Hidden Markov Models

Shreyas Chennamaraja

April 2020

1 Introduction

Hidden Markov Models (HMMs) are types of probabilistic models, a subset/application of a Bayesian classification framework to be exact. Furthermore, they are reliant on time-series data because of their feature of transition probability, as we will explore later. In a real world setting, Hidden Markov Models have numerous applications, such as Object Localization (ex. Robots), Speech/Noise Recognition, Biological sequence analysis (annotating DNA sequences with functionally important elements), and annotating text sequence with the role of words in a sentence, like denoting part of speech. **In general,** Hidden Markov Models tell us the probabilities of the sequence of random variables/states

1.1 Case example

Suppose that for a couple weeks, you are able to go outside and look at the weather. From these observed events, you could probably use your observations to predict what the weather forecast might be for the next couple weeks. This would be an example of a Markov Model. However, let's say you were locked in a room, and whenever your friend visit you with a happy attitude, that event would most likely correspond to a sunny day, when your friend visits you with a grumpy attitude that event would most likely correspond to a stormy day, and when your friend visits with a confused attitude, that would correspond to a snowy day. In this case, the states outside are hidden, but you can observe an "observable", which is the attitude of your friend. This would represent an HMM, where you would have to use the observables to predict the state (the weather).

2 The Basics

2.1 Parts of a Regular Markov Model

In our example, we had three types of weather: sunny, rainy, stormy, which would correspond to the set of states "S_1, S_2, S_3"

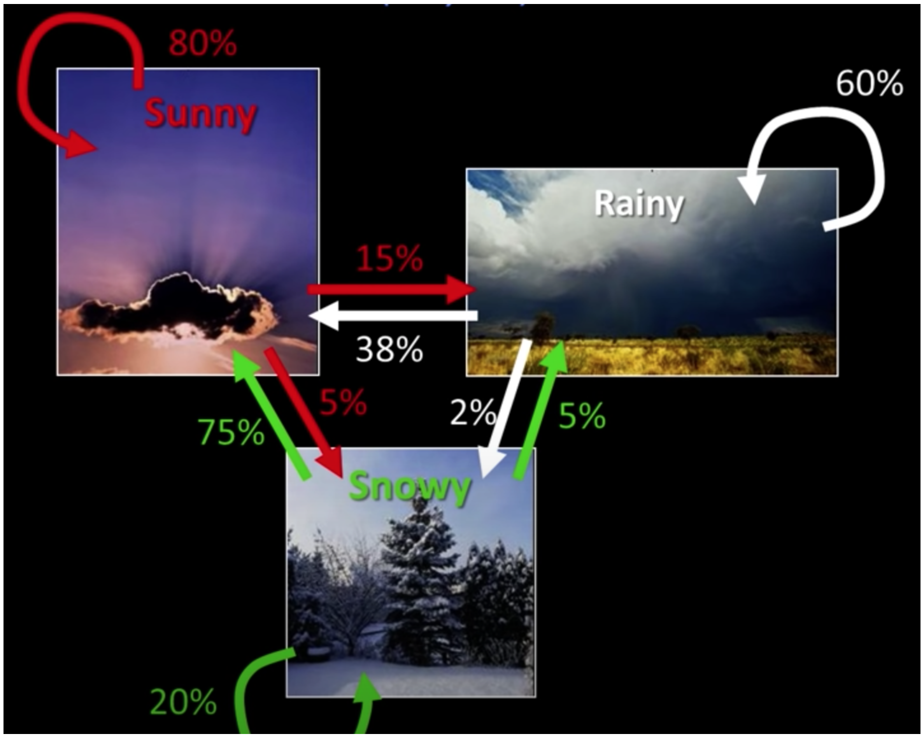

Figure 1: Markov System with Example States

In a Markov System, we would also need to specify the likely next state for any given current state. This shown by the following figure.

