Machine Learning

Hidden Markov Models

Satishkumar L. Varma

Department of Information Technology
SVKM's Dwarkadas J. Sanghvi College of Engineering, Vile Parle, Mumbai.
ORCID | Scopus | Google Scholar | Google Site | Website

Outline

- Classification
  - Bayesian Belief Networks
  - Hidden Markov Models
  - Support Vector Machine
    - Maximum Margin Linear Separators
    - Quadratic Programming solution to finding maximum margin separators
    - Kernels for learning non-linear functions
  - Classification using k Nearest Neighbour Algorithm

Hidden Markov Models

- Hidden Markov models (HMMs) are sequence models.
- That is, given a sequence of inputs, such as words
- HMM will compute a sequence of outputs of the same length.
- HMM model is a graph where nodes are probability distributions over labels and edges give the probability of transitioning from one node to the other.
- When working with sequences of data,
  - we often face situations where we can’t directly see the important factors that influence the datasets.
  - HMM help solve this problem by predicting these hidden factors based on the observable data.

Hidden Markov Models

- Hidden Markov models (HMMs) are sequence models.
- That is, given a sequence of inputs, such as words
- HMM will compute a sequence of outputs of the same length.
- HMM model is a graph where nodes are probability distributions over labels and edges give the probability of transitioning from one node to the other.
- When working with sequences of data,
  - we often face situations where we can't directly see the important factors that influence the datasets.
  - HMM help solve this problem by predicting these hidden factors based on the observable data.

Hidden Markov Models

- HMM in Machine Learning
- It is an statistical model.
- Used to describe probabilistic relationship bet. a sequence of observations and a sequence of hidden states.
- Used in situations where observations is unknown or hidden, hence it known as "Hidden Markov Model."
- HMM consists of two types of variables: hidden states and observations.
  - Hidden states: Variables that generate the observed data, but they are not directly observable.
  - Observations: Variables that are measured and observed.
- Relationship between the hidden states and the observations is modeled using a probability distribution.
- HMM is the relationship between the hidden states and the observations using two sets of probabilities:
  - the transition probabilities and the emission probabilities.
- Transition probabilities: describe the probability of transitioning from one hidden state to another.
- Emission probabilities: describe the probability of observing an output given a hidden state.

Hidden Markov Models

- Hidden Markov Model Algorithm: Implementation Steps:
- Step 1: Define the state space and observation space:
  - The state space is the set of all possible hidden states, and the observation space is the set of all possible observations.
- Step 2: Define the initial state distribution:
  - This is the probability distribution over the initial state.
- Step 3: Define the state transition probabilities:
  - These are the probabilities of transitioning from one state to another.
  - This forms the transition matrix, which describes the probability of moving from one state to another.
- Step 4: Define the observation likelihoods:
  - These are the probabilities of generating each observation from each state.
  - This forms emission matrix, which describes probability of generating each observation from each state.

Hidden Markov Models

- Hidden Markov Model Algorithm: Implementation Steps:
- Step 5: Train the model:
  - The parameters of the state transition probabilities and the observation likelihoods are estimated using the Baum-Welch algorithm, or the forward-backward algorithm.
  - This is done by iteratively updating the parameters until convergence.
- Step 6: Decode the most likely sequence of hidden states:
  - Given observed data, the Viterbi algorithm is used to compute the most likely sequence of hidden states.
  - This can be used to predict future observations, classify sequences, or detect patterns in sequential data.
- Step 7: Evaluate the model:
  - The performance of the HMM can be evaluated using various metrics, such as
    - accuracy,
    - precision,
    - recall, or F1 score.

Hidden Markov Models

- HMM can be used to identify underlying patterns or structures in sequential data.
- HMM application domains:
  - Machine learning
  - Natural language processing
  - Speech recognition
  - Bioinformatics
  - Gene analysis
  - Time-series forecasting
- HMM Example: Speech recognition tasks
  - It helps to measure the probability of a certain word or lack of words occurring in a given audio recording.
  - Occurrence of specific words or silence in the recording can represent states
  - Volume of speech throughout the recording can represent observations
  - By knowing observations (volume), HMM helps to determine
    - the likelihood of hidden states or
    - words and lack of words and
    - predict the most probable word being spoken.

