Bioinformatics
Lecture : HMM
Timo Koski

TK

02.08.2018

KTH Matematik

Contents

This lecture corresponds the sections 11.1−11.3.2 in Ewens and Grant

1) Definition, gene prediction, protein modelling, three problems of HMM.
2) Properties implied by conditional independence
3) Forward-Backward Algorithm and the Scoring/Evaluation/Parsing Problem
4) Alignment (Viterbi algorithm), Learning (Baum - Welch algorith)
5) Learning (Baum - Welch algorith)

KTH Matematik

The three problems of HMM 1

[1]**The Evaluation or Scoring Problem** Compute $P\left(Y_0 = o_0, \ldots, Y_n = o_n; \lambda\right)$. Since the margnalization involves J^{n+1} possible sequences, the total computational requirements are of the order $2(n+1) \cdot J^{n+1}$ operations. The solution is known as the *forward-backward procedure*

KTH Matematik

The three problems of HMM 2

[2] **The Decoding or Alignment Problem** Find the most probable state sequence that led to the observed sequence $(o_0 \ldots o_n)$. This is an *alignment problem*. Find the sequence $j_0^* \ldots j_n^*$ that maximizes

$$P\left(X_0 = j_0, \ldots, X_n = j_n, Y_0 = o_0, \ldots, Y_n = o_n; \lambda\right)$$

for a fixed observed sequence $o_0 \ldots o_n$ (*Viterbi algorithm*).

KTH Matematik

The three problems of HMM 3

[3] **The Learning or Training Problem** Given an observed sequence $\mathbf{o} = o_0 \ldots o_n$, find the 'right' model parameter values

$$\lambda = (A, B, \pi(0))$$

in a fixed topology that specify a model most likely to generate the given sequence

KTH Matematik

HIDDEN MARKOV MODELS: Conditional Independence

Markov property and the conditional independence property **III** imply useful expressions for smoothing, prediction, filtering and evaluation, and these yield the solutions to the three problems stated above.

KTH Matematik

Notations

Here the typesetting is simplified e.g. by writing a conditional probability as

$$P\left(Y_m = o_m, \ldots, Y_N = o_N \mid X_n = j_n \ldots, X_N = j_N\right)$$

simply as

$$P\left(Y_m, \ldots, Y_N \mid X_n, \ldots, X_N\right)$$

$$P\left(Y_0, \ldots, Y_n \mid X_0, \ldots, X_n\right) = \prod_{i=0}^{n} P\left(Y_i \mid X_i\right).$$

KTH Matematik

Proposition 1

Proposition

For all integers n and m such that $0 \leq n \leq m \leq N$

$$P\left(Y_m, \ldots, Y_N \mid X_n \ldots, X_N\right) = P\left(Y_m, \ldots, Y_N \mid X_m, \ldots, X_N\right).$$

KTH Matematik

Proposition 1.

Proof: The left hand side of the asserted identity can be expressed as

$$\frac{1}{P\left(X_n,\ldots,X_N\right)}\sum P\left(Y_m,\ldots,Y_N \mid X_0\ldots,X_N\right)\cdot P\left(X_0\ldots,X_N\right),$$

where the summation is over $j_0,\ldots,j_{n-1}$ (i.e. the values of $X_{j_0},\ldots,X_{j_{n-1}}$). If $n=0$, there is no summation. By conditional independence (and a marginalization argument)

$$P\left(Y_m,\ldots,Y_N \mid X_0\ldots,X_N\right)=P\left(Y_m|X_m\right)\cdot\ldots\cdot P\left(Y_N|X_N\right).$$

This can be taken outside the summation sign $\sum$, since $m \geq n$.

Proof

Then we are dealing with

$$\prod_{l=m}^{N} P\left(Y_l|X_l\right) \frac{1}{P\left(X_n,\ldots,X_N\right)} \sum P\left(X_0\ldots,X_N\right),$$

where the sum equals $P\left(X_n,\ldots,X_N\right)$, since we are summing over $j_0,\ldots,j_{n-1}$. Thus the whole last expression equals

$$= \prod_{l=m}^{N} P\left(Y_l|X_l\right),$$

which is independent of n. Since the right hand side of the above is a special case of the left hand side for $n = m$, this proves the assertion as claimed. □

Proposition 2.

