Machine Learning

4771

Instructor: Tony Jebara

Topic 20

- HMMs with Evidence
- HMM Collect
- HMM Evaluate
- HMM Distribute
- HMM Decode
- HMM Parameter Learning via JTA & EM

Collect

$\psi(q_t, q_{t-1}) = p(q_t \mid q_{t-1})$ Initialization

$\psi^{*}(q_t, q_{t-1}) = p(q_{t-1}, q_t)$ After collection

$\psi^{**}(q_t, q_{t-1}) = p(q_{t-1}, q_t)$ After Distribution.

HMMs: JTA with Evidence

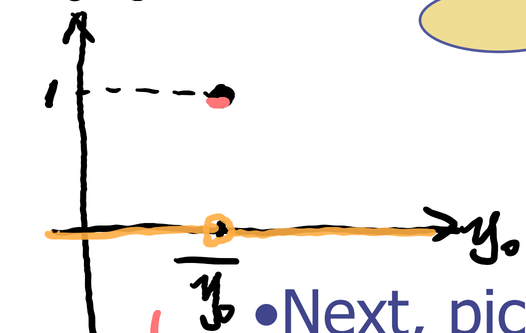
$\bar{y}$: It's observed.

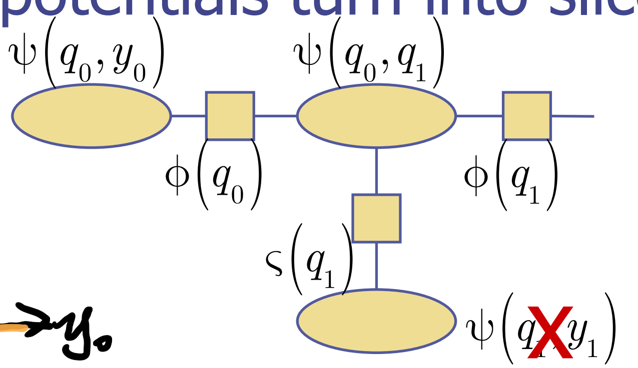
- If y sequence is observed (in problems 1,2,3) get evidence:

$$p(q,\bar{y}) = p(q_0)\prod_{t=1}^{T} p(q_t \mid q_{t-1})\prod_{t=0}^{T} p(\bar{y}_t \mid q_t)$$

- The potentials turn into slices:

$\delta(y-\bar{y}_0)$ Function:





$\longrightarrow \psi(q_1, \bar{y}_1)$

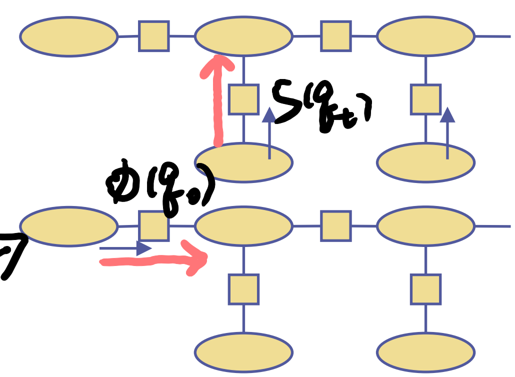
$\varsigma^{*}(q_t) \cancel{=} \sum_{y_t} \psi(q_t, y_t)$ y_t has been observed.

$\varsigma^{*}(q_t) = \psi(q_t, \bar{y}_t) = p(\bar{y}_t \mid q_t)$

- Next, pick a root, for example *rightmost* one: $\psi(q_{T-1}, q_T)$

- Collect all zeta separators bottom up:

$$\varsigma^{*}(q_t) = \psi(q_t, \bar{y}_t) = p(\bar{y}_t \mid q_t)$$



- Collect leftmost phi separator to the right:

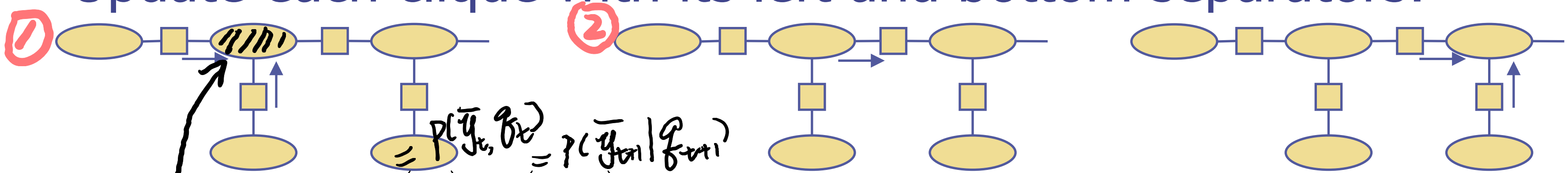
$$\phi^{*}(q_0) = \sum_{y_0} \psi(q_0, \bar{y}_0)\,\delta(y_0 - \bar{y}_0) = p(\bar{y}_0, q_0)$$

$= p(q_0) \cdot p(\bar{y}_0 \mid q_0) \longrightarrow$

collect

HMMs: Collect with Evidence

- Now, we will collect (*) along the backbone left to right
- Update each clique with its left and bottom separators:



1 $$\psi^*(q_t, q_{t+1}) = \frac{\phi^*(q_t)}{1}\frac{\varsigma^*(q_{t+1})}{1}\psi(q_t, q_{t+1}) = \phi^*(q_t)\, p(\bar{y}_{t+1} \mid q_{t+1})\, \alpha_{q_t, q_{t+1}}$$

$\psi(q_t, q_{t+1}) = p(q_{t+1} | q_t) = \alpha_{q_t, q_{t+1}}$

2 $$\phi^*(q_{t+1}) = \sum_{q_t} \psi^*(q_t, q_{t+1}) = \sum_{q_t} \phi^*(q_t)\, p(\bar{y}_{t+1} \mid q_{t+1})\, \alpha_{q_t, q_{t+1}}$$

- Keep going along chain until right most node
- Note: above formula for phi is recursive, could use as is.
- Property: recall we had $\phi^*(q_0) = p(\bar{y}_0, q_0)$

$$\phi^*(q_1) = \sum_{q_0} p(\bar{y}_0, q_0)\, p(\bar{y}_1 \mid q_1)\, p(q_1 \mid q_0) = p(\bar{y}_0, \bar{y}_1, q_1)$$

$$\phi^*(q_2) = \sum_{q_1} p(\bar{y}_0, \bar{y}_1, q_1)\, p(\bar{y}_2 \mid q_2)\, p(q_2 \mid q_1) = p(\bar{y}_0, \bar{y}_1, \bar{y}_2, q_2)$$

$$\phi^*(q_{t+1}) = \sum_{q_t} p(\bar{y}_0, \ldots, \bar{y}_t, q_t)\, p(\bar{y}_{t+1} \mid q_{t+1})\, p(q_{t+1} \mid q_t) = p(\bar{y}_0, \ldots, \bar{y}_{t+1}, q_{t+1})$$

Collect

Tony Jebara, Columbia University

HMMs: Evaluate with Evidence

- Say we are solving the first HMM problem:

Collect = 1) Evaluate: given $y_0,\dots,y_T$ & θ compute $p(y_0,\dots,y_T|\theta)$

- If we want to compute the likelihood, we are already done!
- We really just need to do collect (not even distribute).

$\phi^*(q_t) = p(\bar{y}_{t+1}, q_{t+1}|q_t)$

- From previous slide we had: $\psi^*(q_t, q_{t+1}) = \phi^*(q_t)\cdot p(\bar{y}_{t+1}|q_{t+1})\cdot p(q_{t+1}|q_t)$

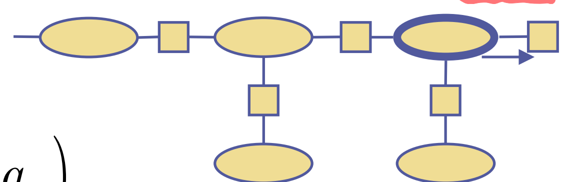
phi ⇒ $$\phi^*\left(q_{t+1}\right) = \sum_{q_t} p\left(\bar{y}_0,\dots,\bar{y}_t,q_t\right)p\left(\bar{y}_{t+1}\mid q_{t+1}\right)p\left(q_{t+1}\mid q_t\right) = p\left(\bar{y}_0,\dots,\bar{y}_{t+1},q_{t+1}\right)$$

p(Evidence)