If the only prediction for the next state, is the current state, that is called a "1st Order Markovian System". The next state we need are the "State transition probabilities". It is the probability that if you are in state i, the next state you would be in is state j. In a real problem, this would be in the shape of a **matrix** This is denoted by:

$$a_{ij} = P(q_{t+1} = S_i | q_t = S_j)$$

The last state is an initial state distribution, which is the probability that "q" is equal to "s". In this case, q sub t is that state at time t, so q sub 1 is the state at the first day/time 1. Basically, it is the probabilties of each event being the **starting probabilities.** This is denoted by:

$$\pi_i = P[q_1 = S_i]$$

2.2 Emission Probabilities and Observation Probability with HMM

In the previous example, we talked about how you use the "observables" to predict the hidden states. Now, the relationship between the observable (attitude) and the state (weather), isn't direct. Therefore, in an HMM, you would need to know the **emission probabilities** of each observed evidence. This would also be in a form of a matrix. This is denoted by the equation for an emission probability:

$$b_j(k) = P(o_t = k | q_t = S)$$

This would be interpreted as, the probability of seeing symbol k, given that at a time, you're in a particular state. In our earlier case, one example could be "The probability that you see a happy attitude given a sunny day." Since we cannot observe the states, we would only simply ask about the likelihood of seeing a specific sequence of observations.

Next, we have an **"Observation probability"**, which is the probability of observing some observation given your true state. If you know the true state, you'd look to the probability function of the state (y), as shown. This would be represented by the observations of the attitudes in the previous example. This is denoted by:

$$P(y_t | x_t)$$

When we put all of these together, we get a probability distribution of all states and observations in our records. First, we have the probability of all the states from time 1 to time "t", and the probabilities of the observations factored into the HMM given what states we want to find, which is shown below.

$$p(X, Y) = p(x_1) \prod_{t=1}^{T-1} p(x_{t+1} | x_t) \prod_{t'=1}^{T} p(y_{t'} | x_{t'})$$

Figure 2: Overall distribution

Furthermore, the below figure shows the graphical representation of the above distribution into what we know as an Hidden Markov Model. On top is the Markov Chain, with an observation that is denoted from the state that we can't see and are trying to predict (shown as x). This is an HMM, because y represents the observations we **do** see, while x are the assumptions we can't observe, and these are told by y.

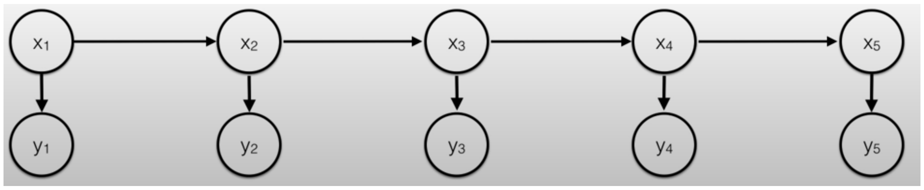

Figure 3: HMM representation

2.3 Worked Example With Probabilities of a Time Series

Lets use what we've just learn for a small time-series probability task. Say that you are given 6 observations of "confused, confused, grumpy, grumpy, happy, grumpy, grumpy? Now our task is to find: "What is the probability of and all happy time-series?" (just to simplify the calculations). Let's also assume we were given the following information:

State Transition Probability:

$$A = \begin{bmatrix} 0.8 & 0.15 & 0.05 \\ 0.38 & 0.6 & 0.02 \\ 0.75 & 0.05 & 0.02 \end{bmatrix}$$

Initial State Distribution: $\pi = (0.7, 0.25, 0.05)$

Emission Probabilities:

$$A = \begin{bmatrix} 0.6 & 0.3 & 0.1 \\ 0.05 & 0.3 & 0.65 \\ 0 & 0.5 & 0.5 \end{bmatrix}$$

Now, the general formula to find the probability of a specific time series would be (where O=Observations and Q=Hidden States):

P(O) = $\sum_{AllQ} P(O|Q)P(Q)$

To find probability of an "all-sun" time series: "$(0.3^2 * 0.1^4 * 0.6) * (0.7 * 0.8^6) + ...$"

Why is this? The first section is the probability of seeing that sequence if every day would be sunny (2 confused, 4 grumpy, 1 happy). The next section contains the probability that you start with happy, followed by all happy for the 6 observations. However, this is only one particular sequence for all happy, you'd need to know all particular sequences of all attitudes during all the days to find the true probability of that one sequence. This may seem like a silly calculation, but calculating HMM Chains like these are important to generalizing probabilities/patterns for future time-series.

3 Hidden State Interference

Our task is to infer hidden state probabilities through observing y and predicting x, simply put as the probability of x **given** y. Specifically, we care about the probability at timestep t given the hidden states given **all** our observations: P(x$_t$|Y)0.2cm

Steps:

1. Compute "Alpha Quantities", the joint probability of the current state variable and our historical observations:

   $\alpha_t(x_t) = p(x_t, y_1, ..., y_t)$

2. Compute "Beta Quantities", the conditional probability of all future probabilities given our current states:

   $\beta B_t(x_t) = p(y_{t+1}, ..., y_t|x_t)$

3. We need these two, because if we multiply them, we get the joint probability of x at time t and past observations and the future observations given x at time t. This turns into the joint probability of x at time t, and all observations. We can normalize this, meaning that we can now find x (our hidden state) given y (our observation)!

