HMM Training

Viterbi and Baum-Welch Examples

Given sequence $\boldsymbol{X}$ and model $\boldsymbol{\theta}$, each cell in … represents P(…)

Viterbi Matrix

optimal state path $\boldsymbol{\pi}$ going to state k at position i, from the start of $\boldsymbol{X}$

Forward Matrix

*any** state path $\boldsymbol{\pi}$ going to state k at position i, from the start of $\boldsymbol{X}$ $P(\pi_i = k \mid X(..i))$

Backward Matrix

*any** state path $\boldsymbol{\pi}$ going from state k at position i, to the end of $\boldsymbol{X}$ $P(\pi_i = k \mid X(i..))$

Forward × Backward Matrix

*any** state path $\boldsymbol{\pi}$ going through state k at position i, in $\boldsymbol{X}$ $P(\pi_i = k \mid X)$

** : values in matrix cells are "weighted" by the probabilities of contributing state paths*

Expectation == Counting

Given model $\boldsymbol{\theta}$ *(prior **A** and **E**)*, we **expect** that sequence(s) ***X*** were generated by ***A'*** transitions and ***E'*** emissions.

In other words:
*"How often do we **see** each transition and emission occur in **X**?"*

Viterbi Training

Quite straightforward!

Baum-Welch

Slightly trickier...

Viterbi Example

Suppose some prior $\boldsymbol{\theta}$ found optimal state paths $\boldsymbol{\pi}$ for the sequences $\boldsymbol{X}$:

$\boldsymbol{X_1}$ **= *ATG***
$\boldsymbol{\pi_1}$ **= *B Q1 Q1 Q1 E***

$\boldsymbol{X_2}$ **= *ACGC***
$\boldsymbol{\pi_2}$ **= *B Q1 Q2 Q2 Q1 E***

$\boldsymbol{X_3}$ **= *GG***
$\boldsymbol{\pi_3}$ **= *B Q2 Q2 E***

What are ***A'*** and ***E''***?

Viterbi Example

Suppose some prior $\boldsymbol{\theta}$ found optimal state paths $\boldsymbol{\pi}$ for the sequences $\boldsymbol{X}$:

$X_1 =$ A T G

$\pi_1 =$ B $\boxed{Q1}$ Q1 Q1 E

$X_2 =$ A C G C

$\pi_2 =$ B $\boxed{Q1}$ Q2 Q2 Q1 E

$X_3 =$ G G

$\pi_3 =$ B $\boxed{Q2}$ Q2 E

What are *A'* and *E'*?

A'	*B*	*Q1*	*Q2*	*E*
B				
Q1				
Q2				
E				

E'	*A*	*T*	*C*	*G*
B				
Q1				
Q2				
E				

Viterbi Example

Suppose some prior $\boldsymbol{\theta}$ found optimal state paths $\boldsymbol{\pi}$ for the sequences $\boldsymbol{X}$:

$X_1 = \quad A \quad T \quad G$

$\pi_1 = B\ \boxed{Q1}\ Q1\ Q1\ E$

$X_2 = \quad A \quad C \quad G \quad C$

$\pi_2 = B\ \boxed{Q1}\ Q2\ Q2\ Q1\ E$

$X_3 = \quad G \quad G$

$\pi_3 = B\ \boxed{Q2}\ Q2\ E$

What are ***A'*** and ***E'***?

A'	*B*	*Q1*	*Q2*	*E*
B	0	2	1	0
Q1				
Q2				
E				

E'	*A*	*T*	*C*	*G*
B				
Q1				
Q2				
E				

Viterbi Example

Suppose some prior $\boldsymbol{\theta}$ found optimal state paths $\boldsymbol{\pi}$ for the sequences $\boldsymbol{X}$:

$X_1 = \quad A \;\; T \;\; G$
$\pi_1 = B \; Q1 \; Q1 \; Q1 \; E$

$X_2 = \quad A \;\; C \;\; G \;\; C$
$\pi_2 = B \; Q1 \; Q2 \; Q2 \; Q1 \; E$

$X_3 = \quad G \;\; G$
$\pi_3 = B \; Q2 \; Q2 \; E$

What are *A'* and *E'*?

