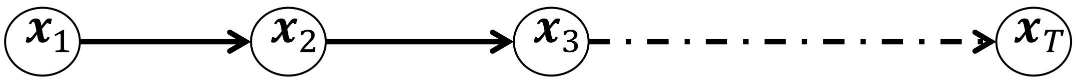
Hidden Markov Models (HMM)

(a) Computing EM in Hidden Markov Models

(b) The Viterbi algorithm

Markov Chains with Discrete Random Variables



Let's assume we have discrete random variables (e.g., taking 3 discrete values $\boldsymbol{x}_t = \{\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}\}$)

Markov Property: $p(\boldsymbol{x}_t|\boldsymbol{x}_1, .., \boldsymbol{x}_{t-1}) = p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$

$$\text{e.g. } p(\boldsymbol{x}_t = \begin{bmatrix}1\\0\\0\end{bmatrix} |\boldsymbol{x}_{t-1} = \begin{bmatrix}0\\1\\0\end{bmatrix})$$

Stationary, Homogeneous or Time-Invariant if the distribution $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$ does not depend on t

Markov Chains with Discrete Random Variables

$$p(\boldsymbol{x}_1, .., \boldsymbol{x}_T) = p(\boldsymbol{x}_1) \prod_{t=2}^{\mathrm{T}} p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1})$$

What do we need in order to describe the whole procedure?

(1) A probability for the first frame/timestamp etc $p(\boldsymbol{x}_1)$. In order to define the probability we need to define the vector $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_{\mathrm{K}})$

$$p(\boldsymbol{x}_1 | \boldsymbol{\pi}) = \prod_{c=1}^{K} \pi_c{}^{x_{1c}}$$

(2) A transition probability $p(\boldsymbol{x_t} | \boldsymbol{x_{t-1}})$. In order to define it we need a KxK transition matrix $\boldsymbol{A} = [a_{ij}]$

$$p(\boldsymbol{x}_t | \boldsymbol{x}_{t-1}, \boldsymbol{A}) = \prod_{j=1}^{K} \prod_{k=1}^{K} a_{jk}{}^{x_{t-1j} x_{tk}}$$

Hidden Markov Models

A casino has two coins (our latent variable is probably the coin ☺)

$$\boldsymbol{z} = \{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\}$$

One is fair and one is not

Each coin has 2 sides.

$$x = \{\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}\}$$

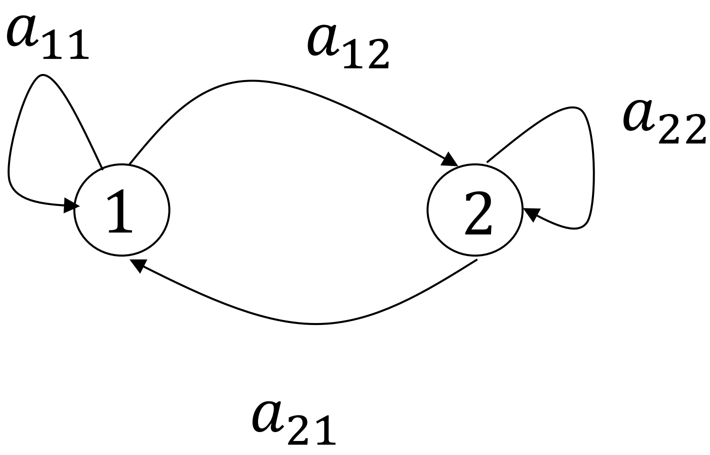
(1) Fair coin $p\left(x_j = 1 | z_1 = 1\right) = \frac{1}{2}$ for all $j = \{1,2\}$

(2) Loaded coin $p(x_1 = 1 | z_2 = 1) = b_{12}$ for $j = \{1,2\}$ and $p(x_2 = 1 | z_2 = 1) = 1 - b_{12}$

Emission probability

Hidden Markov Models

A casino player switches back & forth between fair and loaded coin



$$p(z_{11} = 1) = \pi_1$$

$$p(z_{12} = 1) = \pi_2$$

$$p(\boldsymbol{z}_1|\boldsymbol{\pi}) = \prod_{c=1}^{2} \pi_c{}^{z_{1c}}$$

$$p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, \boldsymbol{A}) = \prod_{j=1}^{2} \prod_{k=1}^{2} a_{jk}{}^{z_{t-1j} z_{tk}}$$

$$p(\boldsymbol{x}_t|\boldsymbol{z}_t) = \prod_{j=1}^{2} \prod_{k=1}^{2} b_{jk}{}^{x_{tj} z_{tk}}$$

EM in HMMs

Given a set of strings of observations (or even one) and the above model:

HHTT, THTH, HHHT, HTHH

Find the parameters $\boldsymbol{A}$, $\boldsymbol{\pi}$, b_{12} by maximizing the probability

$$p(D_1, \dots, D_4, Z_1, \dots, Z_4 | \theta)$$
$$= \prod_{l=1}^{4} p\left(\boldsymbol{x}_1{}^l, \boldsymbol{x}_2{}^l, \cdots, \boldsymbol{x}_4{}^l, \boldsymbol{z}_1{}^l, \boldsymbol{z}_2{}^l, \cdots, \boldsymbol{z}_4{}^l | \theta\right)$$
$$= \prod_{l=1}^{4} \prod_{t=1}^{4} p(\boldsymbol{x}_t{}^l | \boldsymbol{z}_t{}^l)\, p(\boldsymbol{z}_1{}^l) \prod_{t=2}^{4} p(\boldsymbol{z}_t{}^l | \boldsymbol{z}_{t-1}{}^l)$$

EM in HMMs

$$= \prod_{l=1} \prod_{t=1}^{4} p(\boldsymbol{x}_t{}^l | \boldsymbol{z}_t{}^l)\, p(\boldsymbol{z}_1{}^l) \prod_{t=2}^{4} p(\boldsymbol{z}_t{}^l | \boldsymbol{z}_{t-1}{}^l)$$

$$= \prod_{l=1}^{4} \prod_{t=1}^{4} \prod_{j=1}^{2} \prod_{k=1}^{2} b_{jk}{}^{x_{tj}{}^l z_{tk}{}^l} \prod_{k=1}^{2} \pi_k{}^{z_{1k}{}^l} \prod_{t=2}^{4} \prod_{j=1}^{2} \prod_{k=1}^{2} a_{jk}{}^{z_{t-1j}{}^l z_{tk}{}^l}$$

$$\underset{ln}{\Rightarrow} = \sum_{l=1}^{4} \sum_{t=1}^{4} \sum_{j=2}^{2} \sum_{k=1}^{2} x_{tj}{}^l z_{tk}{}^l \ln b_{jk} + \sum_{l=1}^{4} \sum_{k=1}^{2} z_k{}^l \ln \pi_k$$

$$+ \sum_{l=1}^{4} \sum_{t=2}^{4} \sum_{j=1}^{2} \sum_{k=1}^{2} z_{t-1j}{}^l z_{tk}{}^l \ln a_{jk}$$