$\alpha_t(x_t)\beta B_t(x_t) = p(x_t, y_1, ..., y_t)p(y_{t+1}, ..., y_t|x_t) = p(x_t|y)$

3.1 Forward-Backward Inference

In order to find the normalized expression at the end of the previous expression, one would have to compute forward inference (the alpha equation), as well as backward inference (the beta equation), and then fuse them together. Remember that our goal here is to to find the probability of finding our hidden state (x at time t) given y (our observation). After conducting forward-backward inference, our equation looks like this:

$$p(x_t|Y) = \frac{\alpha_t(x_t)\beta_t(x_t)}{\sum_{x_t'} \alpha_t(x_t')\beta_t(x_t')}$$

Figure 4: Equation after Forward-Backward Interference

Now, what would happen if we wanted to find the transitional probability of the Hidden Markov Model – that is, the probability of transitioning from one state to another (x at time t, to x at time t+1). Using a similar process of fusing alpha and beta equations, we get this expression for the probability:

$$p(x_t, x_{t+1}|Y) = \frac{\alpha_t(x_t)p(x_{t+1}|x_t)\beta_{t+1}(x_{t+1})p(y_{t+1}|x_{t+1})}{\sum_{x_T} \alpha_T(x_T)}$$

Figure 5: Equation for Transitional Probability

This can cause the joint probability can be really tiny because of so many variables, causing underflow. To avoid this you would have to re-normalize at each step.

4 Decoding

4.1 The Viterbi Algorithm with NLP Speech Tagging

Suppose that we were asked to find the best sequence of hidden states Q given the observation sequence O. For example, in our aforementioned case, if our friend's attitudes were: happy, happy, grumpy, confused, what would the weather be like during that time period? This is called the **decoding** task. The Viterbi algorithm is a common decoding algorithm for Hidden Markov Models. It is a dynamic programming like concept. The structure is best shown by the figure below:

Starting off, the program processes the observations left to right, and each cell represents the the probability of an HMM being in a certain state after

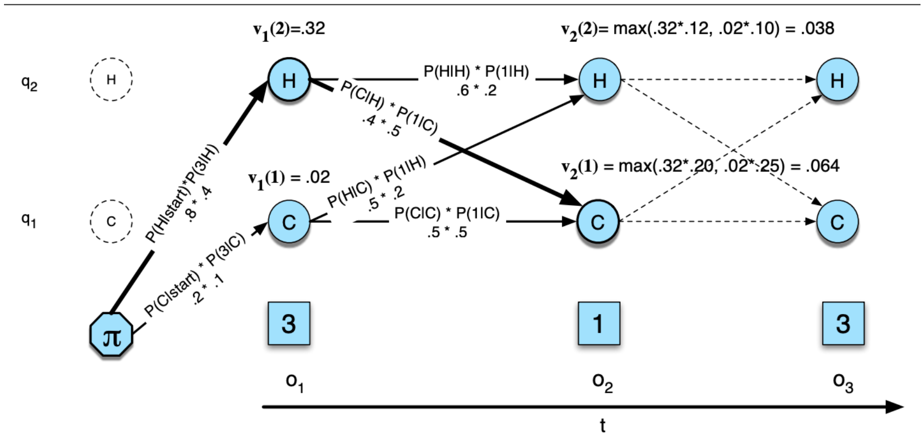

Figure 6: The Viterbi Algorithm Structure

seeing t amount of observations, passing through the state sequence. Each cell is then recursively computed to find the most profitable path to a cell. Generally speaking, you want to find the path of hidden states that gives you the highest probability of that specific sequence. In statistical terms, if Q is our sequence of states from $q_1, ..., q_t$

, we'd want to find the path of states which maximizes the probability of a single state given our probabilities: $P(Q—O_1, ..., O_t)$

This will be shown further in the following demo with NLP speech tagging.

References

Here are the sources used for my lecture, I would recommend checking them out for more information.

1. https://web.stanford.edu/ jurafsky/slp3/A.pdf
2. https://medium.com/@postsanjay/hidden-markov-models-simplified-c3f58728caab
3. http://www.shokhirev.com/nikolai/abc/alg/hmm/hmm.html
4. https://www.youtube.com/watch?v=9yl4XGp5OEg