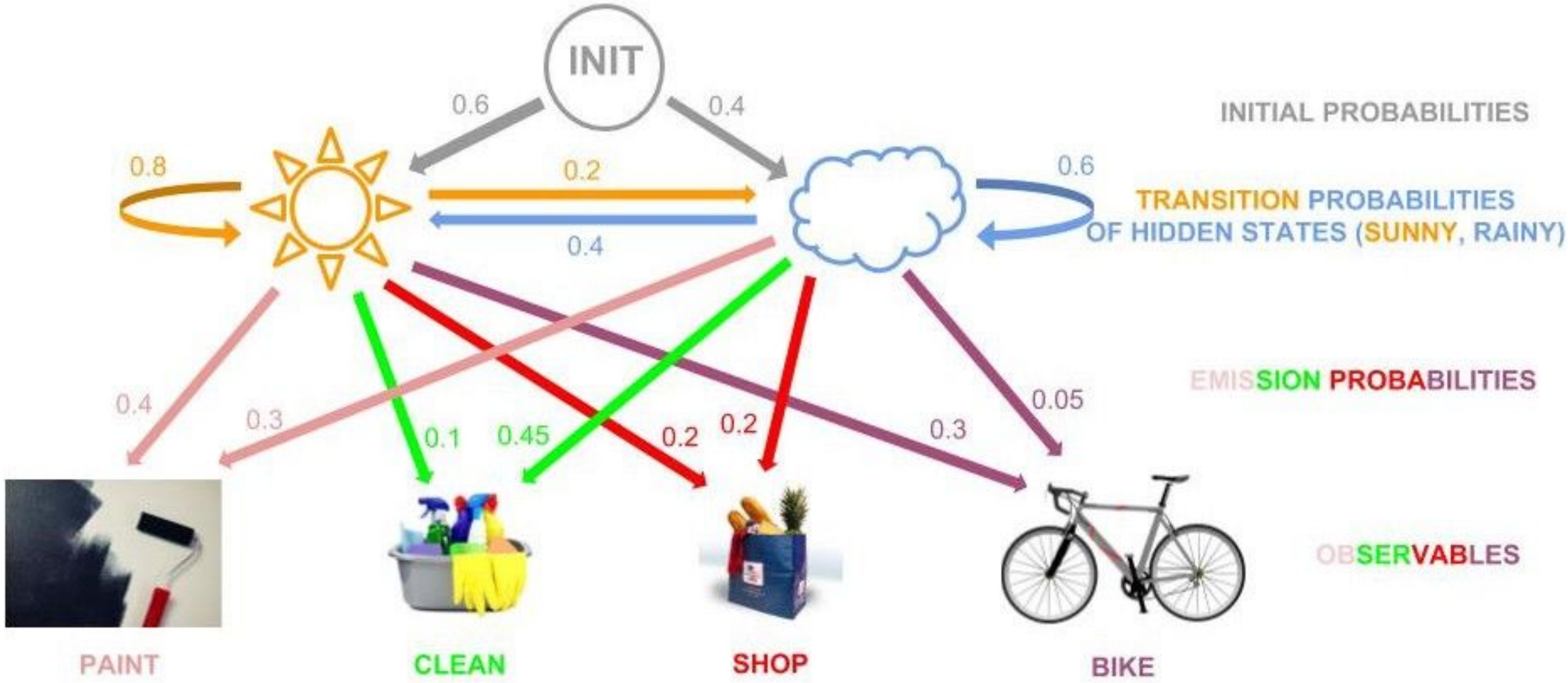
	S	R	C
S	0.8	0.15	0.05
R	0.38	0.6	0.02
C	0.75	0.05	0.2

B =

	Short	Coat	Umb
Sunny	0.6	0.3	0.1
Rainy	0.05	0.30	0.65
Cloudy	0	0.5	0.5

Example Problem-1 (Evaluation/Likelihood Problem: Forward-Backward Algorithm

Lea is our imaginary friend and during the day she does either of these four things: Painting, Cleaning the house, Biking, Shopping for groceries



