Parameter Estimation of Hidden Markov Models

Relating Observables to Hidden States Through the Segmental K-Means Algorithm

Noah Brandt Koesgaard, Emilie Elizabeth Milan Nielsen, Ludvig Marcussen and Michelle Rix

March 2023

Abstract

Not all Markov processes present Markovian features as observables. It is then desirable to infer parameter values of the original process from some observed sequence. In this proceeding we discuss a method for estimating the parameters of a Hidden Markov Model (HMM) from an observed sequence. The article in focus approximates the Baum-Welch approach of estimating the parameters, yielding a more efficient algorithm (the Segmented K-Means Algorithm). This is done considering only the most likely state sequence in the likelihood instead of summing over all possible sequences.

Introduction

A quintessential task in science concerns the elucidation of the relationship between observables of natural phenomena and the underlying mechanisms that produce these observables. A commonly used model to describe real world stochastic phenomena is Markov Chains in which states randomly evolve over time with some probability, subject to the constraint that the probability of going into some other state in the next time step only depends on the current state and not previous states. The simplicity and generality of this model makes it suitable to describe natural processes such as mutations on a DNA string and Brownian motion [3]. However, not all processes governed by Markov Chains necessarily exhibit Markovian observables, as the process of the mechanism may not be the same as the process producing the observables [1]. In such cases the challenge is to infer the structure of the original, or "hidden", Markov process from the observed process. This is the principal task of *Hidden Markov Models* (HMM). In particular, it is the aim of this write up to convey the application of HMM to infer the parameters of the hidden states that constitute the underlying Markov Chain from the observed non-Markov process using an algorithm described by Juang and Rabiner [2].

Definition of Hidden Markov Models

In this proceeding we consider an HMM relating an N-state first order Markov chain with states $S = [s_0, s_1...s_N]$ with a set of n observable states $O = [o_0, o_1...o_n]$ through some set of probability distributions $B = [b_{s0}, b_{s1}...b_{sN}]$. The Markov chain is defined from two parameters: The transition matrix and the initial state probability vector. The transition matrix is as usual defined as an N by N matrix, A, with entrances $a_{i,j}$ that denote the probability of going from the state I to state j within the Markov chain, while the initial state probability vector is defined as the probability of starting in a given state $\boldsymbol{\pi} = [\pi_{s0}, \pi_{s1}...\pi_{sN}]$. The entire HMM may then be completely specified from the Markov parameters and the probability distributions relating the Markov states with the observed states. The entire HMM is then denoted $\lambda = (A, \boldsymbol{\pi}, B)$. The relation between the different parameters is visualized in Figure 1.

Method for Inferring Parameters in HMMs

A natural question to be answered concerns the inference of parameters in the HMM given some observed sequence $x = [x_0, x_1...x_T]$. This is in essence a conditioning problem. If we consider the case in which the hidden states are equal to the observed ones the probability of observing a sequence s is given by the conditional probability in Equation 1.

$$P(s|A, \pi) = \pi_{s_0} \prod_{t=1}^{T} a_{(s_{(t-1)}, s_t)} \tag{1}$$

Since the observed states are dependent on a probability distribution relating the hidden states to the observed states, the probability of observing the sequence x given a HMM λ is given by Equation 2.

$$f(x|\lambda) = \sum_{s} \pi_{s_0} \prod_{t=1}^{T} a_{(s_{(t-1)}, s_t)} b_{s_t}(x_t) \tag{2}$$