EM in HMMs

Taking the expectations with regards to the posterior

$$= \sum_{l=1}^{4} \sum_{t=1}^{4} \sum_{j=2}^{2} \sum_{k=1}^{2} x_{tj}{}^{l} E[z_{tk}{}^{l}] \ln b_{jk}$$

$$+ \sum_{l=1}^{4} \sum_{k=1}^{2} E[z_{1k}{}^{l}] \ln \pi_k$$

$$+ \sum_{l=1}^{4} \sum_{t=2}^{2} \sum_{j=1}^{2} \sum_{k=1}^{2} E[z_{t-1j}{}^{l} z_{tk}{}^{l}] \ln a_{jk}$$

EM in HMMs

$$E[{z_{1k}}^l] = \sum_{{z_{1k}}^l} {z_{1k}}^l p({z_{1k}}^l | {\boldsymbol{x}_1}^l, \cdots, {\boldsymbol{x}_T}^l) = p({z_{1k}}^l = 1 | {\boldsymbol{x}_1}^l, \cdots, {\boldsymbol{x}_T}^l)$$

$$E[{z_{tk}}^l] = \sum_{{z_{tk}}^l} {z_{tk}}^l p({z_{tk}}^l | {\boldsymbol{x}_1}^l, \cdots, {\boldsymbol{x}_T}^l) = p({z_{tk}}^l = 1 | {\boldsymbol{x}_1}^l, \cdots, {\boldsymbol{x}_T}^l)$$

$$E\left[{z_{t-1j}}^l {z_{tk}}^l\right] = \sum_{{z_{t-1j}}^l} \sum_{{z_{tk}}^l} {z_{t-1j}}^l {z_{tk}}^l p({z_{t-1j}}^l {z_{tk}}^l | {\boldsymbol{x}_1}^l, {\boldsymbol{x}_2}^l, \cdots, {\boldsymbol{x}_T}^l)$$

$$= p({z_{t-1j}}^l = 1, {z_{tk}}^l = 1 | {\boldsymbol{x}_1}^l, {\boldsymbol{x}_2}^l, \cdots, {\boldsymbol{x}_T}^l)$$

EM in HMMs

$$p\left(z_{1k}{}^{l}=1\middle|\boldsymbol{x}_1{}^{l},\cdots,x_T{}^{l}\right)$$

$$=\frac{p\left(\boldsymbol{x}_1{}^{l},z_{1k}{}^{l}=1\right)p(\boldsymbol{x}_2{}^{l},\cdots,\boldsymbol{x}_T{}^{l}|z_{1k}{}^{l}=1)}{p(\boldsymbol{x}_1{}^{l},\cdots,\boldsymbol{x}_T{}^{l})}=\frac{\alpha\left(z_{1k}{}^{l}\right)\beta(z_{1k}{}^{l})}{p(\boldsymbol{x}_1{}^{l},\cdots,\boldsymbol{x}_T{}^{l})}$$

$$p\left(z_{tk}{}^{l}=1\middle|\boldsymbol{x}_1{}^{l},\cdots,\boldsymbol{x}_T{}^{l}\right)$$

$$=\frac{p\left(\boldsymbol{x}_1{}^{l},\cdots,\boldsymbol{x}_t{}^{l},z_{tk}{}^{l}=1\right)p(\boldsymbol{x}_{t+1}{}^{l},\cdots,\boldsymbol{x}_T{}^{l}|z_{tk}{}^{l}=1)}{p(\boldsymbol{x}_1{}^{l},\cdots,\boldsymbol{x}_T{}^{l})}=\frac{\alpha\left(z_{tk}{}^{l}\right)\beta(z_{tk}{}^{l})}{p(\boldsymbol{x}_1{}^{l},\cdots,\boldsymbol{x}_T{}^{l})}$$

$$p\left(z_{t-1j}{}^{l}=1,z_{tk}{}^{l}=1\middle|\boldsymbol{x}_1{}^{l},\cdots,\boldsymbol{x}_T{}^{l}\right)$$

$$=\frac{\alpha\left(z_{t-1j}{}^{l}\right)\prod_{r=1}^{2}b_{rk}{}^{x_{tj}{}^{l}}\,a_{jk}\beta(z_{tk}{}^{l})}{p(\boldsymbol{x}_1{}^{l},\cdots,\boldsymbol{x}_T{}^{l})}$$

Computing the E-step

Initialize parameters

$$\boldsymbol{\pi} = \begin{bmatrix} \pi_1 \\ \pi_2 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$$

$$p(\boldsymbol{z}_t \mid \boldsymbol{z}_{t-1})$$

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & 1 - a_{11} \\ 1 - a_{22} & a_{22} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \qquad a_{11} = a_{22} = 1/2$$

$$p(\boldsymbol{x}_t \mid \boldsymbol{z}_t)$$

$$\boldsymbol{b}_1 = \begin{bmatrix} 1/2 \\ 3/4 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} b_{11} & 1 - b_{11} \\ b_{12} & 1 - b_{12} \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 3/4 & 1/4 \end{bmatrix}$$

$$\boldsymbol{b}_2 = \begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix}$$

Computing the E-step

First sequence: HHTT (compute $a(\boldsymbol{z}_i{}^1)$, $\beta(\boldsymbol{z}_i{}^1)$)

$$a(\boldsymbol{z}_1{}^1) = p(\boldsymbol{x}_1{}^1, \boldsymbol{z}_1{}^1) = p(\boldsymbol{z}_1{}^1)\mathrm{p}(x_{11}{}^1 = 1, \boldsymbol{z}_1{}^1) \quad \text{HHTT}$$

$$a(\boldsymbol{z}_1{}^1) = \boldsymbol{\pi} \odot \boldsymbol{b}_1 = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \odot \begin{bmatrix} 1/2 \\ 3/4 \end{bmatrix} = \begin{bmatrix} 1/4 \\ 3/8 \end{bmatrix}$$

$$a(\boldsymbol{z}_2{}^1) = p(x_{21}{}^1 = 1 | z_2{}^1) \sum_{z_1} p(\boldsymbol{z}_2{}^1 | \boldsymbol{z}_1{}^1) a(\boldsymbol{z}_1{}^1) \quad \text{HHTT}$$

$$= \boldsymbol{b}_1 \odot (\boldsymbol{A}^T a(\boldsymbol{z}_1{}^1)) = \begin{bmatrix} 1/2 \\ 3/4 \end{bmatrix} \odot \left(\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1/2 \\ 3/4 \end{bmatrix} \right)$$

$$= \begin{bmatrix} 1/2 \\ 3/4 \end{bmatrix} \odot \begin{bmatrix} 5/16 \\ 5/16 \end{bmatrix} = \begin{bmatrix} 5/32 \\ 15/64 \end{bmatrix}$$

Computing the E-step

HHTT

$$a(\boldsymbol{z}_3{}^1) = p(x_{32}{}^1 = 1 | z_3{}^1) \sum_{\boldsymbol{z}_2} p(\boldsymbol{z}_3{}^1 | \boldsymbol{z}_2{}^1) a(\boldsymbol{z}_2{}^1)$$

$$= \boldsymbol{b}_2 \odot (\boldsymbol{A}^T a(\boldsymbol{z}_2{}^1)) = \begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix} \odot \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 5/32 \\ 15/64 \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix} \odot \begin{bmatrix} \frac{25}{128} \\ \frac{25}{128} \end{bmatrix} = \begin{bmatrix} 25/256 \\ 25/512 \end{bmatrix}$$

