

Hidden Markov model

Krishna Agrawal, 19111031

September 15, 2021

Introduction

Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process – call it X – with unobservable ("hidden") states. HMM assumes that there is another process Y whose behavior "depends" on X . The goal is to learn about X by observing Y .

Hidden Markov models are known for their applications to thermodynamics, statistical mechanics, physics, chemistry, economics, finance, signal processing, information theory, pattern recognition - such as speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges and bioinformatics.

Defination

Let X_n and Y_n be discrete-time stochastic processes and $n \geq 1$. The pair (X_n, Y_n) is a hidden Markov model if

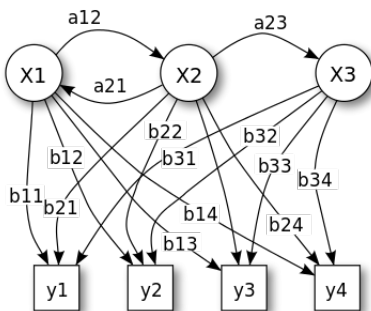
- X_n is a Markov process whose behavior is not directly observable ("hidden");
- $P(Y_n \in A | X_1 = x_1, \dots, X_n = x_n) = P(Y_n \in A | X_n = x_n)$, for every $n \geq 1, x_1, \dots, x_n$ and an arbitrary (measurable) set A .

The states of the process X_n are called hidden states, and $P(Y_n \in A | X_n = x_n)$ is called emission probability or output probability.

History

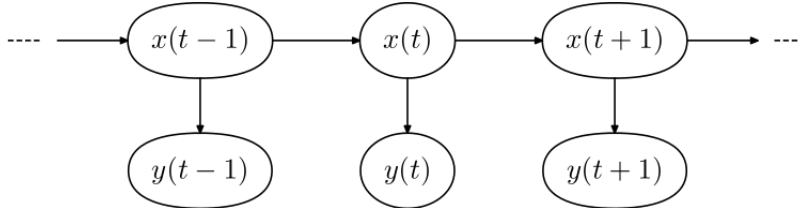
Hidden Markov Models were described in a series of statistical papers by Leonard E. Baum and other authors in the second half of the 1960s. One of the first applications of HMMs was speech recognition, starting in the mid-1970s. In the second half of the 1980s, HMMs began to be applied to the analysis of biological sequences, in particular DNA. Since then, they have become ubiquitous in the field of bioinformatics.

Structural Architecture



The random variable $x(t)$ is the hidden state at time t , $x(t) \in (x_1, x_2, x_3)$. The random variable $y(t)$ is the observation at time t with $y(t) \in (y_1, y_2, y_3, y_4)$. The arrows in the diagram (often called a trellis diagram) denote conditional dependencies.

From the diagram, it is clear that the conditional probability distribution of the hidden variable $x(t)$ at time t , given the values of the hidden variable x at all times, depends only on the value of the hidden variable $x(t-1)$; the values at time $t-2$ and before have no influence. This is called the Markov property. Similarly, the value of the observed variable $y(t)$ only depends on the value of the hidden variable $x(t)$ (both at time t).



The diagram below shows the general architecture of an instantiated HMM. Each oval shape represents a random variable that can adopt any of a number of values.

In the standard type of hidden Markov model considered here, the state space of the hidden variables is discrete, while the observations themselves can either be discrete (typically generated from a categorical distribution) or continuous (typically from a Gaussian distribution). The parameters of a hidden Markov model are of two types, **transition probabilities** and **emission probabilities** (also known as output probabilities). The transition probabilities control the way the hidden state at time t is chosen given the hidden state at time $t-1$.

The hidden state space is assumed to consist of one of N possible values, modelled as a categorical distribution. This means that for each of the N possible states that a hidden variable at time t can be in, there is a transition probability from this state to each of the N possible states of the hidden variable at time $t+1$, for a total of N^2 transition probabilities.

In addition, for each of the N possible states, there is a set of emission probabilities governing the distribution of the observed variable at a particular time given the state of the hidden variable at that time. The size of this set depends on the nature of the observed variable. For example, if the observed variable is discrete with M possible values, governed by a categorical distribution, there will be $M-1$ separate parameters, for a total of $N(M-1)$ emission parameters over all hidden states.

On the other hand, if the observed variable is an M -dimensional vector distributed according to an arbitrary multivariate Gaussian distribution.

Applications

HMMs can be applied in many fields where the goal is to recover a data sequence that is not immediately observable (but other data that depend on the sequence are).

Applications include: Computational finance, Single-molecule kinetic analysis, Speech recognition, Speech synthesis, Document separation in scanning solutions, Machine translation, Gene prediction, Time series analysis, Handwriting recognition, Activity recognition, Sequence classification, Transportation forecasting, Solar irradiance variability.

Conclusion

The parameter learning task in HMMs is to find, given an output sequence or a set of such sequences, the best set of state transition and emission probabilities. The task is usually to derive the maximum likelihood estimate of the parameters of the HMM given the set of output sequences. No tractable algorithm is known for solving this problem exactly, but a local maximum likelihood can be derived efficiently using the Baum–Welch algorithm or the Baldi–Chauvin algorithm. The Baum–Welch algorithm is a special case of the expectation-maximization algorithm.