- Collect 'til root (rightmost node): $\psi^*\left(q_{T-1},q_T\right) = p\left(\bar{y}_0,\dots,\bar{y}_T,q_{T-1},q_T\right) = \phi(q_{T-1})\cdot p(\bar{y}_T, q_T|q_{T-1})$

psi

its normalizer is p(EVIDENCE)!



hypothetical

φ: phi

ψ: psi

phi

Or use hypothetical $\phi^*\left(q_T\right) = p\left(\bar{y}_0,\dots,\bar{y}_T,q_T\right)$

- Can compute the likelihood just by marginalizing this phi

likelihood

$$p\left(\bar{y}_0,\dots,\bar{y}_T\right) = \sum_{q_T} p\left(\bar{y}_0,\dots,\bar{y}_T,q_T\right) = \sum_{q_T}\phi^*\left(q_T\right)$$

- So, adding up the entries in last ϕ^* gives us the likelihood

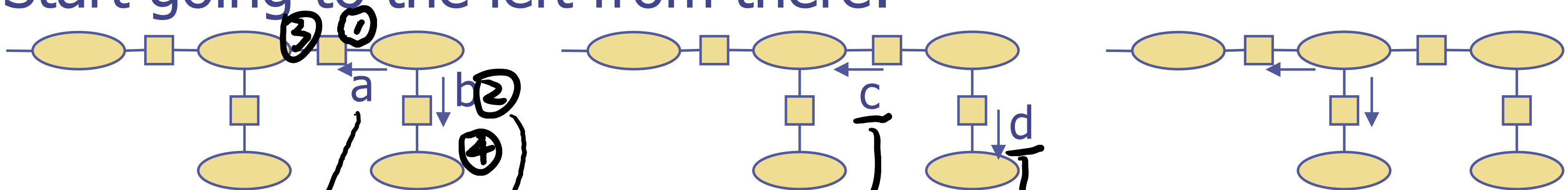
Distribute.

HMMs: Distribute with Evidence

- Back to collecting… say just finished collecting to the root with our last update formula: $= p(\bar{y}_0, \dots, \bar{y}_T, q_{T-1}, q_T)$

$$\psi^*\left(q_{T-1}, q_T\right) = \frac{\phi^*\left(q_{T-1}\right)}{1} \frac{\varsigma^*\left(q_T\right)}{1} \psi\left(q_{T-1}, q_T\right) = \phi^*\left(q_{T-1}\right) p\left(\bar{y}_T \mid q_T\right) \alpha_{q_{T-1}, q_T}$$

- Now, we distribute (**) along the backbone right to left
- Have first ** for root (stays the same): $\psi^{**}\left(q_{T-1}, q_T\right) = \psi^*\left(q_{T-1}, q_T\right)$
- Start going to the left from there:



for t=T-1 to 0

① a) $\phi^{**}\left(q_t\right) = \sum_{q_{t+1}} \psi^{**}\left(q_t, q_{t+1}\right)$

② b) $\varsigma^{**}\left(q_{t+1}\right) = \sum_{q_t} \psi^{**}\left(q_t, q_{t+1}\right)$

③ c) $\psi^{**}\left(q_t, q_{t+1}\right) = \frac{\phi^{**}\left(q_{t+1}\right)}{\phi^*\left(q_{t+1}\right)} \psi^*\left(q_t, q_{t+1}\right)$

④ d) $\psi^{**}\left(y_t, q_t\right) = \frac{\varsigma^{**}\left(q_t\right)}{\varsigma^*\left(q_t\right)} \psi\left(y_t, q_t\right)$

HMM Example

You are given the parameters of a 2-state HMM. You observed the input sequence AB (from a 2-symbol alphabet A or B). In other words, you observe two symbols from your finite state machine, A and then B. Using the junction tree algorithm, ① evaluate the likelihood of this data p(y) given your HMM and its parameters. Also ② compute (for decoding) the individual marginals of the states after the evidence from this sequence is observed: $p(q_0|y)$ and $p(q_1|y)$. The parameters for the HMM are provided below. They are the initial state prior $p(q_0)$ π, the state transition matrix given by $p(q_t| q_{t-1})$ a, and the emission matrix $p(y_t| q_t)$ η respectively.