Computing the E-step

HHTT

$$a(\boldsymbol{z}_4{}^1) = \boldsymbol{p}(\boldsymbol{x}_{42}{}^1 = \mathbf{1} | \boldsymbol{z}_3{}^1) \sum_{z_3} p(\boldsymbol{z}_4{}^1 | \boldsymbol{z}_3{}^1) a(\boldsymbol{z}_3{}^1)$$

$$= \boldsymbol{b}_2 \odot \left(\mathbf{A} a(\boldsymbol{z}_3{}^1)\right) = \begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix} \odot \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 25/256 \\ 25/512 \end{bmatrix}$$

$$= \begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix} \odot \begin{bmatrix} \frac{75}{1024} \\ \frac{75}{1024} \end{bmatrix} = \begin{bmatrix} 75/2048 \\ 75/4096 \end{bmatrix}$$

$$p(\boldsymbol{x}_1{}^1, \boldsymbol{x}_2{}^1, \boldsymbol{x}_3{}^1, \boldsymbol{x}_4{}^1) = p(\text{HHTT}) = \frac{150 + 75}{4096} = \frac{225}{4096}$$

Computing the E-step

HHTT

$$\beta(\boldsymbol{z}_4{}^1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

HHTT $\beta(\boldsymbol{z}_3{}^1) = \sum_{z_4} \beta(\boldsymbol{z}_4{}^1) p(\boldsymbol{x}_{42}{}^1 = 1 | \boldsymbol{z}_4{}^1) \, \boldsymbol{p}(\boldsymbol{z}_4{}^1 | \boldsymbol{z}_3{}^1)$

$$= \boldsymbol{A}(\beta(\boldsymbol{z}_4{}^1) \odot \boldsymbol{b_2})$$

$$= \boldsymbol{A}\left(\begin{bmatrix} 1 \\ 1 \end{bmatrix} \odot \begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix}\right) = \boldsymbol{A}\left(\begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix}\right)$$

$$= \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix} = \begin{bmatrix} 3/8 \\ 3/8 \end{bmatrix}$$

Computing the E-step

HHTT

$$\beta(\boldsymbol{z}_2{}^1) = \sum_{z_3} \beta(\boldsymbol{z}_3{}^1) p(\boldsymbol{x}_{32}{}^1 = 1|\boldsymbol{z}_3{}^1)\, \boldsymbol{p}(\boldsymbol{z}_3{}^1|\boldsymbol{z}_2{}^1)$$

$$= \boldsymbol{A}(\beta(\boldsymbol{z}_3{}^1) \odot \boldsymbol{b}_2)$$

$$= \boldsymbol{A}\left(\begin{bmatrix} 3/8 \\ 3/8 \end{bmatrix} \odot \begin{bmatrix} 1/2 \\ 1/4 \end{bmatrix}\right)$$

$$= \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \begin{bmatrix} 3/16 \\ 3/32 \end{bmatrix} = \begin{bmatrix} 9/64 \\ 9/64 \end{bmatrix}$$

Computing the E-step

HHTT

$$\beta(\boldsymbol{z}_1{}^1) = \boldsymbol{A}(\beta(\boldsymbol{z}_2{}^1) \odot \boldsymbol{b}_1)$$

$$= \boldsymbol{A}\left(\begin{bmatrix} 9/64 \\ 9/64 \end{bmatrix} \odot \begin{bmatrix} 1/2 \\ 3/4 \end{bmatrix}\right)$$

$$= \boldsymbol{A}\begin{bmatrix} 9/128 \\ 27/256 \end{bmatrix} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}\begin{bmatrix} 9/128 \\ 27/256 \end{bmatrix} = \begin{bmatrix} 45/512 \\ 45/512 \end{bmatrix}$$

Computing the E-step

$$\boldsymbol{C_1} = [\alpha(\boldsymbol{z}_1{}^1)\ \ \alpha(\boldsymbol{z}_2{}^1)\ \ \alpha(\boldsymbol{z}_3{}^1)\ \ \alpha(\boldsymbol{z}_4{}^1)]$$

$$= \begin{bmatrix} 1/4 & 5/32 & 25/256 & 75/2048 \\ 3/8 & 15/64 & 25/512 & 75/4096 \end{bmatrix}$$

$$\boldsymbol{C_2} = [\beta(\boldsymbol{z}_1{}^1)\ \ \beta(\boldsymbol{z}_2{}^1)\ \ \beta(\boldsymbol{z}_3{}^1)\ \ \beta(\boldsymbol{z}_4{}^1)]$$

$$= \begin{bmatrix} 43/512 & 9/64 & 3/8 & 1 \\ 43/512 & 9/64 & 3/8 & 1 \end{bmatrix}$$

Computing the E-step

Now we are ready to compute the expectations $E[z_{tk}{}^1]$as

$$\boldsymbol{E} = (\boldsymbol{C_1} \odot \boldsymbol{C_2})/p(\text{HHTT})$$

$$\frac{225}{4096}$$

$$\boldsymbol{E} = \begin{bmatrix} 45/2048 & 45/2048 & 75/2048 & 75/2048 \\ 135/2048 & 135/4096 & 75/4096 & 75/4096 \end{bmatrix} \frac{4096}{225}$$

$$= \begin{bmatrix} 90/225 & 90/225 & 150/225 & 150/225 \\ 135/225 & 135/225 & 75/225 & 75/225 \end{bmatrix}$$

Computing the E-step

HHTT

$$p(\boldsymbol{z}_t{}^1, \boldsymbol{z}_{t-1}{}^1 | \boldsymbol{x}_1{}^1, \dots, \boldsymbol{x}_4{}^1) = \frac{a(\boldsymbol{z}_{t-1}{}^1) p(\boldsymbol{x}_t{}^1 | \boldsymbol{z}_t{}^1) p(\boldsymbol{z}_t{}^1, \boldsymbol{z}_{t-1}{}^1) \beta(\boldsymbol{z}_t{}^1)}{p(\boldsymbol{x}_1{}^1, \dots, \boldsymbol{x}_4{}^1)}$$

$p(\boldsymbol{z}_t{}^1, \boldsymbol{z}_{t-1}{}^1 | \boldsymbol{x}_1{}^1, \dots, \boldsymbol{x}_4{}^1)$ is $2x2$ matrix and in total we have $T - 1 = 4 - 1$ of these matrices