Hidden Markov Models

- HMM are probabilistic frameworks.
- HMM model data as a series of outputs generated by one of several (hidden) internal states.
- HMM pertains to stochastic processes where states can be hidden or not directly visible to the observer.
- HMM used to identify underlying patterns or structures in sequential data.
- HMM is used when we can’t observe the states of a stochastic process.
- However, we can only see the result of some probability function (observation) of the states.

Hidden Markov Models

- HMM: Assumptions
- 1. Output Independence Assumption
    - Output observation is conditionally independent of all other hidden states and
    - all other observations when given the current hidden state.
    - Output independence assumption:
        - $P(x_t = v_i / z_t = s_j) = P(x_t = v_i / x_1, x_2,\ldots x_T, z_1, z_2,\ldots, z_T) = B_{ji}$
- 2. Emission Probability Matrix
    - Probability of hidden state generating output v_i given that state at the corresponding time was s_j

Hidden Markov Models

- MM Assumptions
- 1. Limited Horizon Assumption
  - Probability of being in a state at a time t depend only on the state at the time (t-1).
  - Limited horizon assumption equation
    - $P(z_t, z_{t-1}, z_{t-2}, \ldots, z_1) = (z_t / z_{t-1})$
  - That means state at time t represents enough summary of the past to reasonably predict the future.
  - This assumption is an order-1 Markov process.
  - An order-k Markov process assumes conditional independence of state z_t from the states that are k + 1-time steps before it.
- 2. Stationary Process Assumption
  - Conditional (probability) distribution over the next state, given the current state, doesn't change over time.
  - Stationary process assumption
    - $P(z_t / z_{t-1}) = (z_2 / z_1)$ where $t \in 2,\ldots,T$
  - That means states keep on changing over time but the underlying process is stationary.

Hidden Markov Models

- Markov Model versus HMM
- Both Markov Model and HMM handle data that can be represented as a sequence of observations over time.
- Markov Model (MM) or Markov chain concerns stochastic (random) process states that are visible to the observer. Mathematically: Markov models versus HMM
- MM: Series of (hidden) states $z=\{z_1, z_2 \ldots\ldots\ldots\ldots\}$
    - drawn from state alphabet $S = \{s_1, s_2, \ldots\ldots s_{|S|}\}$ where z_i belongs to S.
- HMM: Series of observed output $x = \{x_1, x_2, \ldots\ldots\}$
    - drawn from an output alphabet $V = \{v_1, v_2, \ldots, v_{|v|}\}$ where x_i belongs to V.
- Questions Answered in a Markov Model
    - What is the probability of particular sequences of state z?
    - How do we estimate parameter of state transition matrix A to maximize likelihood of observed sequence?
- Questions Answered by Hidden Markov Model:
    - What is the probability of an observed sequence?
    - What is the most likely series of states to generate an observed sequence?
    - How can we learn the values for the HMMs parameters A and B given some data?

Hidden Markov Models

- Markov Model
- Questions Answered in a Markov Model
  - What is the probability of particular sequences of state z?
  - How do we estimate parameter of state transition matrix A to maximize likelihood of observed sequence?

Hidden Markov Models

- Notation Convention
  - There is an initial state and an initial observation $z_0 = s_0$
  - s_0: Initial probability distribution over states at time 0.
  - Initial state probability: (π)
  - At t = 1, probability of seeing first real state z_1 is $p(z_1/z_0)$
  - Since $z_0 = s_0$: $P(Z_t / Z_{t-1}, Z_{t-2},\ldots Z_1) = P(Z_t / Z_{t-1}, Z_{t-2},\ldots Z_1, Z_0)$
- State Transition Matrix
- Consider state transition matrix of four states, including the initial state:

$P(S_{hot} / S_0)$

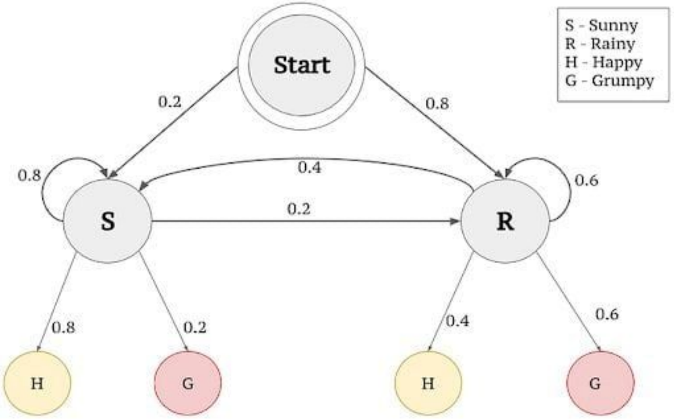
i / j	S_0	S_{hot}	S_{cold}	S_{rain}
S_0		0.33	0.33	0.34
S_{hot}	0	0.8	0.1	0.1
S_{cold}	0	0.2	0.6	0.2
S_{rain}	0	0.1	0.2	0.7

Hidden Markov Models

- Questions Answered in a Markov Model
  - What is the probability of particular sequences of state z?
  - **Ai,j**: probability of transitioning from state i to state j at any time t.
  - $P(Z) = P(Z_t, Z_{t-1}, Z_{t-2},\ldots Z_1; A)$
  - $P(Z) = P(Z_t / Z_{t-1}, Z_{t-2},\ldots Z_1, Z_0; A)$ Initial State
  - $P(Z) = P(Z_t / Z_{t-1}, Z_{t-2},\ldots Z_1, Z_0; A)\, P(Z_{t-1} / Z_{t-2},\ldots, Z_0; A) \ldots P(Z_1 / Z_0; A)$ Chain rule of probability
  - $P(Z) = P(Z_t / Z_{t-1}; A)\, P(Z_{t-1} / Z_{t-2}; A) \ldots P(Z_2 / Z_1; A)\, P(Z_1 / Z_0; A)$ Limited Horizon Assumption
  - $P(Z) = \pi\, P(Z_t, Z_{t-1}; A)$
  - $P(Z) = \pi\, Az_1, z_t$
- Consider the state transition matrix above;
- Q: Finding probability of particular sequence i.e determine the probability of sequence:
  - $\{z_1 = s_{hot}, z_2 = s_{cold}, z_3 = s_{rain}, z_4 = s_{rain}, z_5 = s_{cold}\}$
  - $P(z) = P(s_{hot}|s_0)\, P(s_{cold}|s_{hot})\, P(s_{rain}|s_{cold})\, P(s_{rain}|s_{rain})\, P(s_{cold}|s_{rain})$
  - $= 0.33 \times 0.1 \times 0.2 \times 0.7 \times 0.2 = 0.000924$

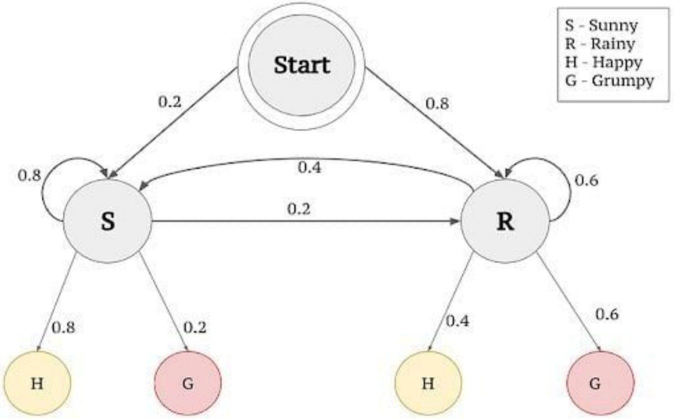
Hidden Markov Models

- Example 1: Markov Model
  - State machine of MM gives person feeling in different climates. (see Markov model as finite state machine)
  - Set of states: (S) = {Happy, Grumpy}
  - Set of hidden states: (Q) = {Sunny , Rainy}
  - State series over time: = $z \in S_T$
  - Observed states for four day: {z_1 = Happy, z_2 = Grumpy, z_3 = Grumpy, z_4 = Happy}
  - Feelings are observed through emotions.
  - Observation: feeling that you understand from a person emoting. (since we observe them.)
  - Hidden state: Weather that influences the feeling of a person. (since we can't observe it.)