Transition & Emission Probabilities

$$A = \text{from} \begin{array}{c} \\ \\ \end{array} \overset{\text{to}}{\begin{vmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{vmatrix}}$$

$$A = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

$$B = \text{from} \begin{array}{c} \\ \\ \end{array} \overset{\text{to}}{\begin{array}{cccc} \text{PAINT} & \text{CLEAN} & \text{SHOP} & \text{BIKE} \end{array}}$$

$$\begin{vmatrix} 0.4 & 0.1 & 0.2 & 0.3 \\ 0.3 & 0.45 & 0.2 & 0.05 \end{vmatrix}$$

$$B = \begin{vmatrix} b_{11} & b_{12} & b_{13} & b_{14} \\ b_{21} & b_{22} & b_{23} & b_{24} \end{vmatrix}$$

$$S = \{ S_{sunny}, S_{rainy} \} \quad (\textit{Hidden States})$$

$$O = \{ O_{clean}, O_{bike}, O_{shop}, O_{paint} \} \quad (\textit{Observables})$$

$$\pi = \begin{vmatrix} 0.6 & 0.4 \end{vmatrix} \quad (\textit{Initial Probabilities})$$

$$A = \begin{vmatrix} 0.8 & 0.2 \\ 0.4 & 0.6 \end{vmatrix} \quad (\textit{Transition Probabilities})$$

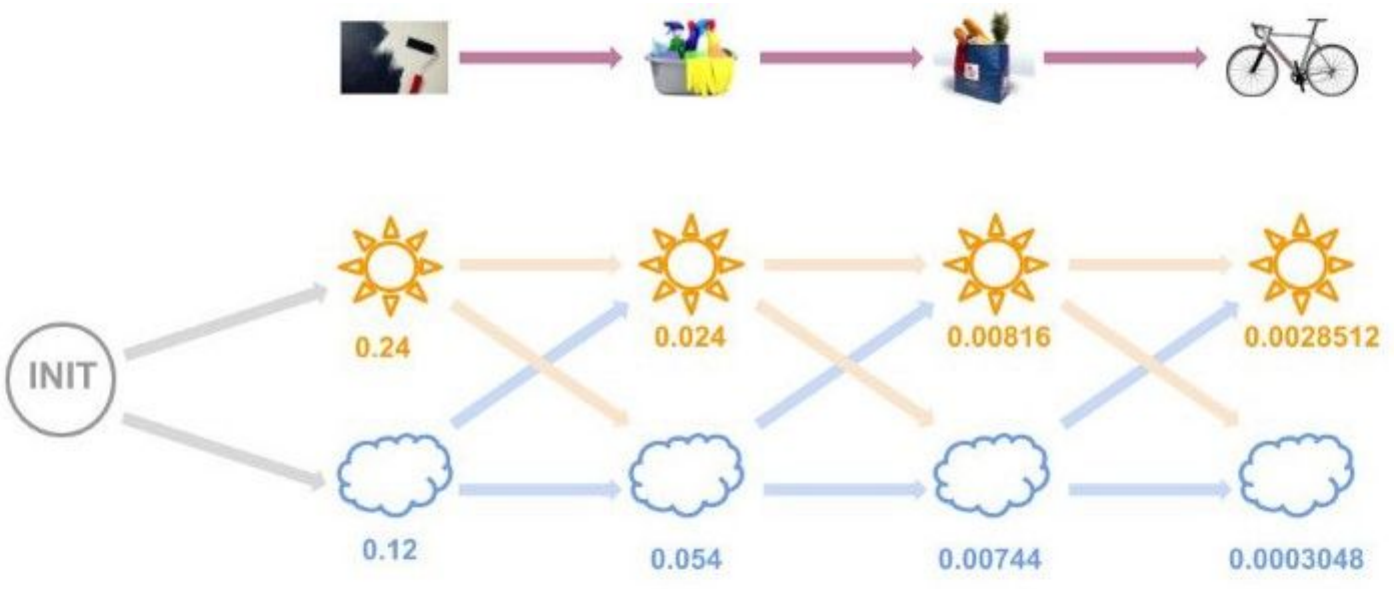
$$B = \begin{vmatrix} 0.4 & 0.1 & 0.2 & 0.3 \\ 0.3 & 0.45 & 0.2 & 0.05 \end{vmatrix} \quad (\textit{Emission Probabilities})$$

Forward Algorithm - 3 Steps

Initialization $$\alpha_1(i) = \pi_i \cdot b_i(O_1)$$ First forward variable

Recursion $$\alpha_{t+1}(j) = \sum_{i=1}^{N} \alpha_t(i) \cdot a_{ij} \cdot b_j(O_{t+1})$$ Finding the next Forward variable