Computing the E-step

HHTT

$$a(\boldsymbol{z}_1{}^1)p(\boldsymbol{x}_2{}^1|\boldsymbol{z}_2{}^1)p(\boldsymbol{z}_2{}^1,\boldsymbol{z}_1{}^1)\beta(\boldsymbol{z}_2{}^1) =$$

$$= \begin{bmatrix} a(z_{11}{}^1 = 1)p(x_{21}{}^1 = 1|z_{21}{}^1 = 1)p(z_{21}{}^1 = 1|z_{11}{}^1 = 1)\beta(z_{21}{}^1 = 1) \\ a(z_{12}{}^1 = 1)p(x_{21}{}^1 = 1|z_{21}{}^1 = 1)p(z_{21}{}^1 = 1|z_{12}{}^1 = 1)\beta(z_{21}{}^1 = 1) \end{bmatrix}$$

$$\begin{matrix} a(z_{11}{}^1 = 1)p(x_{21}{}^1 = 1|z_{22}{}^1 = 1)p(z_{22}{}^1 = 1|z_{11}{}^1 = 1)\beta(z_{22}{}^1 = 1) \\ a(z_{12}{}^1 = 1)p(x_{21}{}^1 = 1|z_{22}{}^1 = 1)p(z_{22}{}^1 = 1|z_{12}{}^1 = 1)\beta(z_{22}{}^1 = 1) \end{matrix}\Bigg]$$

$$= [a(\boldsymbol{z_1}{}^{\boldsymbol{1}})\ a(\boldsymbol{z_1}{}^{\boldsymbol{1}})] \odot \boldsymbol{A} \odot \begin{bmatrix} [\boldsymbol{b}_1 \odot \beta(\boldsymbol{z_2}{}^{\boldsymbol{1}})]^T \\ [\boldsymbol{b}_1 \odot \beta(\boldsymbol{z_2}{}^{\boldsymbol{1}})]^T \end{bmatrix}$$

Computing the E-step

$$a(\boldsymbol{z}_1{}^1)p(\boldsymbol{x}_2{}^1|\boldsymbol{z}_2{}^1)p(\boldsymbol{z}_2{}^1,\boldsymbol{z}_1{}^1)\beta(\boldsymbol{z}_2{}^1)$$

$$= [a(\boldsymbol{z}_1{}^1)\ a(\boldsymbol{z}_1{}^1)] \odot \boldsymbol{A} \odot \begin{bmatrix} [\boldsymbol{b}_1 \odot \beta(\boldsymbol{z}_2{}^1)]^T \\ [\boldsymbol{b}_1 \odot \beta(\boldsymbol{z}_2{}^1)]^T \end{bmatrix}$$

or equivalently

$$= \boldsymbol{A} \odot \left(a(\boldsymbol{z}_{\boldsymbol{1}}{}^{\boldsymbol{1}}) \left[\boldsymbol{b}_1 \odot \beta(\boldsymbol{z}_{\boldsymbol{2}}{}^{\boldsymbol{1}}) \right]^T \right)$$

Computing the E-step

HHTT

$$[\boldsymbol{a}(\boldsymbol{z_1}^{\mathbf{1}})\ \boldsymbol{a}(\boldsymbol{z_1}^{\mathbf{1}})]\odot\boldsymbol{A}\odot[\boldsymbol{b}_1\odot\beta(\boldsymbol{z_2}^{\mathbf{1}})\quad \boldsymbol{b}_1\odot\beta(\boldsymbol{z_2}^{\mathbf{1}})]^T/p(\text{HHTT})$$

$$=\begin{bmatrix}1/4 & 1/4\\ 3/8 & 3/8\end{bmatrix}\odot\begin{bmatrix}1/2 & 1/2\\ 1/2 & 1/2\end{bmatrix}\odot\begin{bmatrix}9/128 & 27/256\\ 9/128 & 27/256\end{bmatrix}/p(\text{HHTT})$$

$$=\begin{bmatrix}36/225 & 54/225\\ 54/225 & 81/225\end{bmatrix}$$

Computing the E-step

HHTT

$$\boldsymbol{p}(\boldsymbol{z_3}^{\mathbf{1}}, \boldsymbol{z_2}^{\mathbf{1}} | \boldsymbol{x_1}^{\mathbf{1}}, \dots, \boldsymbol{x_4}^{\mathbf{1}})$$

$$= [a(\boldsymbol{z_2}^{\mathbf{1}}) \;\; a(\boldsymbol{z_2}^{\mathbf{1}})] \odot \boldsymbol{A} \odot \begin{bmatrix} [\boldsymbol{b}_2 \odot \beta(\boldsymbol{z_3}^{\mathbf{1}})]^T \\ [\boldsymbol{b}_2 \odot \beta(\boldsymbol{z_3}^{\mathbf{1}})]^T \end{bmatrix} / p(\boldsymbol{x_1}^{\mathbf{1}}, \dots, \boldsymbol{x_4}^{\mathbf{1}})$$

$$= \begin{bmatrix} 5/32 & 5/32 \\ 15/64 & 15/64 \end{bmatrix} \odot \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \odot \begin{bmatrix} \frac{3}{16} & \frac{3}{32} \\ \frac{3}{16} & \frac{3}{32} \end{bmatrix} / (\frac{225}{4096})$$

$$= \begin{bmatrix} 60/225 & 30/225 \\ 90/225 & 45/225 \end{bmatrix}$$

Computing the E-step

HHTT

$$p(\boldsymbol{z_4}^{\mathbf{1}}, \boldsymbol{z_3}^{\mathbf{1}} | \boldsymbol{x_1}^{\mathbf{1}}, \dots, \boldsymbol{x_4}^{\mathbf{1}})$$

$$= [a(\boldsymbol{z_3}^{\mathbf{1}}) \;\; a(\boldsymbol{z_3}^{\mathbf{1}})] \odot \boldsymbol{A} \odot \begin{bmatrix} [\boldsymbol{b}_2 \odot \beta(\boldsymbol{z_4}^{\mathbf{1}})]^T \\ [\boldsymbol{b}_2 \odot \beta(\boldsymbol{z_4}^{\mathbf{1}})]^T \end{bmatrix} / p(\boldsymbol{x_1}^{\mathbf{1}}, \dots, \boldsymbol{x_4}^{\mathbf{1}})$$

$$= \begin{bmatrix} 25/256 & 25/256 \\ 25/512 & 25/512 \end{bmatrix} \odot \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} \odot \begin{bmatrix} 1/2 & 1/4 \\ 1/2 & 1/4 \end{bmatrix} / (\frac{225}{4096})$$

$$= \begin{bmatrix} 100/225 & 50/225 \\ 50/225 & 25/225 \end{bmatrix}$$

The scaling problem

$$\boldsymbol{C_1} = [\alpha(\boldsymbol{z}_1{}^1)\ \ \alpha(\boldsymbol{z}_2{}^1)\ \ \alpha(\boldsymbol{z}_3{}^1)\ \ \alpha(\boldsymbol{z}_4{}^1)]$$

$$= \begin{bmatrix} 1/4 & 5/32 & 25/256 & 75/2048 \\ 3/8 & 15/64 & 25/512 & 75/4096 \end{bmatrix}$$

- These $\alpha(\boldsymbol{z}_1{}^1)$ can go to zero exponentially quickly. Also for

medium length chains (i.e., $T = 100$) the calculation of $\alpha\left(\boldsymbol{z}_t\ \ \right)$ will soon exceed the dynamic range of the computer.