Hidden Markov Models

- Example 1: Hidden Markov Model
- Emission Probabilities
  - In this example, feelings (“Happy” or “Grumpy”) can be only observed.
  - A person can observe that a person has an 80 percent chance to be “happy” given that the climate at the particular point of observation is sunny.
  - Similarly there’s a 60 percent chance of a person being “grumpy” given that the climate is rainy.
  - The 80 percent and 60 percent are emission probabilities since they deal with observations.
- Transition Probabilities
  - When we consider the climates (hidden states) that influence the observations,
    - there are correlations between consecutive days being sunny or alternate days being rainy.
  - There is an 80 percent chance for the Sunny climate to be in successive days,
    - whereas there’s a 60 percent chance for it to be rainy on consecutive days.
  - Prob. that explain the transition to/from hidden states are transition prob.



Hidden Markov Models

- Hidden Markov Model
- Questions Answered by Hidden Markov Model:
  - What is the probability of an observed sequence?
  - What is the most likely series of states to generate an observed sequence?
  - How can we learn the values for the HMMs parameters A and B given some data?

Satishkumar L. Varma

Hidden Markov Models

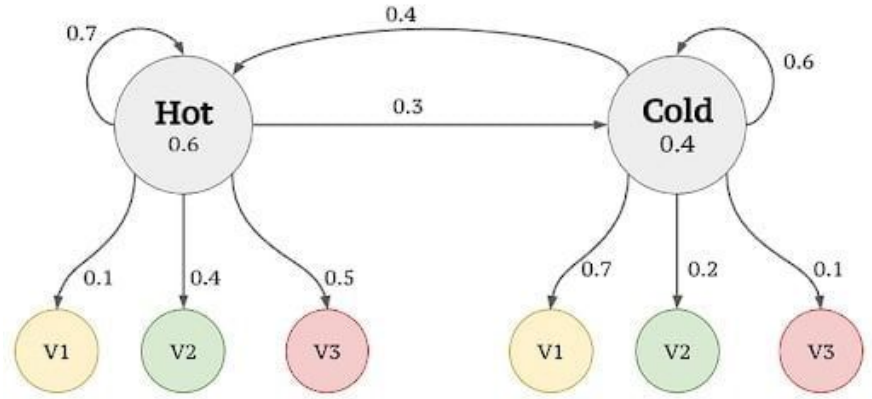
- Example 2: Questions Answered by Hidden Markov Model:
  - What is the probability of an observed sequence?
  - S = {hot, cold}
  - v = {v_1 = 1 ice cream, v_2= 2 ice creams, v_3= 3 ice creams},
  - where V is the Number of ice creams consumed in a day.
  - Example Sequence: = {x_1=v_2, x_2=v_3, x_3= v_1, x_4=v_2}
  - Consider this given data as matrices:
  - Consider the generated finite state machines (FSM) for HMM.

$$\underset{\text{(State Transition Matrix)}}{A} = \begin{array}{c c} & \begin{array}{cc} H & C \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{Bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{Bmatrix} \end{array}$$

$$\underset{\text{(Emission Matrix)}}{B} = \begin{array}{c c} & \begin{array}{ccc} V1 & V2 & V3 \end{array} \\ \begin{array}{c} H \\ C \end{array} & \begin{Bmatrix} 0.1 & 0.4 & 0.5 \\ 0.7 & 0.2 & 0.1 \end{Bmatrix} \end{array}$$

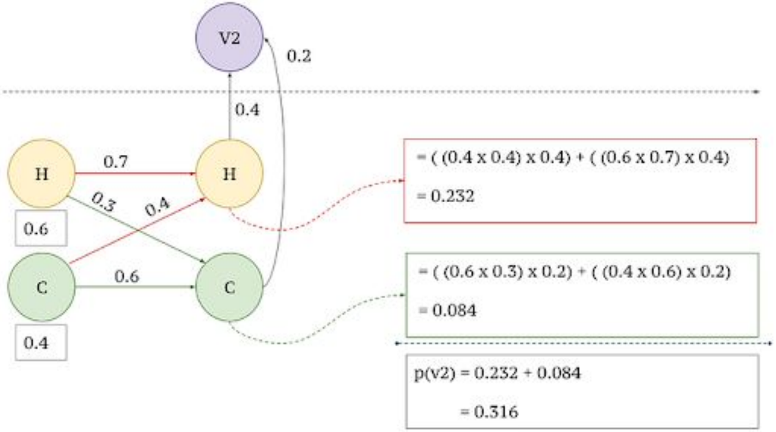
$$\underset{\text{(Initial state S0)}}{\pi} = \begin{array}{c} \begin{array}{cc} H & C \end{array} \\ \begin{Bmatrix} 0.6 & 0.4 \end{Bmatrix} \end{array}$$