① Evaluate

② Decoding.

$$\pi = p(q_0) = \begin{array}{cc} 1 & 2 \\ \end{array} \\ \begin{bmatrix} 1/3 & 2/3 \end{bmatrix}$$

$$a^T = p(q_t | q_{t-1}) = \begin{array}{c|cc} & 1 & 2 \\ 1 & 3/4 & 1/2 \\ 2 & 1/4 & 1/2 \end{array} \qquad \eta^T = p(y_t | q_t) = \begin{array}{c|cc} & 1 & 2 \\ A & 1/2 & 1/3 \\ B & 1/2 & 2/3 \end{array}$$

Given $p(\bar{y}_t | q_t)$

Step1: Initialization:

$\psi(q_0,y_0)=p(q_0)\cdot p(y_0|q_0)$ $\Phi(q_t)=1$

$\psi(q_{t+1},q_t)=p(q_{t+1}|q_t)=a_{q_t,q_{t+1}}$ $Z=1$ $S(q_t)=1$

$p(q_t,y_t)=p(y_t|q_t)$

↓

All are given

Step2. Evaluate: collect:

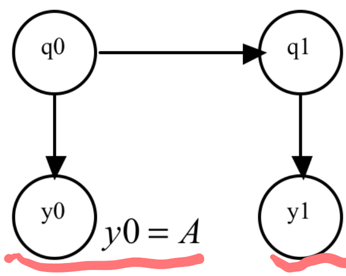
$\psi^*(q_t,q_{t+1}) = \frac{\Phi^*(q_t)}{1}\cdot\frac{S^*(q_{t+1})}{1}\psi(q_t,q_{t+1}) = \Phi^*(q_t)\cdot p(\bar{y}_{t+1}|q_{t+1})\cdot p(q_{t+1}|q_t)$

$\Phi^*(q_{t+1}) = \sum_{q_t}\psi^*(q_t,q_{t+1}) = \sum_q p(\bar{y}_0,\bar{y}_1,\bar{y}_2,\dots,q_t)\cdot p(\bar{y}_{t+1}|q_{t+1})\cdot p(q_{t+1}|q_t) = p(\bar{y}_0,\bar{y}_1,\dots,\bar{y}_{t+1},q_{t+1})$

HMM Example

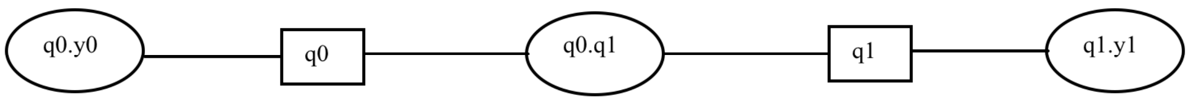
collect: $\psi(q_0,y_0)=p(q_0)\cdot p(y_0|q_0)=p(q_0,y_0)$

$\phi(q_0)=$



$y0 = A$ $y1 = B$ observed

Step1: Initialized Junction Tree



$y0 = A....B$		
$q0 = 1$	1/6	1/6
$q0 = 2$	2/9	4/9

$q0 = 1...2$	
1	1

$q1 = 1...2$		
$q0 = 1$	3/4	1/4
$q0 = 2$	1/2	1/2

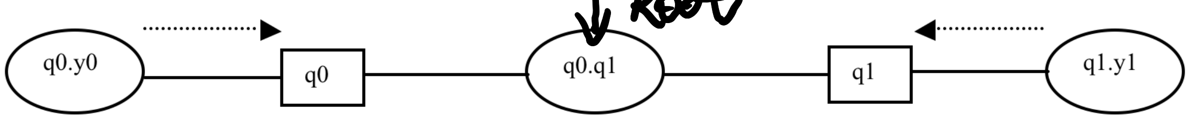
$q1 = 1...2$	
1	1

$y1 = A...B$		
$q1 = 1$	1/2	1/2
$q1 = 2$	1/3	2/3

Step2:

↓ Root

find the right position!