- We compute instead the normalized version which is well-behaved

$$\hat{a}(\boldsymbol{z}_t) = p(\boldsymbol{z}_t | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_t) = \frac{a(\boldsymbol{z}_t)}{p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)}$$

The scaling problem

- We need a recursion with regards to $\hat{a}(\boldsymbol{z}_t)$
- To do so we need to make $p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t)$ in to a product

$$p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t) = \prod_{m=1}^{t} c_m \qquad c_t = p(\boldsymbol{x}_t | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_{t-1})$$

- Hence

$$a(\boldsymbol{z}_t) = p(\boldsymbol{z}_t | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_t) p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_t) = \left(\prod_{m=1}^{t} c_m \right) \hat{a}(\boldsymbol{z}_t)$$

and the recursive formula becomes

$$c_t \hat{a}(\boldsymbol{z}_t) = p(\boldsymbol{x}_t | \boldsymbol{z}_t) \sum_{\boldsymbol{z}_{t-1}} \hat{a}(\boldsymbol{z}_{t-1}) p(\boldsymbol{z}_t | \boldsymbol{z}_{t-1})$$

The scaling problem

- Similarly for β

$$\boldsymbol{C_2} = [\beta(\boldsymbol{z}_1{}^1) \;\; \beta(\boldsymbol{z}_2{}^1) \;\; \beta(\boldsymbol{z}_3{}^1) \;\; \beta(\boldsymbol{z}_4{}^1)]$$

$$= \begin{bmatrix} 43/512 & 9/64 & 3/8 & 1 \\ 43/512 & 9/64 & 3/8 & 1 \end{bmatrix}$$

- We can define re-scaled variables $\hat{\beta}(\boldsymbol{z}_t)$

$$\beta(\boldsymbol{z}_t) = \left(\prod_{m=t+1}^{T} c_m\right)\hat{\beta}(\boldsymbol{z}_t)$$

The scaling problem

$$\hat{\beta}(\boldsymbol{z}_t) = \frac{p(\boldsymbol{x}_{t+1}, \dots, \boldsymbol{x}_T | \boldsymbol{z}_t)}{p(\boldsymbol{x}_{t+1}, \dots, \boldsymbol{x}_T | \boldsymbol{x}_1, \dots, \boldsymbol{x}_t)}$$

and the recursive formula becomes

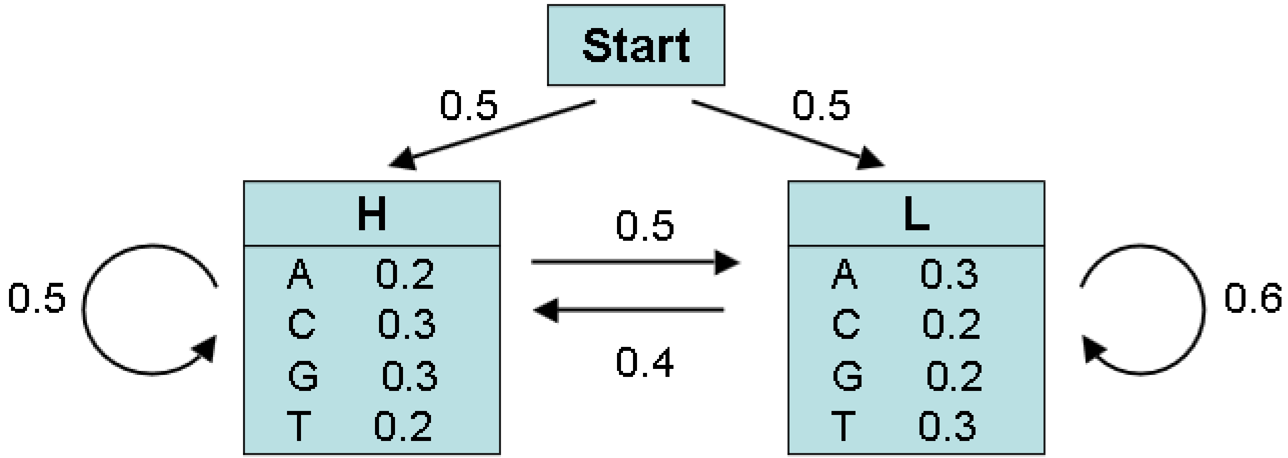
$$c_{t+1}\hat{\beta}(\boldsymbol{z}_t) = \sum_{\boldsymbol{z}_{t+1}} \hat{\beta}(\boldsymbol{z}_{t+1}) p(\boldsymbol{x}_{t+1} | \boldsymbol{z}_{t+1}) p(\boldsymbol{z}_{t+1} | \boldsymbol{z}_t)$$

Hence the posteriors become

$$p(\boldsymbol{z}_t | \boldsymbol{x}_{1:T}) = \hat{a}(\boldsymbol{z}_t)\hat{\beta}(\boldsymbol{z}_t)$$

$$\mathrm{p}(\boldsymbol{z}_{t-1}, \boldsymbol{z}_t | \boldsymbol{x}_{1:T}) = {c_t}^{-1} \hat{a}(\boldsymbol{z}_{t-1})\, p(\boldsymbol{x}_t | \boldsymbol{z}_t)\, p(\boldsymbol{z}_t | \boldsymbol{z}_{t-1}) \hat{\beta}(\boldsymbol{z}_t)$$

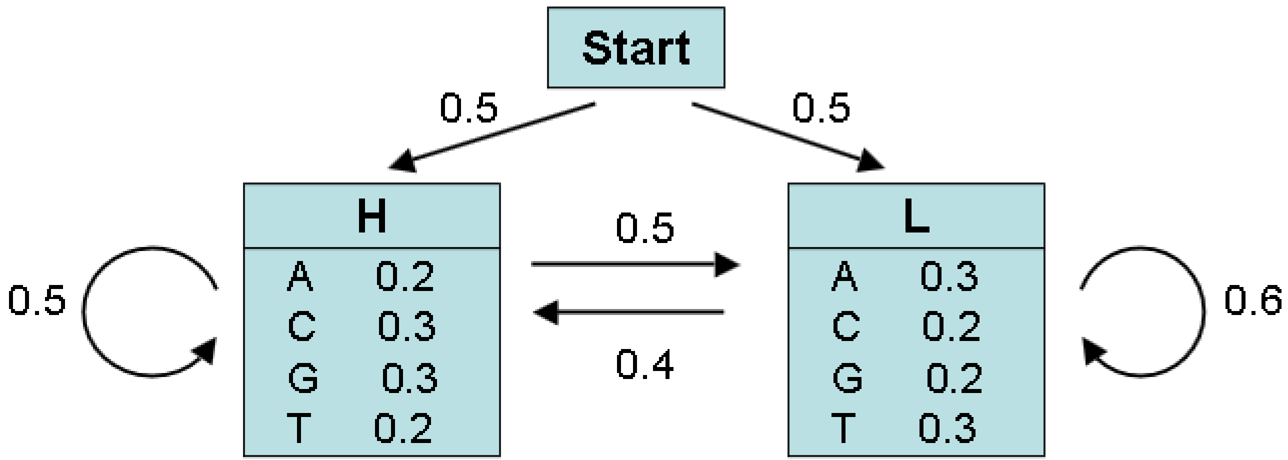
Decoding (Viterbi algorithm)



Let us consider the above HMM (3 states, 4 different letters as observations). State H – coding DNA, state L- Non-coding DNA. Can be used to predict the region of coding DNA from a given sequence.

