Hidden Markov Models

See also Durban, Eddy, Krogh, Mitchison

Definition

A Hidden Markov Model is a Markov Chain over an ‘unobservable’ state space, S, together with a set of emission probabilities, defined on the symbol space B, and indexed by S.

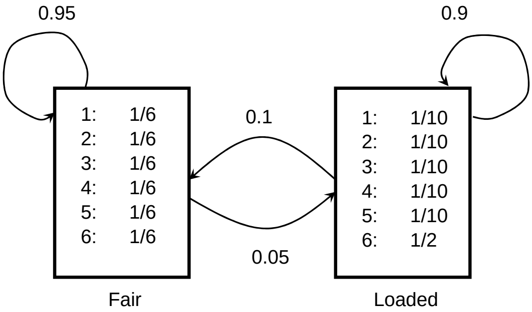
Consider a path ρ through S, resulting in the observed symbol sequence x_i.

Define $a_{kl} = P(\rho_i = l|\rho_{i-1} = k)$ and $e_k(b) = P(x_i = b|\rho_i = k)$

Example: Occasionally dishonest casino

Casino which uses a fair die most of the time but occasionally switches to a loaded die.

Model:



Viterbi algorithm

Calculates the most probable path through hidden states Objective:

$$\boldsymbol{\rho}^* = \underset{\boldsymbol{\rho}}{\operatorname{argmax}}\ P(x, \boldsymbol{\rho})$$

Define $v_m(i)$, the probability of the most probable path ending in state m at observation i ($v_m(i) = \max\limits_{\{\boldsymbol{\rho}|\rho_i=m\}} P(x_1, \ldots, x_i, \boldsymbol{\rho})$)

then we can calculate $v_m(i+1) = e_m(x_{i+1}) \max\limits_{k} (v_k(i)a_{km})$

Viterbi Algorithm

Initialisation: $v_{\mathcal{B}}(0) = 1, v_k(0) = 0\ \forall\ k \neq \mathcal{B}$

Recursion : For $(i = 1 \ldots L)$ $\quad v_m(i) = e_m(x_i) \max_k(v_k(i-1)a_{km};$

$\qquad \mathrm{ptr}_i(m) = \operatorname{argmax}_k(v_k(i-1)a_{km})$

Termination : $\quad P(x, \boldsymbol{\rho}^*) = \max_k(v_k(L)a_{k\mathcal{E}});$

$\qquad \rho_L^* = \operatorname{argmax}_k(v_k(L)a_{k\mathcal{E}})$

Traceback $(i = L \ldots 1)$: $\quad \rho_{i-1}^* = \mathrm{ptr}_i(\rho_i^*)$

Example

Given the following sequence of observations from the occasionally dishonest casino in the earlier slide, what is the most probably sequence of hidden states (i.e. the most probably sequence of fair/loaded dice).

$$4, 4, 5, 2, 3, 4, 1, 3, 6, 5, 4, 5, 6, 6, 4, 5, 3, 2, 6, 6, 1, 6, 5, 4$$

Forward algorithm

Calculates the probability of a sequence of observations, by summing over all possible paths through the hidden states - $P(x) = \sum_{\rho} P(x, \rho)$

The algorithm is similar to Viterbi but with maximisations replaced by sums:

Define $f_m(i)$, the joint probability of the sequence of observations up to position i and that the state at i is m

Forward Algorithm

Initialisation ($i = 0$):	$f_{\mathcal{B}}(0) = 1, f_k(0) = 0$ for $k \neq \mathcal{B}$
Recursion ($i = 1 \ldots L$):	$f_l(i) = e_l(x_i) \sum_k f_k(i-1) a_{kl}$
Termination:	$P(x) = \sum_k f_k(L) a_{k\mathcal{E}}$

Example

Calculate the probability of generating the following sequence of observations in the occasionally dishonest casino.

$$4, 4, 5, 2, 3, 4, 1, 3, 6, 5, 4, 5, 6, 6, 4, 5, 3, 2, 6, 6, 1, 6, 5, 4$$

Backward algorithm

Can also do this in the reverse direction

Define $b_m(i) = P(x_{i+1}, \ldots, x_L | \rho_i = m)$

Backward Algorithm

Initialisation ($i = L$):	$b_k(L) = a_{k\mathcal{E}}$ for all k
Recursion ($i = L-1, \ldots, 1$):	$b_m(i) = \sum_k a_{mk} e_k(x_{i+1}) b_k(i+1)$
Termination:	$P(x) = \sum_k a_{\mathcal{B}k} e_k(x_1) b_k(1)$

Forward-Backward procedure

Calculate the posterior probability that the hidden state at position i is m - i.e. $P(\rho_i = m|\mathbf{x})$

$$
\begin{aligned}
P(x, \rho_i = m) \quad &= \quad P(x_1, ...x_i, \rho_i = m)P(x_{i+1} \ldots, x_L|x_1 \ldots, x_i, \rho_i = m) \\
&= \quad P(x_1, ...x_i, \rho_i = m)P(x_{i+1} \ldots, x_L|\rho_i = m) \\
&\Longrightarrow \quad P(\rho_i = m|\mathbf{x}) = \frac{f_k(i)b_k(i)}{P(\mathbf{x})}
\end{aligned}
$$

Training a Hidden Markov Model when the state sequence is known

ML estimates of parameters can be calculated easily as

$$\hat{a}_{km} = \frac{A_{km}}{\sum_j A_{kj}} \text{ and } \hat{e}_k(b) = \frac{E_k(b)}{\sum_j E_k(j)}$$

Training a Hidden Markov Model - Baum-Welch training

Baum-Welch estimates HMM parameters iteratively.
Informal description: given an initial set of values for the HMM parameters, obtain new parameter estimates from the frequency with which transitions/emmissions are observed across the probably paths. Repeat.

Training a Hidden Markov Model - Baum-Welch training

Baum-Welch estimates HMM parameters iteratively. Given a set of values θ for the HMM parameters

$$P(\rho_i = k, \rho_{i+1} = m|\mathbf{x}, \theta) = \frac{f_k(i)a_{km}e_m(x_{i+1})b_m(i+1)}{P(x)} \ldots \text{Prove}$$

We can derive estimates of A_{kl} by summing over all positions of the sequence.

$E_k(b)$ can be estimated from

$$\sum_{\{i|x_i=b\}} \frac{f_k(i)b_k(i)}{P(x)}$$

Semi-Markov models

In a markov chain the time spent in any state is described by a geometric (or exponential in continuous-time case) distribution.

This can be generalized - e.g. semi-Markov HMM, each state emits a sequence of observations. Semi-Markov HMMs have been applied to the problem of finding genes in unannotated genomic DNA sequences (e.g. Burge & Karlin, *Journal of Molecular Biology*, 1997).