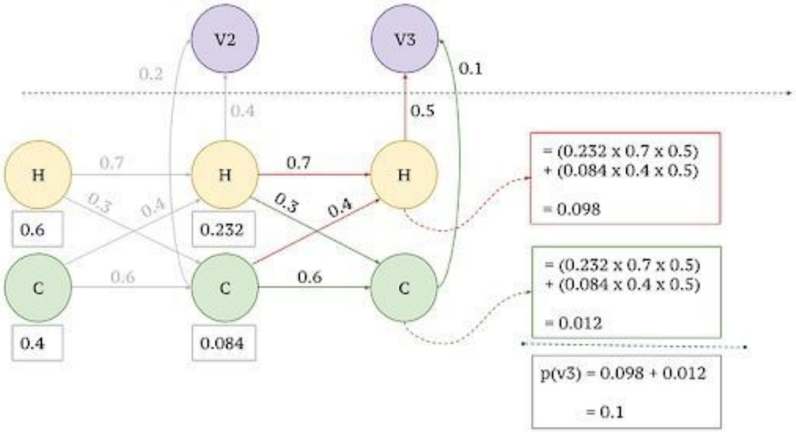
$$\underset{\text{(State Transition Matrix)}}{A} = \begin{array}{c c} & \begin{array}{cc} H & C \end{array} \\ \begin{array}{c} \pi \\ H \\ C \end{array} & \begin{Bmatrix} 0.6 & 0.4 \\ 0.7 & 0.3 \\ 0.4 & 0.6 \end{Bmatrix} \end{array}$$



Hidden Markov Models

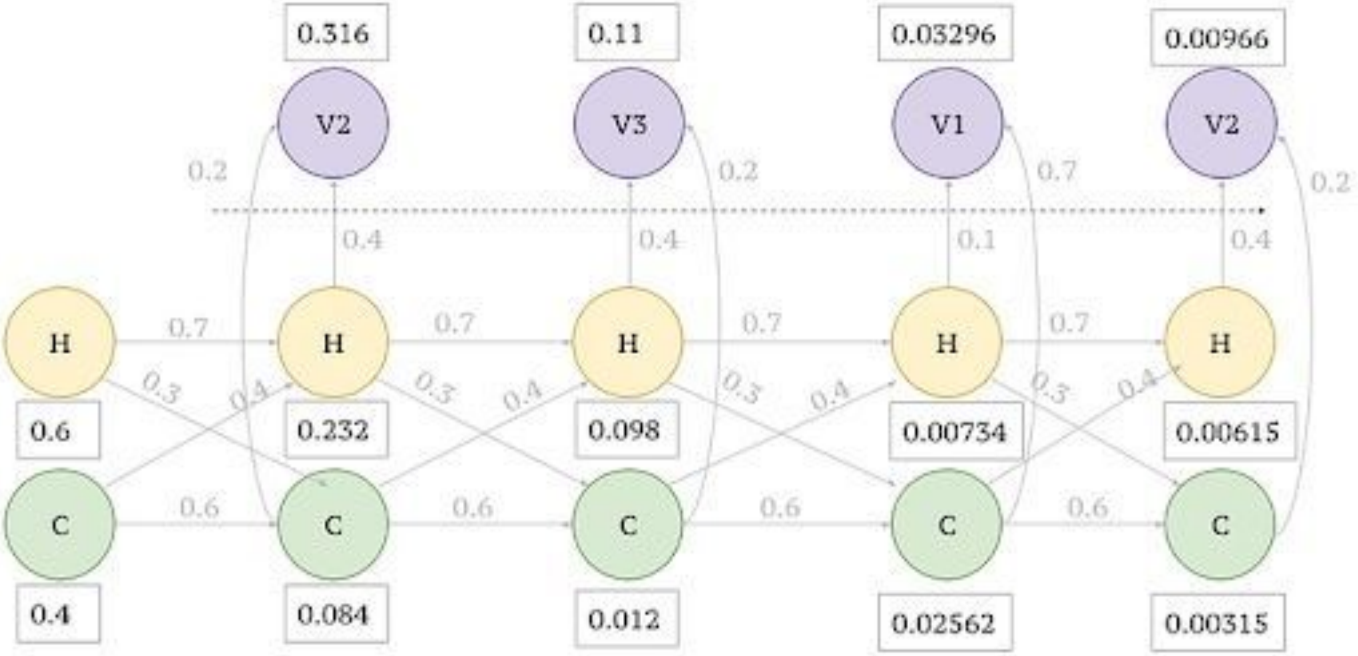
- Step 1: Calculate the prior probabilities i.e probability of being hot or cold previous to any actual observation.
  - This can be obtained from S_0 or π.
  - From the given data as matrices, S_0 is provided as 0.6 and 0.4 (which are prior probabilities).
- Step 2: Calculate the probability of a given sequence based on Markov and HMM assumptions (4 steps of HMM)
- Step 2.1: Observed Output for $x_1 = v_2$ (see left figure for 1st step of HMM)
- Step 2.2: Observed Output for $x_2 = v_3$ (see right figure for 2nd step of HMM)
- Step 2.3: Observed Output for $x_3 = v_1$ (simply multiply the paths that lead to v_1.)
- Step 2.4: Observed Output for $x_4 = v_2$ (simply multiply the paths that lead to v_2.)
- Step 3. Find Maximum Likelihood