$$\boldsymbol{z} = \{(H)\begin{bmatrix}1\\0\end{bmatrix}, (L)\begin{bmatrix}0\\1\end{bmatrix}\}$$

Decoding (Viterbi algorithm)



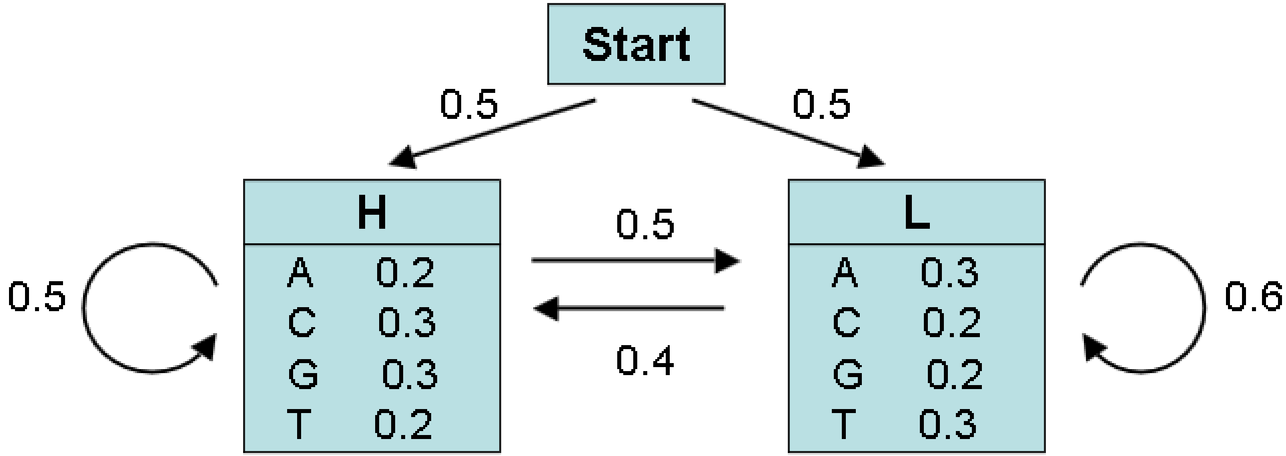
Consider the sequence S=GGCACTGAA.

There are several paths through the hidden states (H and L) that lead to the given sequence.

Example: LLHHHHLLL

$$p(x_1, \ldots, x_9, z_{12} = 1, \ldots, z_{92} = 1)$$
$$= p(z_{12} = 1) p(x_{14} = 1 | z_{12} = 1) p(z_{22} = 1 | z_{12} = 1) p(x_{24} = 1 | z_{22} = 1) \ldots$$
$$= \quad 0.5 * \quad 0.2 * \quad 0.6 * \quad 0.2$$

Decoding (Viterbi algorithm)

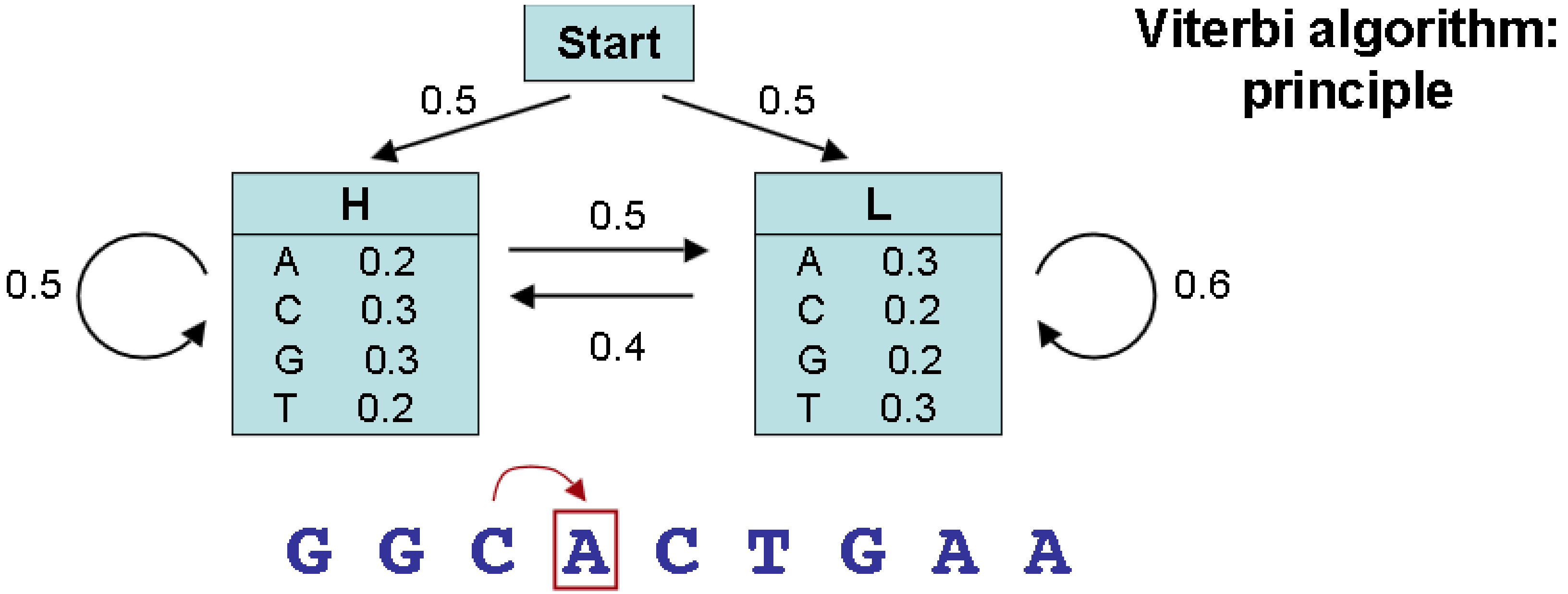


GGCACTGAA

There are several paths through the hidden states (H and L) that lead to the given sequence, but they do not have the same probability.

The Viterbi algorithm is a dynamic programming algorithm that allows us to compute the most probable path.

Decoding (Viterbi algorithm)



The probability of ending up in a state j at time t, given that we take the most probable path.

$$\delta_t(j) \triangleq max_{z_1,\ldots,z_{t-1}} p(z_{1:t-1}, z_{tj} = 1 | x_{1:t})$$

Decoding (Viterbi algorithm)