A'	*B*	*Q1*	*Q2*	*E*
B	0	2	1	0
Q1				
Q2				
E				

E'	*A*	*T*	*C*	*G*
B	0	0	0	0
Q1				
Q2				
E				

Viterbi Example

Suppose some prior $\boldsymbol{\theta}$ found optimal state paths $\boldsymbol{\pi}$ for the sequences $\boldsymbol{X}$:

$X_1 = \quad A \; T \; G$

$\pi_1 = B\ Q1\ \boxed{Q1}\ \boxed{Q1}\ \boxed{E}$

$X_2 = \quad A \; C \; G \; C$

$\pi_2 = B\ Q1\ \boxed{Q2}\ Q2\ Q1\ \boxed{E}$

$X_3 = \quad G \; G$

$\pi_3 = B\ Q2\ Q2\ E$

What are ***A'*** and ***E'***?

A'	*B*	*Q1*	*Q2*	*E*
B	0	2	1	0
Q1	0	2	1	2
Q2				
E				

E'	*A*	*T*	*C*	*G*
B	0	0	0	0
Q1				
Q2				
E				

Viterbi Example

Suppose some prior $\boldsymbol{\theta}$ found optimal state paths $\boldsymbol{\pi}$ for the sequences $\boldsymbol{X}$:

$X_1 =$ A T G
$\pi_1 =$ B Q1 Q1 Q1 E

$X_2 =$ A C G C
$\pi_2 =$ B Q1 Q2 Q2 Q1 E

$X_3 =$ G G
$\pi_3 =$ B Q2 Q2 E

What are *A'* and *E'*?

A'	*B*	*Q1*	*Q2*	*E*
B	0	2	1	0
Q1	0	2	1	2
Q2				
E				

E'	*A*	*T*	*C*	*G*
B	0	0	0	0
Q1	2	1	1	1
Q2				
E				

Viterbi Example

Suppose some prior $\boldsymbol{\theta}$ found optimal state paths $\boldsymbol{\pi}$ for the sequences $\boldsymbol{X}$:

$X_1 =$ *A T G*
$\pi_1 =$ **B Q1 Q1 Q1 E**

$X_2 =$ *A C G C*
$\pi_2 =$ **B Q1 Q2 Q2 Q1 E**

$X_3 =$ *G G*
$\pi_3 =$ **B Q2 Q2 E**

What are ***A'*** and ***E'***?

A'	*B*	*Q1*	*Q2*	*E*
B	0	2	1	0
Q1	0	2	1	2
Q2	0	1	2	1
E				

E'	*A*	*T*	*C*	*G*
B	0	0	0	0
Q1	2	1	1	1
Q2	0	0	1	3
E				

Viterbi Example

Suppose some prior $\boldsymbol{\theta}$ found optimal state paths $\boldsymbol{\pi}$ for the sequences $\boldsymbol{X}$:

X_1 = A T G
π_1 = B Q1 Q1 Q1 E

X_2 = A C G C
π_2 = B Q1 Q2 Q2 Q1 E

X_3 = G G
π_3 = B Q2 Q2 E

What are ***A'*** and ***E'***?

A'	*B*	*Q1*	*Q2*	*E*
B	0	2	1	0
Q1	0	2	1	2
Q2	0	1	2	1
E	0	0	0	0

E'	*A*	*T*	*C*	*G*
B	0	0	0	0
Q1	2	1	1	1
Q2	0	0	1	3
E	0	0	0	0

Expectation == Counting

Given model $\boldsymbol{\theta}$ *(prior **A** and **E**)*, we **expect** that sequence(s) ***X*** were generated by ***A'*** transitions and ***E'*** emissions.