$y0 = A....B$		
$q0 = 1$	1/6	
$q0 = 2$	2/9	

$q0 = 1...2$	
1/6	2/9

$\begin{bmatrix} \frac{1}{6} \\ \frac{2}{9} \end{bmatrix}$

$q1 = 1...2$		
$q0 = 1$	3/4	1/4
$q0 = 2$	1/2	1/2

$q1 = 1...2$	
1/2	2/3

$y1 = A...B$		
$q1 = 1$		1/2
$q1 = 2$		2/3

$$\begin{bmatrix} \frac{3}{4}\times\frac{1}{6}\times\frac{1}{2} & \frac{1}{4}\times\frac{2}{3}\times\frac{1}{6} \\ \frac{1}{2}\times\frac{2}{9}\times\frac{1}{2} & \frac{1}{2}\times\frac{2}{9}\times\frac{2}{3} \end{bmatrix}$$

clique (q_0,q_1)

$$\begin{array}{cc|l} q_1=1 & q_1=2 & \\ \hline \frac{3}{4} & \frac{1}{4} & q_0=1 \\ \frac{1}{2} & \frac{1}{2} & q_0=2 \end{array} \qquad \psi(q_0,q_1) \Rightarrow$$

$$\Phi^*(q_0)=\begin{bmatrix}\frac{1}{6}\\ \frac{2}{9}\end{bmatrix}\begin{matrix}q_0=1\\ q_0=2\end{matrix}$$

$\psi(q_0,q_1)$ After ① update from left to right

$$=\frac{\Phi^*(q_0)}{\Phi(q_0)=1}\cdot\psi(q_0,q_1)=\begin{bmatrix}\frac{1}{6}\times\frac{3}{4} & \frac{1}{4}\times\frac{1}{6}\\ \frac{1}{2}\times\frac{2}{9} & \frac{1}{2}\times\frac{2}{9}\end{bmatrix}\begin{matrix}q_0=1\\ q_0=2\end{matrix} \quad (q_1=1,\ q_1=2)$$

$\Downarrow$

$\psi^{**}(q_0,q_1)$ After ② update from right to left

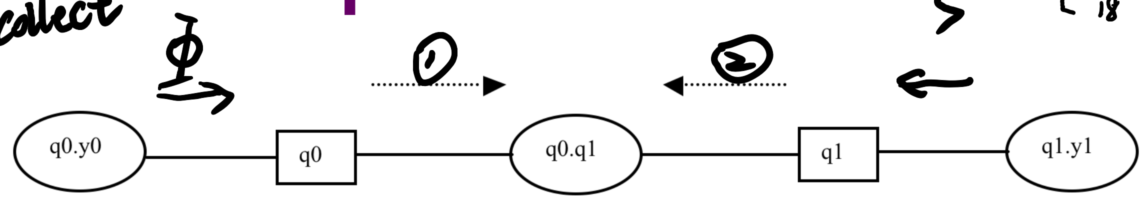
$$=\frac{\Phi^*(q_1)}{\Phi(q_1)=1}\cdot\psi(q_0,q_1)=\begin{bmatrix}\frac{1}{2}\times\frac{1}{6}\times\frac{3}{4} & \frac{1}{4}\times\frac{1}{6}\times\frac{2}{3}\\ \frac{1}{2}\times\frac{1}{2}\times\frac{2}{9} & \frac{1}{2}\times\frac{2}{9}\times\frac{2}{3}\end{bmatrix}=\begin{bmatrix}\frac{1}{16} & \frac{1}{36}\\ \frac{1}{18} & \frac{2}{27}\end{bmatrix}$$

Not Algebra!

HMM Example

Tony Jebara, Columbia University

step2: collect



$y0 = A....B$		
$q0=1$	1/6	
$q0=2$	2/9	

$q0 = 1...2$	
1/6	2/9

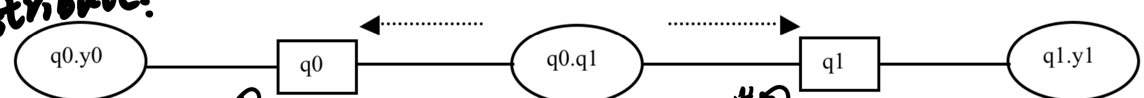
	$q1 = 1...2$	
$q0=1$	1/16	1/36
$q0=2$	1/18	2/27

$q1 = 1...2$	
1/2	2/3

$y1 = A...B$		
$q1=1$		1/2
$q1=2$		2/3

$\left[\frac{1}{2} \quad \frac{2}{3}\right]$ $q_1=1$ $q_1=2$

step3. Distribute.