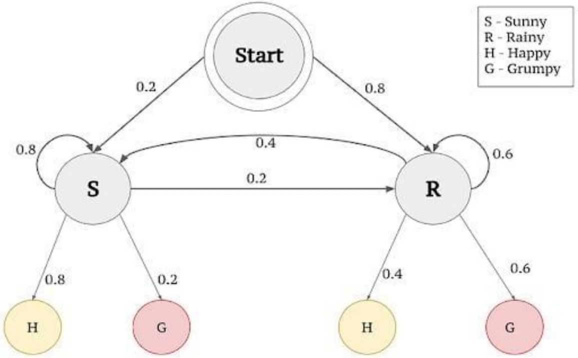
Hidden Markov Models

- Step 2: Calculate the probability of a given sequence (last 2 steps of HMM)
- Step 2.3: Observed Output for $x_3 = v_1$ (simply multiply the paths that lead to v_1) (see right figure for 3rd step of HMM)
- Step 2.4: Observed Output for $x_4 = v_2$ (simply multiply the paths that lead to v_2) (see right figure for 4th step of HMM)



Hidden Markov Models

- Step 3. Find Maximum Likelihood (using the forward procedure).
  - For a given observed sequence of outputs $x \epsilon V_T$
  - we intend to find the most likely series of states $z \epsilon S_T$
- Observe (right top/ left bottom) given data for MM as finite state machine.
- Markov model as a finite state machine (see, right bottom).



S - Sunny
R - Rainy
H - Happy
G - Grumpy

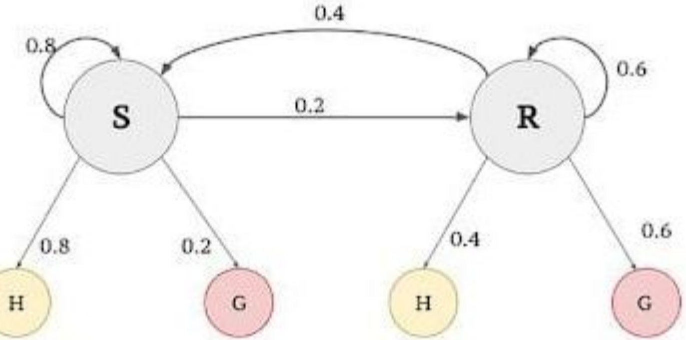
Climate	S	S	S	S	R	R	R	S	S	S	S	R	R	S	S	S
Emotion	G	H	H	H	G	H	H	G	H	H	H	G	H	H	H	H