Viterbi Example

Divide our *A'* and *E'* counts by their respective row sums:

A'	*B*	*Q1*	*Q2*	*E*
B	0	2	1	0
Q1	0	2	1	2
Q2	0	1	2	1
~~*E*~~	~~0~~	~~0~~	~~0~~	~~0~~

E'	*A*	*T*	*C*	*G*
~~*B*~~	~~0~~	~~0~~	~~0~~	~~0~~
Q1	2	1	1	1
Q2	0	0	1	3
~~*E*~~	~~0~~	~~0~~	~~0~~	~~0~~

Viterbi Example

Divide our *A'* and *E'* counts by their respective row sums:

Posterior Transition and Emission probabilities!

a	*B*	*Q1*	*Q2*	*E*
B	0	⅔ 0.666	⅓ 0.333	0
Q1	0	⅖ 0.4	⅕ 0.2	⅖ 0.4
Q2	0	¼ 0.25	½ 0.5	¼ 0.25
E	0	0	0	0

e	*A*	*T*	*C*	*G*
Q1	⅖ 0.4	⅕ 0.2	⅕ 0.2	⅕ 0.2
Q2	0	0	¼ 0.25	¾ 0.75

… can be **priors** for the next iteration!

Baum-Welch Example

Given this $\boldsymbol{X}$ and $\boldsymbol{\theta}$, what are $\boldsymbol{A'}$ and $\boldsymbol{E'}$?

X_1 = **CCH**

X_2 = **CHH**

X_3 = **HPC**

A	*B*	*L*	*D*	*E*
B	0	0.5	0.5	0
L	0	0.7	0.2	0.1
D	0	0.2	0.7	0.1
E	0	0	0	0

E	*H*	*P*	*C*
L	0.5	0	0.5
D	0	0.5	0.5

Baum-Welch Example

$$E_k(b) = \sum_j \frac{1}{P(x^j)} \sum_{\{i|x_i^j=b\}} \boxed{f_k^j(i)}\boxed{b_k^j(i)}$$

Emission counts for X_1

Forward

i	0	1	2	3	4
X	-	C	C	H	-
B	1	0	0	0	0
L	0	0.25	0.112	0.0506	0
D	0	0.25	0.112	0	0
E	0	0	0	0	0.00506

Backward

i	0	1	2	3	4
X	-	C	C	H	-
B	0.00506	0.012	0.025	0	0
L	0.00534	0.0132	0.035	0.1	0
D	0.00377	0.007	0.01	0.1	0
E	0	0	0	0	1*

* Initialized to 0 by Durbin

Baum-Welch Example

$$E_k(b) = \sum_j \frac{1}{P(x^j)} \sum_{\{i|x_i^j=b\}} \boxed{f_k^j(i) b_k^j(i)}$$

Emission counts for X_1

Forward × Backward

i	0	1	2	3	4
X	-	C	C	H	-
B	0.00506	0	0	0	0
L	0	0.0033	0.00392	0.00506	0
D	0	0.00175	0.00112	0	0
E	0	0	0	0	0.00506

Baum-Welch Example

$$E_k(b) = \sum_j \frac{1}{\boxed{P(x^j)}} \sum_{\{i|x_i^j=b\}} \boxed{f_k^j(i) b_k^j(i)}$$

Emission counts for X_1

Forward × Backward

i	0	1	2	3	4
X	-	C	C	H	-
B	0.00506	0	0	0	0
L	0	0.0033	0.00392	0.00506	0
D	0	0.00175	0.00112	0	0
E	0	0	0	0	0.00506

Forward × Backward / P(X_1)

i	0	1	2	3	4
X	-	C	C	H	-
B	1	0	0	0	0
L	0	0.667	0.775	1	0
D	0	0.333	0.221	0	0
E	0	0	0	0	1

Baum-Welch Example

$$E_k(b) = \sum_j \frac{1}{P(x^j)} \sum_{\{i|x_i^j=b\}} f_k^j(i) b_k^j(i)$$

Emission counts for $X_{\boxed{1}}$

E'	H	P	C
L	$\boxed{1}$		
D			

k = **L**
b = **H**
j = **1**

Forward × Backward / P(X_1)

i	0	1	2	3	4
X	-	C	C	$\boxed{\text{H}}$	-
B	1	0	0	0	0
$\boxed{\text{L}}$	0	0.667	0.775	1	0
D	0	0.333	0.221	0	0
E	0	0	0	0	1