$y0 = A....B$		
$q0=1$	1/6	
$q0=2$	2/9	

#①

$q0 = 1...2$	
13/144	7/54
$\frac{1}{16}+\frac{1}{36}$	$\frac{1}{18}+\frac{2}{27}$

marginalize

	$q1 = 1...2$	
$q0=1$	1/16	1/36
$q0=2$	1/18	2/27

#②

$q1 = 1...2$	
17/144	11/108
$\frac{1}{16}+\frac{1}{18}$	$\frac{1}{36}+\frac{2}{27}$

marginalize

$y1 = A...B$		
$q1=1$		1/2
$q1=2$		2/3

- Evaluate:

So the likelihood $p(y)=\frac{13}{144}+\frac{7}{54}=\frac{1}{16}+\frac{1}{18}+\frac{1}{36}+\frac{2}{27}=\frac{17}{144}+\frac{11}{108}=\frac{95}{432}=0.2199$

- Decode:

#① $p(q_0=1\,|\,y)=\frac{13/144}{13/144+7/54}=\frac{39}{95}$, $p(q_0=2\,|\,y)=\frac{7/54}{13/144+7/54}=\frac{56}{95}$

#② $p(q_1=1\,|\,y)=\frac{17/144}{17/144+11/108}=\frac{51}{95}$, $p(q_1=2\,|\,y)=\frac{11/108}{17/144+11/108}=\frac{44}{95}$

HMMs: Marginals & MaxDecoding

- Now that JTA is finished, we have the following:

$$\phi^{**}\left(q_t\right) \propto p\left(q_t \mid \bar{y}_1,\ldots,\bar{y}_T\right) \qquad \varsigma^{**}\left(q_{t+1}\right) \propto p\left(q_{t+1} \mid \bar{y}_1,\ldots,\bar{y}_T\right)$$

$$\psi^{**}\left(q_t,q_{t+1}\right) \propto p\left(q_t,q_{t+1} \mid \bar{y}_1,\ldots,\bar{y}_T\right)$$

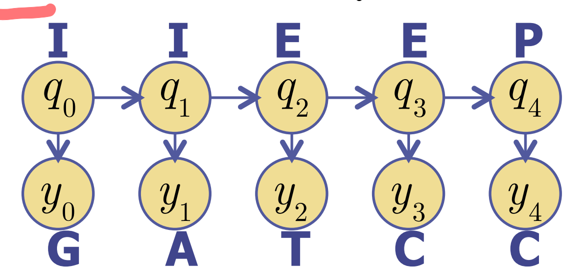
- The separators define a distribution over the hidden states
- This gives the probability the DNA symbol y_t was $q_t=\{I,E,P\}$
- We've done 2) Decode: given $y_0,\ldots,y_T$ & θ find $p(q_0),\ldots,p(q_T)$

- Can also do 2) Decode: given $y_0,\ldots,y_T$ & θ find $q_0,\ldots,q_T$
- We can also decode to find the most likely path $q_0 \ldots q_T$
- Here, we use the ArgMax JTA algorithm

class 18x . P11/P12

- Run JTA but replace sums with max
- Then, find biggest entry in separators:

$$\hat{q}_t = \arg\max_{q_t} \phi^{**}\left(q_t\right) \quad \forall t = 0\ldots T$$



Tony Jebara, Columbia University

HMMs: EM Learning

- Finally 3) Max Likelihood: given $y_0,\ldots,y_T$ learn parameters θ
- Recall max likelihood: $\underline{\hat{\theta} = \arg\max_\theta \log p(\bar{y} \mid \theta)}$
- If observe q, it's easy to maximize the *complete* likelihood:

$p(q_0) = \pi$
$p(q_t \mid q_{t-1}) = a$
$p(\bar{y}_t \mid q_t) = \eta$

$$l(\theta) = \log\left(p\left(q,y\right)\right)$$

$$= \log\left(p\left(q_0\right)\prod_{t=1}^{T} p\left(q_t \mid q_{t-1}\right)\prod_{t=0}^{T} p\left(\bar{y}_t \mid q_t\right)\right)$$

$$= \log p\left(\overset{=\pi}{q_0}\right) + \sum_{t=1}^{T}\log p\left(q_t \mid q_{t-1}\right) + \sum_{t=0}^{T}\log p\left(\bar{y}_t \mid q_t\right) \Rightarrow \text{Lecture 19 Page 10.}$$