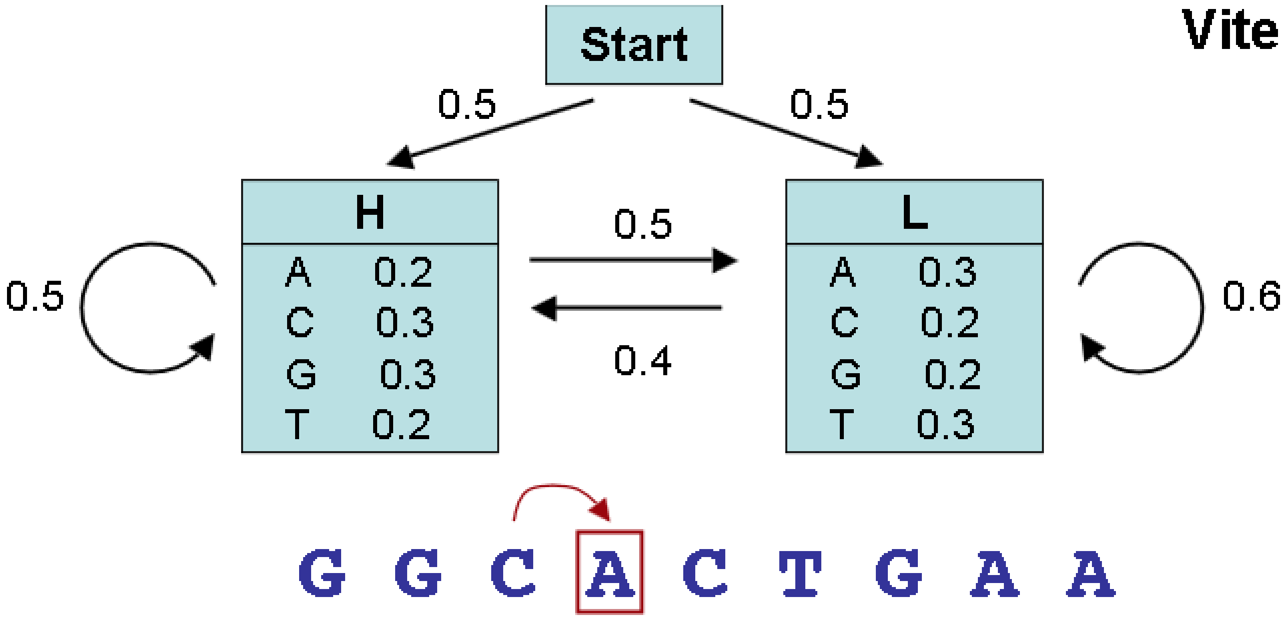
The key insight is that the most probable path to state j at time t must consist of the most probable path to some other state i at time $t-1$, followed by a transition from i to j.

$$\delta_t(j) = {c_t}^{-1} p\left(x_t \middle| z_{tj} = 1\right) max_i\, \delta_{t-1}(i) p\left(z_{tj} = 1 | z_{t-1i} = 1\right)$$

We also keep track of the most likely previous state, for each possible state that we end up

$$a_t(j) = \arg max_i\, p\left(x_t \middle| z_{tj} = 1\right) \delta_{t-1}(i) p\left(z_{tj} = 1 | z_{t-1i} = 1\right)$$

Decoding (Viterbi algorithm)

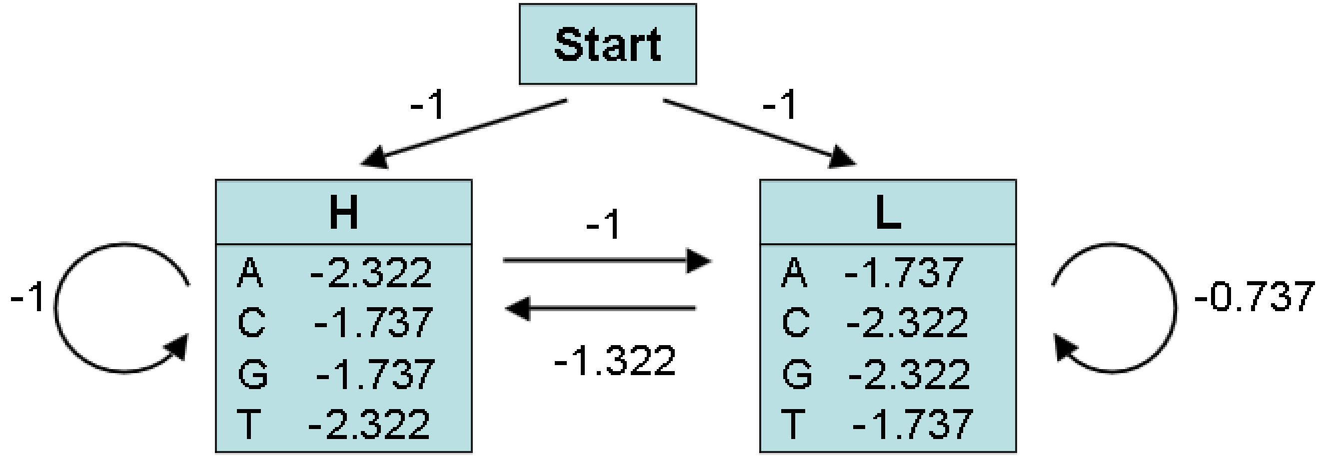


Viterbi algorithm: principle

The probability of the most probable path ending in state H with "A" at the 4th position (recursively)

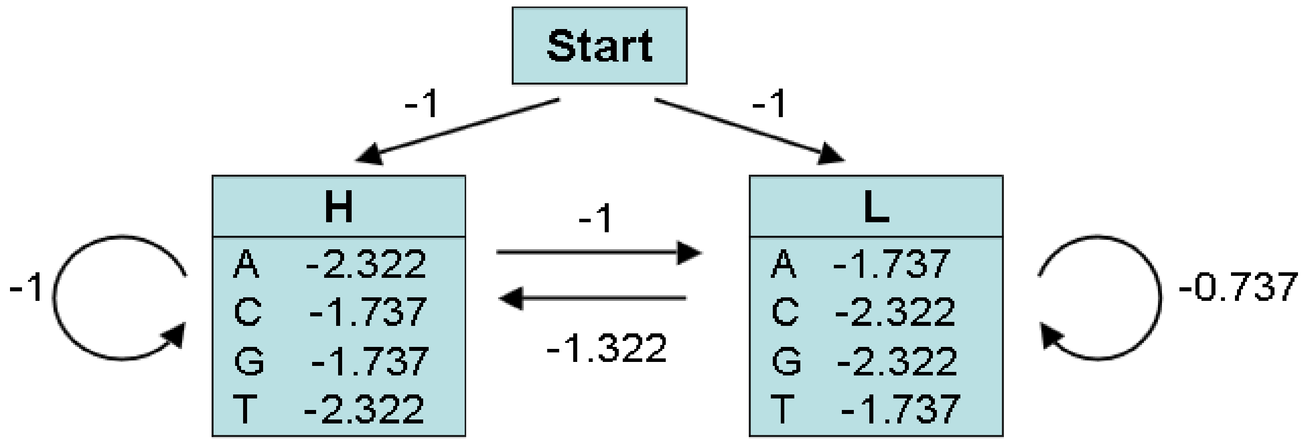
$$\delta_4(j=1) = p\left(x_{41}=1 \middle| z_{4j}=1\right)$$
$$\max(\delta_3(j=2)p(z_{41}=1|z_{32}=1), \delta_3(j=1)p(z_{41}=1|z_{31}=1))$$

Decoding (Viterbi algorithm)



Remark: for the calculations, it is convenient to use the log of the probabilities. Indeed, this allows us to compute sums instead of products which is more efficient and accurate (log_2).