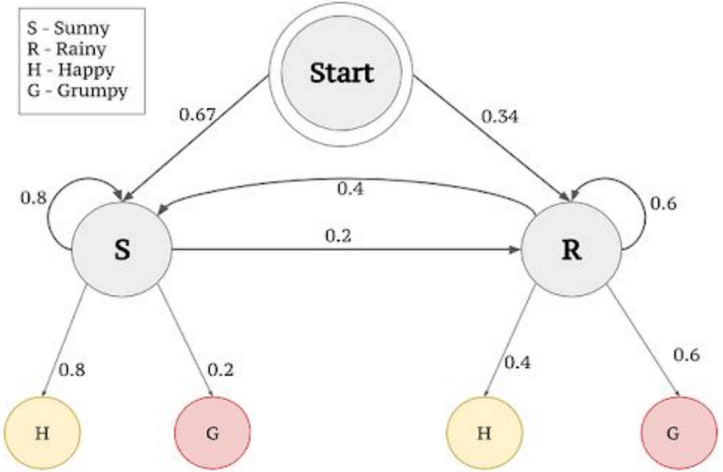
S -> S = 8 0.8

S -> R = 2 0.2

R -> R = 3 0.6

R -> S = 2 0.4





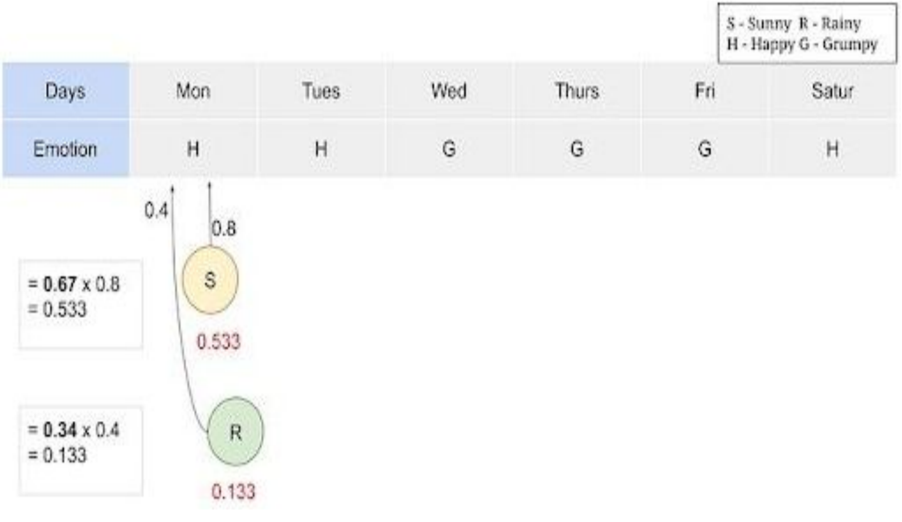
Hidden Markov Models

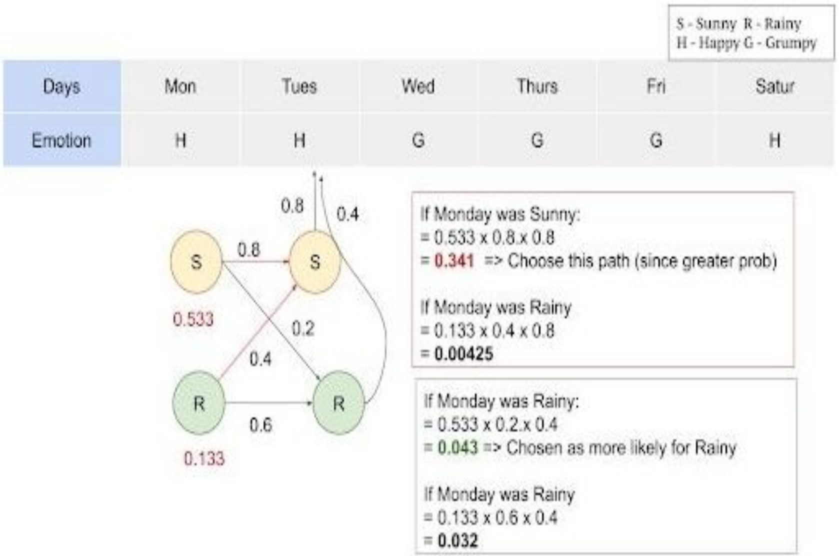
- OR Step 3. Find Maximum Likelihood using Viterbi algorithm in Two Steps
  - It is a dynamic programming algorithm similar to the forward procedure.
  - Here, instead of tracking the total probability of generating the observations,
- Viterbi algorithm tracks the maximum probability and the corresponding state sequence.
- Example
  - Let us take a sequence of emotions: H, H, G, G, G, H for six consecutive days.
  - Using the Viterbi algorithm we will find out the more likelihood of the series.
- The Viterbi algorithm requires to choose the best path (See Figure)
  - There will be several paths that will lead to sunny on Saturday, and
    - many paths that lead to rainy on Saturday.
    - Here, identify the best path to sunny or rainy Saturday and
    - multiply with the transition emission probability of “Happy,”
    - since Saturday makes the person feel “Happy.”