Proposition

For all integers $n = 0, \ldots, N-1$

$$P\left(Y_{n+1}, \ldots, Y_N \mid X_0, \ldots, X_n\right) =$$

$$P\left(Y_{n+1}, \ldots, Y_N \mid X_n\right).$$

KTH Matematik

Proof of proposition 2.

Set

$$X^{(t)} = (X_0, \ldots, X_t), \; Y^{(t)} = (Y_0, \ldots, Y_t),$$

Proof: The left hand side is

$$\frac{1}{P\left(X^{(n)}\right)} \sum P\left(X^{(N)}\right) \cdot P\left(Y_{n+1}, \ldots, Y_N \mid X^{(N)}\right),$$

where the summation is over $j_{n+1}, \ldots, j_N$. By the first proposition, (with $m = n+1$, $n = 0$), we have

$$P\left(Y_{n+1}, \ldots, Y_N \mid X^{(N)}\right) = P\left(Y_{n+1}, \ldots, Y_N \mid X_{n+1}, \ldots, X_N\right)$$

and using the same proposition and equation once more (with $m = n+1$) we have

$$P\left(Y_{n+1}, \ldots, Y_N \mid X_{n+1}, \ldots, X_N\right) =$$

$$P\left(Y_{n+1}, \ldots, Y_N \mid X_n, \ldots, X_N\right).$$

KTH Matematik

Proof of proposition2.

Thus

$$\sum P\left(X^{(N)}\right)\cdot P\left(Y_{n+1},\ldots,Y_N \mid X^N\right) =$$

$$\sum P\left(X^N\right)\cdot P\left(Y_{n+1},\ldots,Y_N \mid X_n,\ldots,X_N\right).$$

By conditional probability $P\left(X^{(N)}\right) =$

$$P\left(X_{n+1},\ldots,X_N|X^{(n)}\right)\cdot P\left(X^{(n)}\right)$$

and by a consequence of *Markov property* we have

$$P\left(X_{n+1},\ldots,X_N|X^{(n)}\right) = P\left(X_{n+1},\ldots,X_N|X_n\right).$$

KTH Matematik

Proof of proposition2.

Thus the sum equals, since we are summing over $j_{n+1}, \ldots, j_N$,

$$\sum P\left(X^{(N)}\right) \cdot P\left(Y_{n+1}, \ldots, Y_N \mid X_n, \ldots, X_N\right) =$$

$$= \sum \frac{P\left(Y_{n+1}, \ldots, Y_N, X_n, \ldots, X_N\right) \cdot P\left(X^{(n)}\right)}{P\left(X_n\right)}.$$

KTH Matematik

Proof of proposition 2.

$$= P\left(X^{(n)}\right)\sum \frac{P\left(Y_{n+1},\ldots,Y_N,X_n,\ldots,X_N\right)}{P\left(X_n\right)}.$$

And as we are summing over $j_{n+1},\ldots,j_N$, we have here that

$$\sum \frac{P\left(Y_{n+1},\ldots,Y_N,X_n,\ldots,X_N\right)}{P\left(X_n\right)} = \frac{P\left(Y_{n+1},\ldots,Y_N,X_n\right)}{P\left(X_n\right)} =$$

KTH Matematik

Proof of proposition 2.

$$= P\left(Y_{n+1},\ldots,Y_N|X_n\right).$$

We have that

$$\frac{1}{P\left(X^{(n)}\right)}\sum P\left(X^{(N)}\right)\cdot P\left(Y_{n+1},\ldots,Y_N \mid X^{(N)}\right) =$$

$$\frac{1}{P\left(X^{(n)}\right)} P\left(Y_{n+1},\ldots,Y_N|X_n\right)\cdot P\left(X^{(n)}\right),$$

which proves the assertion as claimed. □

KTH Matematik

Proposition 3.