Baum-Welch Example

$$E_k(b) = \sum_j \frac{1}{P(x^j)} \sum_{\{i|x_i^j=b\}} f_k^j(i) b_k^j(i)$$

Emission counts for $X_{\boxed{1}}$

E'	**H**	**P**	**C**
L	1	$\boxed{0}$	
D			

k = **L**
b = **P**
j = **1**

Forward × Backward / P(X_1)

i	**0**	**1**	**2**	**3**	**4**
X	**-**	**C**	**C**	**H**	**-**
B	1	0	0	0	0
$\boxed{\textbf{L}}$	0	0.667	0.775	1	0
D	0	0.333	0.221	0	0
E	0	0	0	0	1

Baum-Welch Example

$$E_k(b) = \sum_j \frac{1}{P(x^j)} \sum_{\{i|x_i^j=b\}} f_k^j(i) b_k^j(i)$$

Emission counts for $X_{\boxed{1}}$

E'	H	P	C
L	1	0	$\boxed{1.441}$
D			

k = **L**
b = **C**
j = **1**

Forward × Backward / P(X_1)

i	0	1	2	3	4
X	-	C	C	H	-
B	1	0	0	0	0
L	0	0.667	0.775	1	0
D	0	0.333	0.221	0	0
E	0	0	0	0	1

Baum-Welch Example

$$E_k(b) = \sum_j \frac{1}{P(x^j)} \sum_{\{i|x_i^j=b\}} f_k^j(i) b_k^j(i)$$

Emission counts for $X_{\boxed{1}}$

E'	H	P	C
L	1	0	1.441
D	$\boxed{0}$		

k = **D**
b = **H**
j = **1**

Forward × Backward / P(X_1)

i	0	1	2	3	4
X	-	C	C	$\boxed{\text{H}}$	-
B	1	0	0	0	0
L	0	0.667	0.775	1	0
$\boxed{\text{D}}$	0	0.333	0.221	0	0
E	0	0	0	0	1

Baum-Welch Example

$$E_k(b) = \sum_j \frac{1}{P(x^j)} \sum_{\{i|x_i^j=b\}} f_k^j(i) b_k^j(i)$$

Emission counts for $\boldsymbol{X}_{\boxed{1}}$

E'	H	P	C
L	1	0	1.441
D	0	$\boxed{0}$	

k = **D**
b = **P**
j = **1**

Forward × Backward / P(*X*$_1$)

i	**0**	**1**	**2**	**3**	**4**
X	**-**	**C**	**C**	**H**	**-**
B	1	0	0	0	0
L	0	0.667	0.775	1	0
$\boxed{\textbf{D}}$	0	0.333	0.221	0	0
E	0	0	0	0	1

Baum-Welch Example

$$E_k(b) = \sum_j \frac{1}{P(x^j)} \sum_{\{i|x_i^j=b\}} f_k^j(i) b_k^j(i)$$

Emission counts for $X_{\boxed{1}}$

E'	H	P	C
L	1	0	1.441
D	0	0	$\boxed{0.554}$

k = **D**
b = **C**
j = **1**

(Actual values differ slightly due to rounding errors.)

Forward × Backward / P(X_1)

i	0	1	2	3	4
X	**-**	$\boxed{\textbf{C}}$	$\boxed{\textbf{C}}$	**H**	**-**
B	1	0	0	0	0
L	0	0.667	0.775	1	0
$\boxed{\textbf{D}}$	0	0.333	0.221	0	0
E	0	0	0	0	1

Baum-Welch Example

Transition counts for $\boldsymbol{X_1}$

$$P(\pi_i = k, \pi_{i+1} = l \mid x, \theta) = \frac{\boxed{f_k(i)}\boxed{a_{kl}}\boxed{e_l(x_{i+1})}\boxed{b_l(i+1)}}{\boxed{P(x)}}$$

To get A', iterate over every possible combination of i, k and l and add up the resulting $P(\pi_i = k, \pi_{i+1} = l)$

Do this for every sequence X_j in X. Counts add up together!

