

Hidden markov model

Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process.

Learning :

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. If the HMMs are used for time series prediction, more sophisticated Bayesian inference methods, like Markov chain Monte Carlo (MCMC) sampling are proven to be favorable over finding a single maximum likelihood model both in terms of accuracy and stability. Since MCMC imposes significant computational burden, in cases where computational scalability is also of interest, one may alternatively resort to variational approximations to Bayesian inference. Indeed, approximate variational inference offers computational efficiency comparable to expectation-maximization, while yielding an accuracy profile only slightly inferior to exact MCMC-type Bayesian inference.

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.

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 , cryptanalysis,

Speech recognition , speech synthesis , parts of speech tagging , machine translation, chromatin state discovery , transportation forecast.

Inference :

Several inference problems are associated with hidden Markov models, as outlined below .

Probability of latent variables :

Filtering :

The task is to compute, given the model's parameters and a sequence of observations, the distribution over hidden states of the last latent variable at the end of the sequence . This task is normally used when the sequence of latent variables is thought of as the underlying states that a process moves through at a sequence of points of time, with corresponding observations at each point in time. Then, it is natural to ask about the state of the process at the end.

This problem can be handled efficiently using the forward algorithm .

Most likely explanation :

The task, unlike the previous two, asks about the joint probability of the entire sequence of hidden states that generated a particular sequence of observations (see illustration on the right). This task is generally applicable when HMM's are applied to different sorts of problems from those for which the tasks of filtering and smoothing are applicable. An example is part-of-speech tagging, where the hidden states represent the underlying parts of speech corresponding to an observed sequence of words. In this case, what is of interest is the entire sequence of parts of speech, rather than simply the part of speech for a single word, as filtering or smoothing would compute. This task requires finding a maximum over all possible state sequences, and can be solved efficiently by the Viterbi algorithm.

Statistical significance :

For some of the above problems, it may also be interesting to ask about statistical significance. What is the probability that a sequence drawn from some null distribution will have an HMM probability (in the case of the forward algorithm) or a maximum state sequence probability (in the case of the Viterbi algorithm) at least as large as that of a particular output sequence? When an HMM is used to evaluate the relevance of a hypothesis for a particular output sequence, the statistical significance indicates the false positive rate associated with failing to reject the hypothesis for the output sequence.