Proposition

For all integers $n = 0, \ldots, N$

$$P\left(Y_0, \ldots, Y_n \mid X_0, \ldots, X_N\right) =$$

$$P\left(Y_0, \ldots, Y_n \mid X_0, \ldots, X_n\right).$$

KTH Matematik

Proposition 4

Proposition

For all integers $n = 0, \ldots, N$

$$P\left(Y_0, Y_1, \ldots, Y_N \mid X_n\right) =$$

$$P\left(Y_0, Y_1, \ldots, Y_n \mid X_n\right) \cdot P\left(Y_{n+1}, \ldots, Y_N \mid X_n\right).$$

KTH Matematik

Backward variable

The conditional probability $P(Y_{n+1}, \ldots, Y_N \mid X_n)$ is called the **backward variable**. The next proposition is used to find a recursion for this backward variable.

A Proposition for the Backward variable

Proposition

For all integers $n = 0, \ldots, N$

$$P\left(Y_n, Y_{n+1}, \ldots, Y_N \mid X_n\right) =$$

$$P\left(Y_n \mid X_n\right) \cdot P\left(Y_{n+1}, \ldots, Y_N \mid X_n\right).$$

KTH VETENSKAP OCH KONST
KTH Matematik

One More Proposition

Proposition

For all integers $n = 0, \ldots, N-1$

$$P\left(Y_0, Y_1, \ldots, Y_N \mid X_n, X_{n+1}\right) =$$

$$P\left(Y_0, Y_1, \ldots Y_n \mid X_n\right) \cdot P\left(Y_{n+1}, \ldots, Y_N \mid X_{n+1}\right).$$

The problem is to compute the simultaneous probability for the a sequence of emitted symbols, $\mathbf{o} = o_0 \ldots o_N$, conditioned on some model $\lambda = (A, B, \pi(0))$,

$$L_N = P\left(Y_0 = o_0 \ldots, Y_n = o_N; \lambda\right) =$$

$$\sum_{j_o=1}^{J} \cdots \sum_{j_N=1}^{J} \pi_{j_0}(0)\, b_{j_0}(0) \prod_{l=1}^{N} a_{j_{l-1}|j_l} b_{j_l}(l),$$

so that the exponential growth of operations in N involved in the marginalization is avoided. In order to simplify the notation, the reference to the model λ is omitted.

Forward Algorithm

Let

$$P\left(Y_0 = o_0, \ldots, Y_N = o_N, X_n = j\right) =$$

$$P\left(X_n = j\right) \cdot P\left(Y_0 = o_0, \ldots, Y_N = o_N | X_n = j\right).$$

But the right hand side is factorized as

$$P\left(Y_0 = o_0, \ldots, Y_N = o_N | X_n = j\right) =$$

$$P\left(Y_0 = o_0, \ldots, Y_n = o_n | X_n = j\right) \cdot P\left(Y_{n+1} = o_{n+1}, \ldots, Y_N = o_N | X_n = j\right)$$

KTH Matematik

Forward Algorithm

This gives

$$P\left(Y_0 = o_0, \ldots, Y_n = o_n, X_n = j\right) \cdot P\left(Y_{n+1} = o_{n+1}, \ldots, Y_N = o_N | X_n = j\right).$$

Since

$$P\left(Y_0 = o_0, \ldots, Y_N = o_N\right)$$

$$= \sum_{j=1}^{J} P\left(Y_0 = o_0, \ldots, Y_N = o_N, X_n = j\right),$$

we get

$$P\left(Y_0 = o_0, \ldots, Y_N = o_N\right) = \sum_{j=1}^{J} \alpha_n(j) \cdot \beta_n(j),$$

Forward Algorithm

$$P\left(Y_0 = o_0, \ldots, Y_N = o_N\right) = \sum_{j=1}^{J} \alpha_n(j) \cdot \beta_n(j),$$

where

$$\alpha_n(j) = P\left(Y_0 = o_0, \ldots, Y_n = o_n, X_n = j\right).$$

$$\beta_n(j) = P\left(Y_{n+1} = o_{n+1}, \ldots, Y_N = o_N | X_n = j\right).$$

We take $\beta_N(j) = 1$ for every j arbitrarily.