Baum-Welch Example

Transition counts for X_1

$$P(\pi_i = k, \pi_{i+1} = l \mid x, \theta) = \frac{\boxed{f_k(i)}\boxed{a_{kl}}\boxed{e_l(x_{i+1})}\boxed{b_l(i+1)}}{\boxed{P(x)}}$$

Forward

i	0	1	2	3	4
X	-	C	C	H	-
B	1	0	0	0	0
L	0	0.25	0.112	0.0506	0
D	0	0.25	0.112	0	0
E	0	0	0	0	0.00506

Backward

i	0	1	2	3	4
X	-	C	C	H	-
B	0.00506	0.012	0.025	0	0
L	0.00534	0.0132	0.035	0.1	0
D	0.00377	0.007	0.01	0.1	0
E	0	0	0	0	1*

Prior θ

A	*B*	*L*	*D*	*E*
B	0	0.5	0.5	0
L	0	0.7	0.2	0.1
D	0	0.2	0.7	0.1
E	0	0	0	0

E	*H*	*P*	*C*
L	0.5	0	0.5
D	0	0.5	0.5

For example, with *i* = 1, *k* = D and *l* = L …

$P(\pi_1 = D, \pi_2 = L) = 0.25 \times 0.2 \times 0.5 \times 0.035 \,/\, 0.00506 = 0.173$

Baum-Welch Example

Transition counts for X_1

$$P(\pi_i = k, \pi_{i+1} = l | x, \theta) = \frac{\boxed{f_k(i)}\boxed{a_{kl}}\boxed{e_l(x_{i+1})}\boxed{b_l(i+1)}}{\boxed{P(x)}}$$

Forward

i	0	1	2	3	4
X	-	C	C	H	-
B	1	0	0	0	0
L	0	0.25	0.112	0.0506	0
D	0	0.25	0.112	0	0
E	0	0	0	0	0.00506

Backward

i	0	1	2	3	4
X	-	C	C	H	-
B	0.00506	0.012	0.025	0	0
L	0.00534	0.0132	0.035	0.1	0
D	0.00377	0.007	0.01	0.1	0
E	0	0	0	0	1*

Prior θ

A	*B*	*L*	*D*	*E*
B	0	0.5	0.5	0
L	0	0.7	0.2	0.1
D	0	0.2	0.7	0.1
E	0	0	0	0

E	*H*	*P*	*C*
L	0.5	0	0.5
D	0	0.5	0.5

For every other ***i***, at least one factor = 0...

$P(\pi_1 = D, \pi_2 = L) = 0.25 \times 0.2 \times 0.5 \times 0.035 / 0.00506 = 0.173$

Across all positions ***i*** in ***X₁***, we **count** a total of

0.173 + 0.222 = **0.395** transitions from **D** to **L**.

Baum-Welch Example

Given $\boldsymbol{X}$ and $\boldsymbol{\theta}$, what are A' and E'?

E'_1	H	P	C
L	1	0	1.432
D	0	0	0.568

E'_2	H	P	C
L	2	0	0.778
D	0	0	0.222

E'_3	H	P	C
L	1	0	0.222
D	0	1	0.778

A'_1	B	L	D	E
B	0	0.654	0.346	0
L	0	1.383	0.049	1
D	0	0.395	0.173	0
E	0	0	0	0

This is the 0.395 we found before!

A'_3	B	L	D	E
B	0	1	0	0
L	0	0	1	0.222
D	0	0.222	0.778	0.778
E	0	0	0	0

A'_2	B	L	D	E
B	0	0.778	0.222	0
L	0	1.778	0	1
D	0	0.222	0	0
E	0	0	0	0

Baum-Welch Example

Given ***X*** and prior ***θ***, what is posterior ***θ*** (***a*** and ***e***)?

*Add up A' and E' across all **j** in **X** and divide by the row sums again!*

Prior *Expectation* *Maximization*

X₁ **= CCH**

X₂ **= CHH**

X₃ **= HPC**

A	**B**	**L**	**D**	**E**
B	0	0.5	0.5	0
L	0	0.7	0.2	0.1
D	0	0.2	0.7	0.1
E	0	0	0	0

E	**H**	**P**	**C**
L	0.5	0	0.5
D	0	0.5	0.5