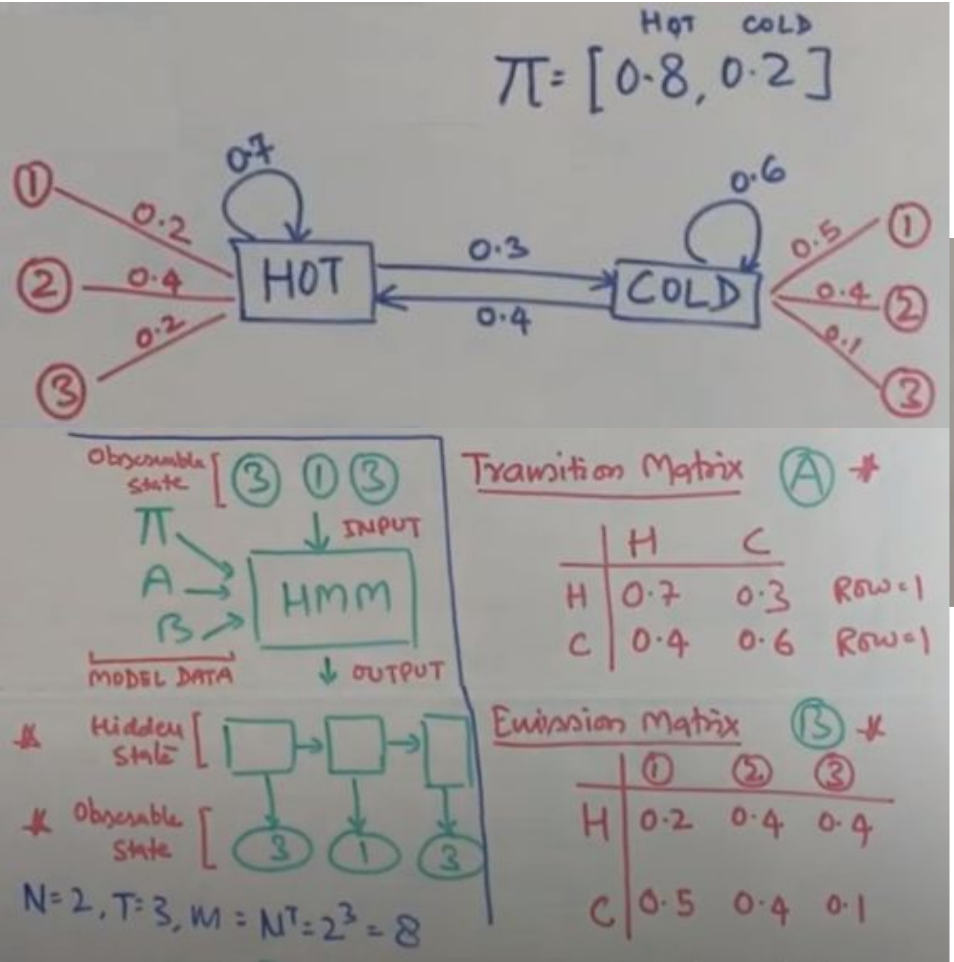
Termination $$P(O|\lambda) = \sum_{i=1}^{N} \alpha_T(i)$$ Sum up all the forward variables at time T, i.e. all the variables of every state at the end of the observation sequence.

