

Hidden Markov Models

Aman Khandwe, 19111006

September 15, 2021

Summary

The Hidden Markov Model (HMM) is a statistical Markov model in which the system being represented is considered to be a Markov process with unobservable ("hidden") states \mathbf{X} . HMM implies that another process, \mathbf{Y} , has behaviour that "depends" on \mathbf{X} . The objective is to learn about \mathbf{X} by observation of \mathbf{Y} . HMM requires that, for each time occurrence n , the conditional probability distribution of Y_n given the history $X_n = x_{n \leq n}$ must not depend on $X_n = x_{n < n}$. Hidden Markov models have been used in thermodynamics, statistical mechanics, physics, chemistry, economics, finance, signal processing, information theory, pattern recognition - including speech, handwriting, gesture recognition, part-of-speech tagging, musical score following, partial discharges, and bioinformatics. 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");

$$\mathbf{P}(Y_n \in A \mid X_1 = x_1, \dots, X_n = x_n) = \mathbf{P}(Y_n \in A \mid X_n = x_n),$$

for every $n \geq 1, x_1, \dots, x_n, x_1, \dots, x_n$, and an arbitrary (measurable) set A . Here the state of process X_n is known as hidden states and

$$\mathbf{P}(Y_n \in A \mid X_n = x_n) \mathbf{P}(Y_n \in A \mid X_n = x_n)$$

is called emission probability or output probability. The learning objective in HMMs is to identify the best set of state transition and emission probabilities given the output sequence or a group of such sequences. Given a set of output sequences, the objective is generally to calculate the highest likelihood estimate of the HMM's parameters. There is no tractable method for solving this issue exactly, but a local maximum likelihood may be rapidly obtained using the Baum–Welch or the Baldi–Chauvin algorithms. The

Baum–Welch algorithm is an example of an expectation-maximization algorithm. When using HMMs for time series prediction, more complex Bayesian inference methods, such as Markov chain Monte Carlo (MCMC) sampling, have been shown to be superior to find a single maximum likelihood model in terms of accuracy and stability because MCMC imposes a considerable computing load, in instances where computational scalability is also desired. Hidden Markov models are generative models that describe both the prior distribution of hidden states (the transition probabilities) and the conditional distribution of observations given states (the emission probabilities). A Dirichlet process is used instead of a Dirichlet distribution as an extension of the previously described hidden Markov models with Dirichlet priors. This model has an unknown and potentially unlimited number of states. A two-level Dirichlet process, comparable to the previously described model with two layers of Dirichlet distributions, is commonly used. A hierarchical Dirichlet process hidden Markov model, or HDP-HMM for short, is one such model. It was first characterised as the "Infinite Hidden Markov Model". A different form of extension substitutes a discriminative model for the generative model included in ordinary HMMs. Rather of modelling the joint distribution, this sort of model explicitly models the conditional distribution of the hidden states given the observations. The maximum entropy Markov model (MEMM), which represents the conditional distribution of states using logistic regression, is an example of this model (also known as a "maximum entropy model"). The benefit of this sort of model is that it may model any characteristics (i.e. functions) of the data, allowing domain-specific knowledge of the problem at hand to be fed into the model. The linear-chain conditional random field is a variation of the discriminative model. Instead of the directed graphical models seen in MEMMs and comparable models, this one employs an undirected graphical model (aka Markov random field). The benefit of this sort of model is that it does not suffer from MEMM's so-called label bias problem, and hence may produce more accurate predictions. The drawback is that training may take longer than with MEMMs. The factorial hidden Markov model, rather than a single Markov chain, allows a single observation to be conditioned on the corresponding hidden variables of a collection of K independent Markov chains. It is equal to a single HMM with N^K states (assuming N states for each chain), and therefore learning in such a model is difficult: given a sequence of length T a simple Viterbi method has complexity $O(N^{2K} T)$. A junction tree technique may be used to discover an accurate answer, but it has a $O(N^{K+1} K T)$ complexity. Approximate techniques, such as variational approaches can be used in it.