KTH Matematik

Forward Algorithm

First

$$\alpha_{n+1}(j) = P\left(Y_0 = o_0, \ldots, Y_{n+1} = o_{n+1}, X_{n+1} = j\right) =$$

$$= \sum_{i=1}^{J} P\left(Y_0 = o_0, \ldots, Y_{n+1} = o_{n+1}, X_n = i, X_{n+1} = j\right)$$

$$= \sum_{i=1}^{J} P\left(X_n = i, X_{n+1} = j\right) \cdot P\left(Y_0 = o_0, \ldots, Y_{n+1} = o_{n+1} | X_n = i, X_{n+1} = j\right).$$

KTH Matematik

Forward Algorithm

Here, by the properties derived ('One More Proposition')

$$= \sum_{i=1}^{J} P\left(X_n = i, X_{n+1} = j\right) \cdot P\left(Y_0 = o_0, \ldots, Y_n = o_n | X_n = i\right) \cdot P\left(Y_{n+1} = o_{n+1} | X_{n+1} = j\right).$$

$$P\left(X_n = i, X_{n+1} = j\right) \cdot P\left(Y_{n+1} = o_{n+1} | X_{n+1} = j\right) = a_{i|j} \cdot b_j\left(o_{n+1}\right) \cdot P\left(X_n = i\right).$$

Hence we have

$$\sum_{i=1}^{J} P\left(X_n = i, X_{n+1} = j\right) \cdot P\left(Y_0 = o_0, \ldots, Y_n = o_n | X_n = i\right) \cdot P\left(Y_{n+1} = o_{n+1} | X_{n+1} = j\right) =$$

$$= \sum_{i=1}^{J} P\left(Y_0 = o_0, \ldots, Y_n = o_n, X_n = i\right) \cdot a_{i|j} \cdot b_j\left(o_{n+1}\right).$$

KTH Matematik

Final step

This last expression is by our definition of the forward variable equal to

$$= \sum_{i=1}^{J} \alpha_n(i) \cdot a_{i|j} \cdot b_j\left(o_{n+1}\right) =$$

$$\left[\sum_{i=1}^{J} \alpha_n(i) \cdot a_{i|j}\right] \cdot b_j\left(o_{n+1}\right).$$

This completes the derivation of the forward algorithm. We summarize the result in a formal way.

KTH Matematik

The Forward Recursion

Consider the forward variable $\alpha_n(j)$ defined as

$$\alpha_n(j) = P\left(Y_0 = o_0, \ldots, Y_n = o_n, X_n = j\right),$$

which is the probability of the emitted subsequence $\mathbf{o} = o_0 \ldots o_n$ and of the hidden chain being in the state j at time n (given the model λ).

Start:

$$\alpha_0(j) = b_j\left(o_0\right)\pi_j(0),\ j = 1, \ldots, J.$$

Recursion:

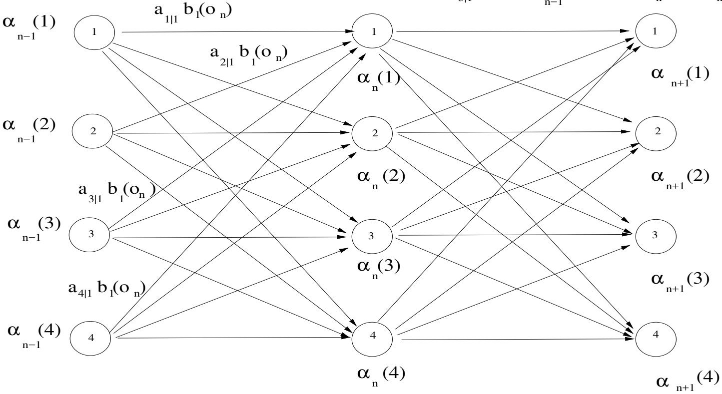
$$\alpha_{n+1}(j) = \left[\sum_{i=1}^{J} \alpha_n(i) \cdot a_{i|j}\right] \cdot b_j\left(o_{n+1}\right).$$

$j = 1, \ldots, J$, $1 \leq n \leq N-1$.