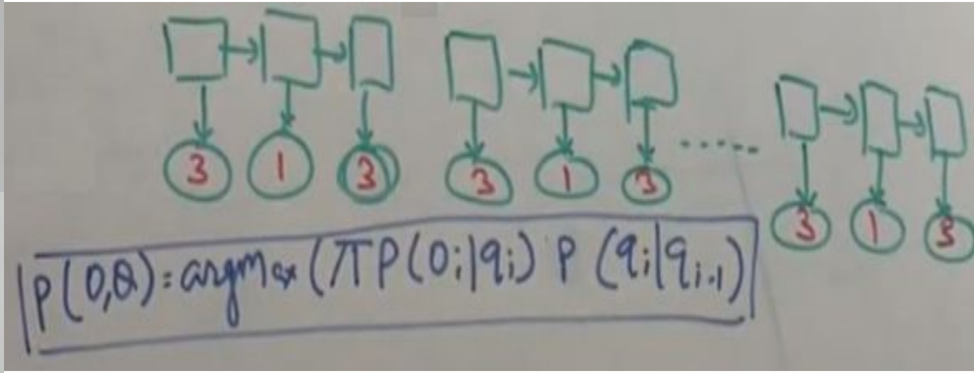


$$P(O|\lambda) = 0.0028512 + 0.0003048 = 0.003156$$

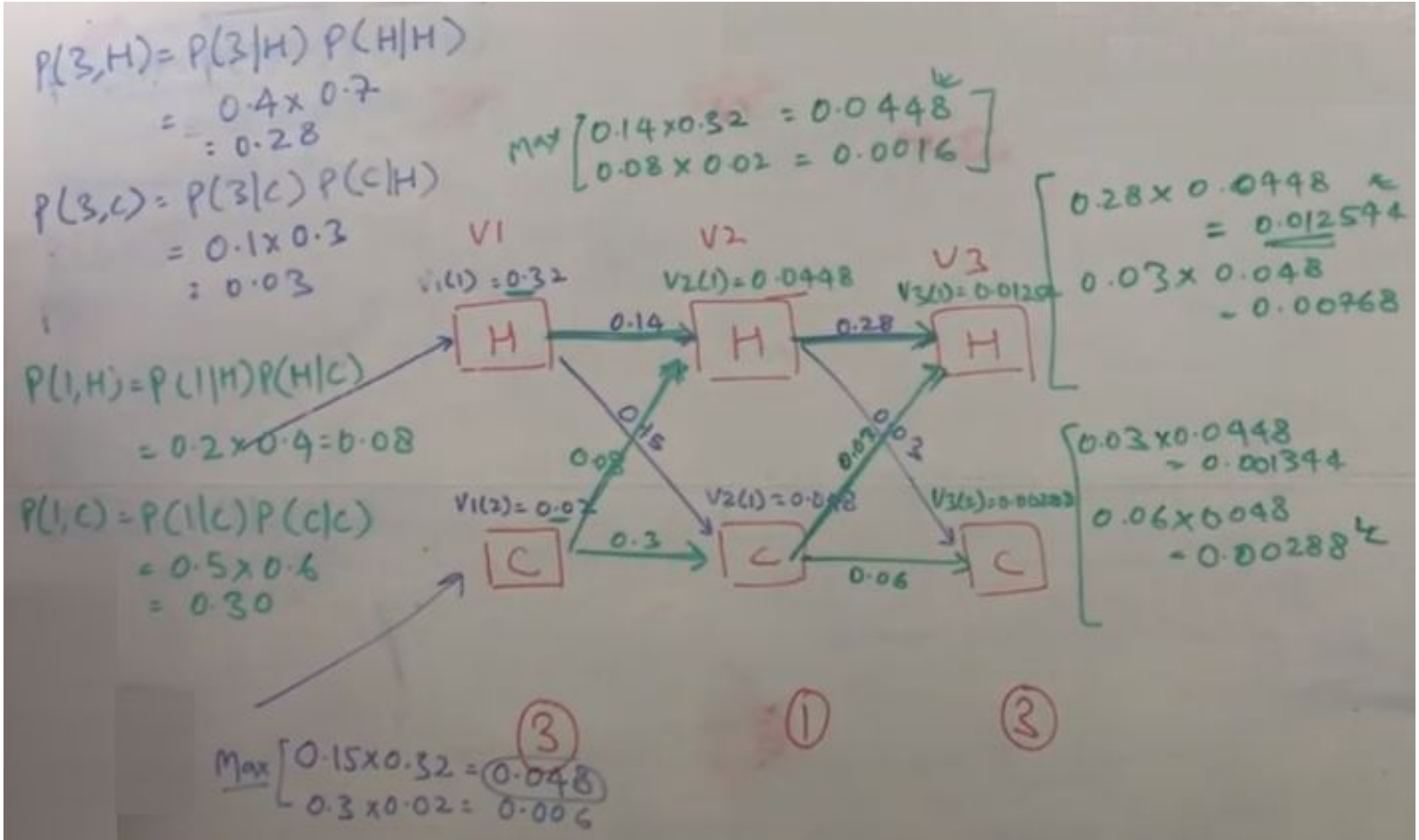
Hence, we can say that the probability that Lea has spent the last four days painting, cleaning, shopping and biking, is 0.003156 given the HMM model we have built for her.

Example Problem 2 - (Decoding Problem: Viterbi Algorithm)





Viterbi Algorithm



Important Links

1. https://www.youtube.com/watch?v=onSi24IM47U&list=PLIRnO_sdVuEfNSksORUz5xzlI79AVtAkz
2. https://medium.com/@Ayra_Lux/hidden-markov-models-part-1-the-likelihood-problem-8dd1066a784e
3. https://medium.com/@Ayra_Lux/hidden-markov-models-part-2-the-decoding-problem-c628ba474e69
4.

???