Decoding (Viterbi algorithm)



GGCACTGAA

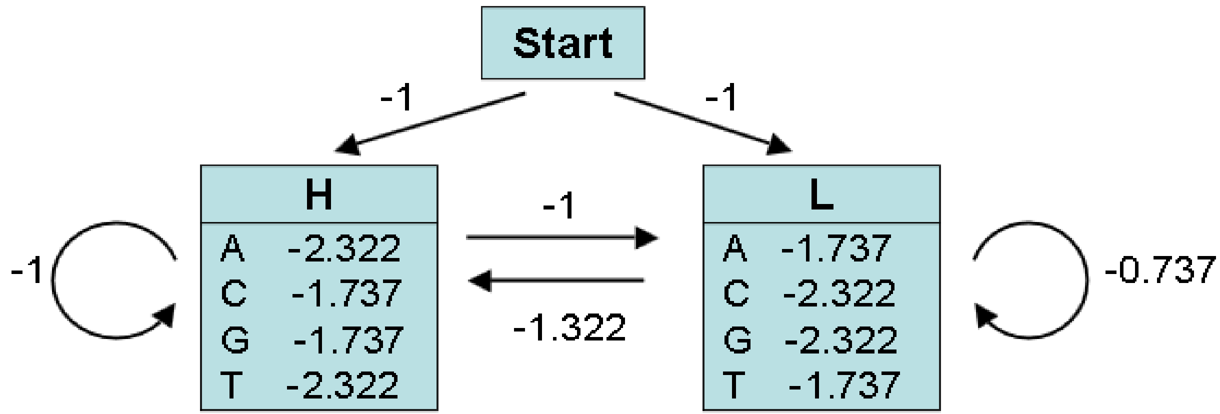
Probability (in $\log_2$) that G at the first position was emitted by state **H**

$p_H(G,1) = -1 -1.737 = -2.737$

Probability (in $\log_2$) that G at the first position was emitted by state **L**

$p_L(G,1) = -1 -2.322 = -3.322$

Decoding (Viterbi algorithm)



GGCACTGAA

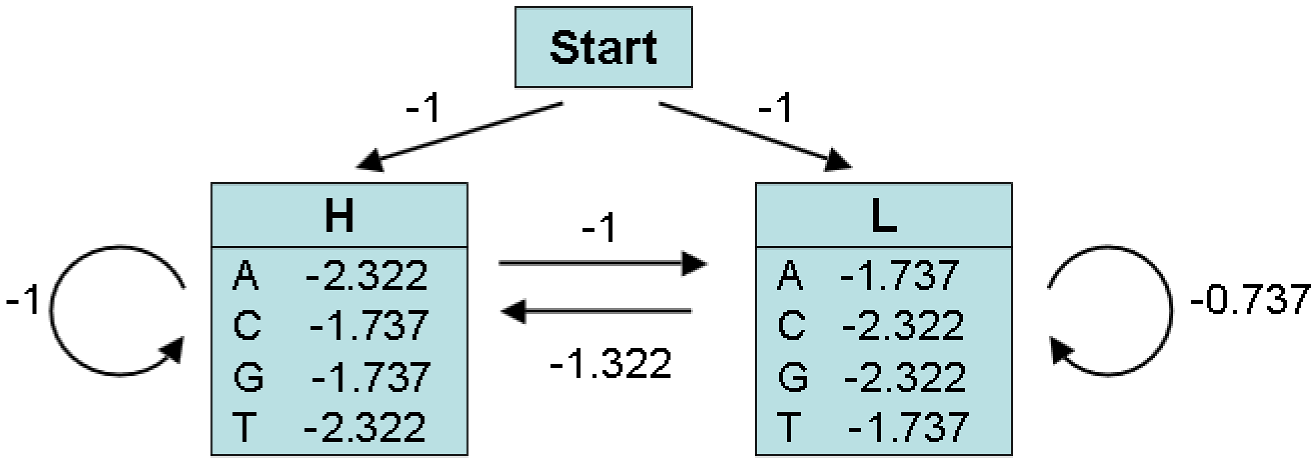
Probability (in $\log_2$) that G at the 2nd position was emitted by state **H**

$$p_H(G,2) = -1.737 + \max(p_H(G,1)+p_{HH},\ p_L(G,1)+p_{LH})$$
$$= -1.737 + \max(-2.737 - 1,\ -3.322 - 1.322)$$
$$= -5.474 \text{ (obtained from } p_H(G,1))$$

Probability (in $\log_2$) that G at the 2nd position was emitted by state **L**

$$p_L(G,2) = -2.322 + \max(p_H(G,1)+p_{HL},\ p_L(G,1)+p_{LL})$$
$$= -2.322 + \max(-2.737 - 1.322,\ -3.322 - 0.737)$$
$$= -6.059 \text{ (obtained from } p_H(G,1))$$

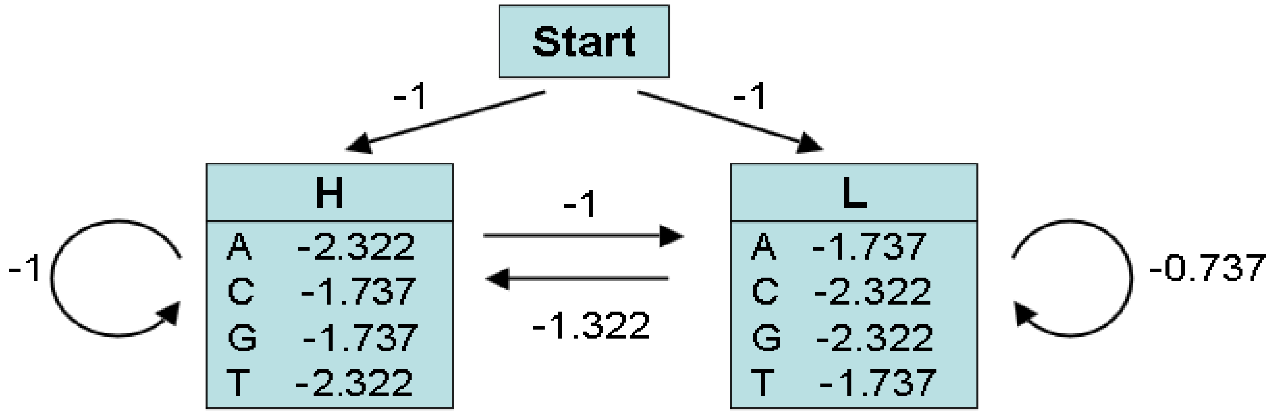
Decoding (Viterbi algorithm)



GGCACTGAA

	G	G	C	A	C	T	G	A	A
H	-2.73	-5.47	-8.21	-11.53	-14.01	...			-25.65
L	-3.32	-6.06	-8.79	-10.94	-14.01	...			-24.49

Decoding (Viterbi algorithm)



back-tracking
(= finding the path which corresponds to the highest probability, -24.49)

GGCACTGAA

	G	G	C	A	C	T	G	A	A
H	**-2.73**	**-5.47**	**-8.21**	-11.53	-14.01	...			-25.65
L	-3.32	-6.06	-8.79	**-10.94**	-14.01	...			**-24.49**

The most probable path is: **HHHLLLLLL**

Its probability is $2^{-24.49} = 4.25E-8$
(remember that we used $\log_2(p)$)