KTH Matematik

The Forward Recursion Trellis

$$\alpha_n(1) = a_{1|1} b_1(o_n)\, \alpha_{n-1}(1) + a_{2|1} b_n(o_1)\, \alpha_{n-1}(2) + a_{3|1} b_1(o_n)\, \alpha_{n-1}(3) + a_{4|1} b_n(o_1)\, \alpha_{n-1}(4)$$



Backward Algorithm

By definition

$$\beta_n(j) = \sum_{i=1}^{J} \frac{P\left(Y_{n+1} = o_{n+1}, \ldots, Y_N = o_N, X_n = j, X_{n+1} = i\right)}{P\left(X_n = j\right)}$$

$$= \sum_{i=1}^{J} \frac{P\left(Y_{n+1} = o_{n+1}, \ldots, Y_N = o_N | X_n = j, X_{n+1} = i\right) P\left(X_n = j, X_{n+1} = i\right)}{P\left(X_n = j\right)}.$$

Here $\frac{P(X_n=j,X_{n+1}=i)}{P(X_n=j)} = a_{j|i}$ and

$$P\left(Y_{n+1} = o_{n+1}, \ldots, Y_N = o_N | X_n = j, X_{n+1} = i\right) =$$

$$= P\left(Y_{n+1} = o_{n+1}, \ldots, Y_N = o_N | X_{n+1} = i\right).$$

Backward Recursion

We apply furthermore one of the previous properties

$$P\left(Y_{n+1}=o_{n+1},\ldots,Y_N=o_N|X_{n+1}=i\right)=$$

$$=P\left(Y_{n+1}=o_{n+1}|X_{n+1}=i\right)\cdot P\left(Y_{n+2}=o_{n+2},\ldots,Y_N=o_N|X_{n+1}=i\right).$$

KTH Matematik

Backward Recursion

Hence it follows that

$$\beta_n(j) = \sum_{i=1}^{J} P\left(Y_{n+1} = o_{n+1} | X_{n+1} = i\right) \cdot P\left(Y_{n+2} = o_{n+2}, \ldots, Y_N = o_N | X_{n+1} = i\right) \cdot a_{j|i}$$

Recalling the definition of the backward variable and the emission probability $b_i\left(o_{n+1}\right) = P\left(Y_{n+1} = o_{n+1} | X_{n+1} = i\right)$ we have

$$\beta_n(j) = \sum_{i=1}^{J} b_i\left(o_{n+1}\right) \cdot \beta_{n+1}(i) \cdot a_{j|i}.$$

KTH Matematik

The Backward Procedure

Consider the backward variable $\beta_n(j)$ defined as

$$\beta_n(j) = P\left(Y_{n+1} = o_{n+1}, \ldots, Y_N = o_N | X_n = j\right),$$

which is the probability of the emitted subsequence $o_{n+1} \ldots o_N$ (from $n+1$ till the end) conditioned on the hidden chain being in the state j at time n (conditional on the model λ).

Start:

$$\beta_N(j) = 1 \; j = 1, \ldots, J.$$

Recursion :

$$\beta_n(j) = \sum_{i=1}^{J} b_i\left(o_{n+1}\right) \cdot \beta_{n+1}(i) \cdot a_{j|i}. \; j = 1, \ldots, J, \; n = N-1, N-2, \ldots, 0.$$

□

KTH Matematik

The Scoring (Evaluation) Problem

$$L_N = P\left(Y_0 = o_0, \ldots, Y_N = o_N\right) = \sum_{j=1}^{J} \alpha_n(j) \cdot \beta_n(j).$$

Hence we have for any $n = 0, \ldots, N$ a respective way of computing L_N. For example with $n = N$ we have

$$L_N = \sum_{j=1}^{J} \alpha_N(j)$$

by the convention $\beta_N(j) = 1$.

KTH Matematik

We wish to find the state sequence that maximizes the probability

$$P\left(Y_0 = o_0, \ldots, Y_N = o_N, X_0 = j_0, \ldots, X_N = j_N\right)$$

by selection of $j_0 \ldots j_N$, when the sequence $o_0 \ldots o_N$ is fixed and the model λ is known and omitted in the notation.

Alignment Problem

Let us set

$$\delta_n(j) = \max_{j_0 \cdots j_{n-1}} P\left(Y_0 = o_0, \ldots, Y_n = o_n, X_0 = j_0, \ldots, X_n = j\right),$$

which is the highest probability along a single subsequence of states that at time n is in state j and accounts for the first

$n + 1 \leq N$ emitted symbols.

KTH Matematik

Bellman's optimality principle.