Days	M	T	W	Th	F	S
Emotion	H	H	G	G	G	H

Hidden Markov Models

- Step 3. Find Maximum Likelihood using Viterbi algorithm in Two Steps (Step 3.1 and Step 3.2)
- Let's consider a sunny Saturday.
- The previous day, Friday, can be sunny or rainy.
- Then we need to know the best path up to Friday, and
- then multiply with emission probabilities that lead to a grumpy feeling.
- Iteratively, we need to figure out the best path at each day ending up in more likelihood of the series of days.
- Step 3.1 (see left figure) and Step 3.2 (see right figure)



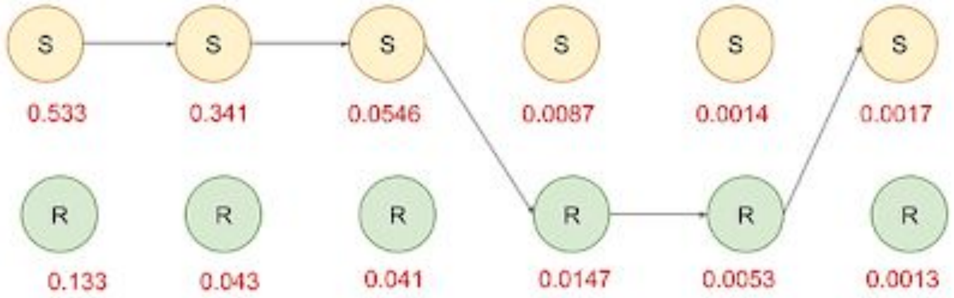


Hidden Markov Models

- Iterate the algorithm to choose the best path.
- The algorithm leaves you with maximum likelihood values and
- Now, we can produce the sequence with a maximum likelihood for a given output sequence.

S - Sunny R - Rainy
H - Happy G - Grumpy

Days	Mon	Tues	Wed	Thurs	Fri	Satur
Emotion	H	H	G	G	G	H



More likelihood sequence = **S S S R R S**
For the given output sequence H H G G G H

Hidden Markov Models

- Summary
- HMM algorithm involves
  - defining the state space, observation space, and the parameters of the state transition probabilities and observation likelihoods,
  - training the model using the Baum-Welch algorithm or the forward-backward algorithm,
  - decoding the most likely sequence of hidden states using the Viterbi algorithm, and
  - evaluating the performance of the model.

References

Text books:

1. Ethem Alpaydin, “Introduction to Machine Learning , 4th Edition, The MIT Press, 2020.
2. Peter Harrington, “Machine Learning in Action”, 1st Edition, Dreamtech Press, 2012."
3. Tom Mitchell, “Machine Learning”, 1st Edition, McGraw Hill, 2017.
4. Andreas C, Müller and Sarah Guido, “Introduction to Machine Learning with Python: A Guide for Data Scientists”, 1ed, O'reilly, 2016.
5. Kevin P. Murphy, “Machine Learning: A Probabilistic Perspective”, 1st Edition, MIT Press, 2012."

Reference Books:

6. Aurélien Géron, “Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow”, 2nd Edition, Shroff/O'Reilly, 2019.
7. Witten Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal., “Data Mining: Practical machine learning tools and techniques”, 1st Edition, Morgan Kaufmann, 2016.
8. Han, Kamber, “Data Mining Concepts and Techniques”, 3rd Edition, Morgan Kaufmann, 2012.
9. Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar, “Foundations of Machine Learning”, 1ed, MIT Press, 2012.
10. H. Dunham, “Data Mining: Introductory and Advanced Topics”, 1st Edition, Pearson Education, 2006.
11. Help for images: Vivek Vinushanth Christopher

Thank You.