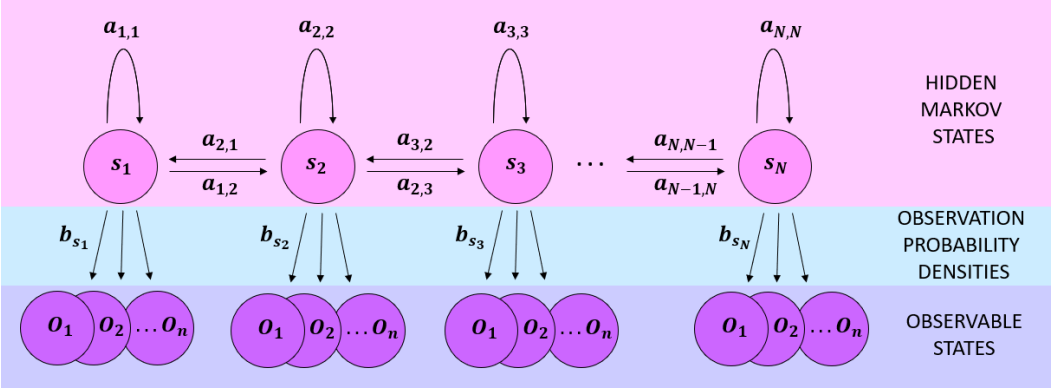

Figure 1: Depiction of an extended Markov process called a doubly stochastic HMM. Emitting observed states, makes it more flexible and capable of capturing complex patterns in the data.

The goal is then to maximize this likelihood over all given parameters in the model to determine the most likely model. This may be done using the Baum-Welch algorithm (sometimes called the forward-backward algorithm) [2]. However, this optimization relies on summing over all possible states which becomes computationally expensive and may encounter numerical issues as well. Instead, the likelihood that is maximized is given in Equation 3.

$$\max_s[f(x,s|\lambda)] = \max_s[\pi_{s_0}\prod_{t=1}^{T} a_{(s_{(t-1)},s_t)}b_{s_t}(x_t)] \tag{3}$$

That is, rather than summing over all possible state sequences only the most likely state sequence is considered. This likelihood is maximized using the segmental K-means algorithm. The algorithm consist of two steps: Segmentation and optimization. In the first step, the likelihood in Equation 3 is found for some initial model λ using the Viterbi algorithm [2]. In the second step the model parameters are varied so as to maximize this likelihood and these parameters are extracted as seen in Equation 4.

$$\overline{\lambda} = \arg\max_{\lambda}\{\max_{s}[f(x,s|\lambda)]\} \tag{4}$$

These new parameters are now used in step one again. This processes is repeated until the likelihood converges within a desired threshold. It can be shown mathematically that the algorithm is guaranteed to converge, however not necessarily to the correct "hidden" values [2]. While the algorithm requires less computational power, it does come with a caveat: We are not considering all possible paths in the hidden state sequence, but only the most probable. This may cause trouble in the event of multiple probable paths. Therefore, the Baum-Welch algorithm is expected to perform better in general since every paths is weighted with some probability.

Conclusion

HMM allows us to statistically model a system assumed to be Markovian, but with unobservable states. Assuming the structure of the HMM is known, the problem lies in estimating the parameters of the HMM model. In this write up, the method of the K-means algorithm has been discussed. The K-means is a method for faster estimation of the HHM parameters, than that of the Baum-Welch method. However, it results in a less precise algorithm, since not all states are saved in the algorithm, and only the most likely state-sequence is taken into account.

References

[1] Mariette Awad and Rahul Khanna. "Hidden Markov Model". In: 2015, pp. 81–104. DOI: 10.1007/978-1-4302-5990-9_5.

[2] B.-H. Juang and L.R. Rabiner. "The segmental K-means algorithm for estimating parameters of hidden Markov models". In: *IEEE Transactions on Acoustics, Speech, and Signal Processing* 38.9 (1990), pp. 1639–1641. ISSN: 0096-3518. DOI: 10.1109/29.60082.

[3] Carolin Kosiol and Nick Goldman. "Markovian and Non-Markovian Protein Sequence Evolution: Aggregated Markov Process Models". In: *Journal of Molecular Biology* 411.4 (2011), pp. 910–923. ISSN: 0022-2836. DOI: https://doi.org/10.1016/j.jmb.2011.06.005. URL: https://www.sciencedirect.com/science/article/pii/S0022283611006309.