Proposition

$$\delta_n(j) = \left[\max_{i=1,\ldots,J} \delta_{n-1}(i) \cdot a_{i|j} \right] \cdot b_j\left(o_n\right).$$

KTH Matematik

Alignment Problem

Proof: Using the notational conventions we set

$$P\left(Y_0 = o_0, \ldots, Y_n = o_n, X_0 = j_0, \ldots, X_n = j\right) = P\left(Y^{(n)} \mid X^{(n)}\right) \cdot P\left(X^{(n)}\right).$$

By the conditional independence of the emitted symbols and the definition of conditional probability

$$= \prod_{j=1}^{n} P\left(Y_j \mid X_j\right) \cdot P\left(X_n \mid X^{(n-1)}\right) P\left(X^{(n-1)}\right).$$

KTH Matematik

Alignment Problem

But the Markov property of the hidden chain and some reorganization give

$$= b_j\left(o_n\right) \prod_{j=1}^{n-1} P\left(Y_j \mid X_j\right) \cdot P\left(X_n \mid X_{n-1}\right) P\left(X^{(n-1)}\right) =$$

$$= a_{j_{n-1}|j} \cdot b_j\left(o_n\right) \cdot \prod_{j=1}^{n-1} P\left(Y_j \mid X_j\right) \cdot P\left(X^{(n-1)}\right).$$

Reverting back to the unabridged notation this equals

$$b_j\left(o_n\right)\left[a_{j_{n-1}|j} P\left(Y_0 = o_0, \ldots, Y_{n-1} = o_{n-1}, X_0 = j_0, \ldots, X_{n-1} = j_{n-1}\right)\right].$$

KTH Matematik

Alignment Problem

For each $j \in S$ at time n we have to find the transition to that state from every state $i \in S$ at time $n-1$ giving the best score (in the sense above). There are many paths leading to i at time $n-1$. But we see that the score for the current transition is factorized as the product

$$a_{j_{n-1}|j} P\left(o_0, \ldots, o_{n-1}, X_0, \ldots, X_{n-1}\right).$$

But this shows again that if we do not choose at time $n-1$ for every i that special subsequence leading to i with maximal probability, we cannot obtain

$$\max_{j_0 \cdots j_{n-1}} a_{j_{n-1}|j} P\left(o_0, \ldots, o_{n-1}, X_0, \ldots, X_{n-1} = i\right).$$

KTH Matematik

Hence

$$\max_{j_0 \ldots j_{n-1}} a_{j_{n-1}|j} P\left(o_0, \ldots, o_{n-1}, X_0 = j_0, \ldots, X_{n-1} = j_{n-1}\right) =$$

$$\left[\max_{i=1,\ldots,J} \delta_{n-1}(i) \cdot a_{i|j}\right]$$

as was to be proved. □

KTH Matematik

Alignment Problem

The subsequence yielding $\delta_n(j)$ is called a *survivor* and denoted by

$$\psi_n(j) = \mathrm{argmax}_{i=1,\ldots,J} \delta_{n-1}(i) \cdot a_{i|j}$$

and consists of the prefix yielding $\delta_{n-1}(i)$ concatenated by the best scoring transition between times $n-1$ and n. Hence we need at any j and any n only remember the survivor and no other path leading to this state through the trellis.

KTH Matematik

Alignment Problem

The complete procedure yielding the best decoded state sequence (path) is now formalizable in the following manner.

Viterbi Algorithm (VA)

Storage: n time index, for each $j \in S$ the survivor $\psi_n(j)$ and the corresponding scores $\delta_n(j)$, $j \in S$.

Start: $n = 0$. Compute for each $j \in S$

$$\delta_0(j) = \pi_j(0) \cdot b_j(o_0),$$

$$\psi_0(j) = \emptyset.$$

Recursion: Compute

$$\delta_{n+1}(j) = \left[\max_{i=1,\ldots,J} \delta_n(i) \cdot a_{i|j} \right] \cdot b_j(o_{n+1}).$$

Store the survivors

$$\psi_n(j) = \mathrm{argmax}_{i=1,\ldots,J} \delta_{n-1}(i) \cdot a_{i|j}.$$

for $j = 1, \ldots, N$. Set $n+1$ to n unless $n = N$ and repeat.

KTH Matematik

Viterbi Algorithm (VA)

Termination:

$$P^* = \max_{i=1,\ldots,J} \delta_N(i)$$

$$j_*(N) = argmax_{i=1,\ldots,J} \delta_N(i).$$

Backtracking: The best path is found by

$$j_*(n) = \psi_{n+1}\left(j_*\left(n+1\right)\right), n = N-1, N-2, \ldots, 0.$$

□

KTH Matematik

Quasiloglikelihood for HMM

Let now t denote the number of state sequences $\mathbf{x} = j_0 j_1 \ldots j_n$ of length $n+1$ *that have positive probability with regard to the model* λ *with the given sequence of emission symbols* $\mathbf{o}$.

KTH Matematik

Quasiloglikelihood for HMM

We enumerate the state sequences $(j_0, \ldots, j_n)$ by the index s, $s = 1, \ldots, t$. Then we set

$$u_s = P\left(Y_0 = o_0 \ldots, Y_n = o_n, X_0 = j_0, \ldots, X_n = j_n; \lambda\right)$$

$$\text{if } \left(j_0 \ldots j_n\right) \mapsto s.$$

Quasiloglikelihood for HMM

For any other model λ^* we set

$$v_s = P\left(Y_0 = o_0 \ldots, Y_n = o_n, X_0 = j_0, \ldots, X_n = j_n; \lambda^*\right)$$

$$\text{if } (j_0 \ldots j_n) \mapsto s.$$

Note that some v_s may be in fact be equal to zero, since we are checking state paths with positive probability with regard to λ.

We have to exclude the converse situation and thus make the following assumption.

KTH Matematik

Quasiloglikelihood for HMM

We assume that the model λ^* does not assign a positive probability, conditioned on the given $\mathbf{o}$, to a state path in S^{n+1} that has probability zero with regard to the model λ or, if we have $\mathbf{x}^\dagger = j_0^\dagger \dots j_n^\dagger$ such that

$$P\left(Y_0 = o_0 \dots, Y_n = o_n, X_0 = j_0^\dagger, \dots, X_n = j_n^\dagger; \lambda\right) = 0$$

then

$$P\left(Y_0 = o_0 \dots, Y_n = o_n, X_0 = j_0^\dagger, \dots, X_n = j_n^\dagger; \lambda^*\right) = 0.$$

KTH Matematik

$$\ln \frac{P\left(Y_0 = o_0 \ldots, Y_n = o_n; \lambda^*\right)}{P\left(Y_0 = o_0 \ldots, Y_n = o_n; \lambda\right)},$$

which is comparing the plausibility of the two models for the fixed sequence of emitted symbols.

Under the assumptions above for $s = 1, \ldots, t$ we have

$$u_s > 0$$

and

$$\ln \frac{P\left(Y_0 = o_0 \ldots, Y_n = o_n; \lambda^*\right)}{P\left(Y_0 = o_0 \ldots, Y_n = o_n; \lambda\right)} \geq$$

$$\frac{Q\left(\lambda, \lambda^*\right) - Q\left(\lambda, \lambda\right)}{P\left(Y_0 = o_0 \ldots, Y_n = o_n; \lambda\right)},$$

Quasiloglikelihood for HMM

$$Q(\lambda, \lambda^*) = Q(\lambda, \lambda^* \mid \mathbf{o}) = \sum_{s=1}^{t} u_s \ln v_s$$

and

$$Q(\lambda, \lambda) = Q(\lambda, \lambda \mid \mathbf{o}) = \sum_{s=1}^{t} u_s \ln u_s.$$

KTH Matematik

Quasiloglikelihood for HMM

$$Q(\lambda, \lambda^*) = \sum_{s=1}^{t} u_s \ln v_s =$$

$$= \sum_{s=1}^{t} u_s \left[\sum_{j=1}^{J} r_j(s) \ln \pi_j^*(0) + \right.$$

$$\left. \sum_{j=1}^{J} \sum_{k=1}^{K} m_{j|k}(s) \ln b_j^*(o_k) + \sum_{j=1}^{J} \sum_{i=1}^{J} n_{i|j}(s) \ln a_{i|j}^* \right] =$$

(interchanging the order of the finite summations)

KTH Matematik

Baum-Welch

We maximize

$$\sum_{j=1}^{J}\left[\sum_{s=1}^{t} u_s r_j(s)\right] \ln \pi_j^*(0) +$$

$$\sum_{j=1}^{J}\sum_{k=1}^{K}\left[\sum_{s=1}^{t} u_s m_{j|k}(s)\right] \ln b_j^*(o_k) +$$

$$\sum_{j=1}^{J}\sum_{i=1}^{J}\left[\sum_{s=1}^{t} u_s n_{i|j}(s)\right] \ln a_{i|j}^*.$$

as function of the unknown parameters. This gives:

Baum-Welch

1. For $j = 1, \ldots, J$,

$$\pi_j^*(0) = \frac{e_j}{P\left(Y_0 = o_0 \ldots, Y_n = o_n; \lambda\right)}.$$

2. For $j = 1, \ldots, J$ and for $k = 1, \ldots, K$

$$b_j^*(o_k) = \frac{d_{j|k}}{n_j}.$$

3. For $j = 1, \ldots, J$ and for $i = 1, \ldots, J$

$$a_{i|j}^* = \frac{c_{i|j}}{\sum_{j=1}^{J} c_{i|j}}.$$

KTH Matematik

Baum-Welch

1. For $j = 1, \ldots, J$,

$$\pi_j^*(0) = \frac{\alpha_0(j) \cdot \beta_0(j)}{P\left(Y_0 = o_0, \ldots, Y_n = o_n\right)}. \tag{1}$$

π_j^* is the expected frequency of j at starting time given $o_0 \ldots o_n$ and conditioned on the current model λ.

KTH Matematik

Baum-Welch

2. For $j = 1, \ldots, J$ and for $k = 1, \ldots, K$

$$b_j^*(o_k) = \frac{\sum_{l=0}^{n} I_{\{Y_l = o_k\}} \alpha_l(j) \cdot \beta_l(j)}{\sum_{l=0}^{n} \alpha_l(j) \cdot \beta_l(j)}. \tag{2}$$

$b_j^*(o_k)$ **is the expected number of visits in state** j **and emitting the symbol** o_k **divided by the expected number of transitions from state** j**, given** $o_0 \ldots o_n$ **and conditioned on the current model** λ.

KTH Matematik

Baum-Welch

3. For $j = 1, \ldots, J$ and for $i = 1, \ldots, J$

$$a^*_{i|j} = \frac{a_{i|j} \cdot \sum_{l=0}^{n-1} \alpha_l(i) \cdot b_j(o_{l+1}) \cdot \beta_{l+1}(j)}{\sum_{l=0}^{n-1} \alpha_l(i) \cdot \beta_l(i)}. \tag{3}$$

$a^*_{i|j}$ **is the ratio of the expected number of transitions from state** i **to state** j **divided by the expected number of transitions from state** i **given** $o_0 \ldots o_n$ **and conditioned on the current model** λ.

Consider a Markov chain $(X_k)_{k=0}^{\infty}$ with the state space $\{0, 1\}$ and with the transition probability matrix

$$A = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

and with the initial distribution

$$\pi(0) = (1-a, a) \, .$$

The emitted sequence $(Y_k)_{k=0}^{\infty}$ is given by

$$Y_k = \begin{cases} 1 & \text{if } X_k + V_k \geq 1 \\ 0 & \text{if } X_k + V_k \leq 0, \end{cases}$$

where $(V_k)_{k=0}^{\infty}$ is a sequence of independent, identically distributed discrete random variables, which are independent of of $(X_k)_{k=0}^{\infty}$, too. The variables V_k assume values in the alphabet $\{-1, 0, 1\}$ with the probabilities

$$1 - p_0 - p_1, p_0, p_1$$

respectively.

(a) Show that this is a hidden Markov model in the sense of our definition. Give the emission probability matrix B.

(b) Let for $j = 0, 1$

$$\widehat{\pi}_j \left(n|m\right) = P\left(X_n = j|\, Y_0, \ldots, Y_m\right)$$

be the prediction ($n > m$) or filtering ($n = m$) probability. Show that

$$\widehat{\pi}_1 \left(n+1|n\right) = p - \left(p - (1-q)\right) \cdot \widehat{\pi}_1 \left(n|n\right).$$