$f(\pi) = \sum_{i=1}^{M} q_0^i \log \pi_i + \gamma \cdot \left(\sum_{i=1}^{M} \pi_i - 1\right)$

$$= \log\prod_{i=1}^{M}\left[\underline{\pi_i}\right]^{q_0^i} + \sum_{t=1}^{T}\log\prod_{i=1}^{M}\prod_{j=1}^{M}\left[\underline{\alpha_{ij}}\right]^{q_{t-1}^i q_t^j} + \sum_{t=0}^{T}\log\prod_{i=1}^{M}\prod_{j=1}^{N}\left[\underline{\eta_{ij}}\right]^{q_t^i y_t^j}$$

$\frac{\partial f(\pi)}{\partial \pi} = 0 \Rightarrow$ Derivative to get the optimal

$$= \sum_{i=1}^{M} q_0^i \log \pi_i + \sum_{t=1}^{T}\sum_{i,j=1}^{M} q_{t-1}^i q_t^j \log \alpha_{ij} + \sum_{t=0}^{T}\sum_{i=1}^{M}\sum_{j=1}^{N} q_t^i y_t^j \log \eta_{ij}$$

marginals   marginals

Introduce Lagrange & take derivatives $\longrightarrow$ $\sum_{i=1}^{M}\pi_i = 1 \qquad \sum_{j=1}^{M}\alpha_{ij} = 1 \qquad \sum_{j=1}^{N}\eta_{ij} = 1$

$\frac{q_0^i}{\pi_i} + \gamma = 0$

$$\hat{\pi}_i = q_0^i \qquad \hat{\alpha}_{ij} = \frac{\sum_{t=0}^{T-1} q_t^i q_{t+1}^j}{\sum_{k=1}^{M}\sum_{t=0}^{T-1} q_t^i q_{t+1}^k} \qquad \hat{\eta}_{ij} = \frac{\sum_{t=0}^{T} q_t^i y_t^j}{\sum_{k=1}^{N}\sum_{t=0}^{T} q_t^i y_t^k}$$

HMMs: EM Learning

- But, we don't observe the q's, incomplete…

$$p\left(\bar{y} \mid \theta\right) = \sum_{q} p\left(q, \bar{y} \mid \theta\right) = \sum_{q_0} \cdots \sum_{q_T} p\left(q_0\right) \prod_{t=1}^{T} p\left(q_t \mid q_{t-1}\right) \prod_{t=0}^{T} p\left(\bar{y}_t \mid q_t\right)$$

- EM: Max expected complete likelihood given current p(q)

$$E\left\{l\left(\theta\right)\right\} = E_{p\left(q_0,\ldots,q_T \mid y\right)}\left\{\log p\left(q,y\right)\right\} = \boxed{\sum_{q_0} \cdots \sum_{q_T} p\left(q \mid y\right) \log p\left(q,y\right)}$$

classPX.Page5. $Q(\theta|\theta_t)$

$$= E\left\{\sum_{i=1}^{M} q_0^i \log \pi_i + \sum_{t=1}^{T} \sum_{i,j=1}^{M} q_{t-1}^i q_t^j \log \alpha_{ij} + \sum_{t=0}^{T} \sum_{i=1}^{M} \sum_{j=1}^{N} q_t^i y_t^j \log \eta_{ij}\right\}$$

$$= \sum_{i=1}^{M} E\left\{q_0^i\right\} \log \pi_i + \sum_{t=1}^{T} \sum_{i,j=1}^{M} E\left\{q_{t-1}^i q_t^j\right\} \log \alpha_{ij} + \sum_{t=0}^{T} \sum_{i=1}^{M} \sum_{j=1}^{N} E\left\{q_t^i\right\} y_t^j \log \eta_{ij}$$

- M-step is maximizing as before:

$$\hat{\pi}_i = E\left\{q_0^i\right\} \qquad \hat{\alpha}_{ij} = \frac{\sum_{t=0}^{T-1} E\left\{q_t^i q_{t+1}^j\right\}}{\sum_{k=1}^{M} \sum_{t=0}^{T-1} E\left\{q_t^i q_{t+1}^k\right\}} \qquad \hat{\eta}_{ij} = \frac{\sum_{t=0}^{T} E\left\{q_t^i\right\} y_t^j}{\sum_{k=1}^{N} \sum_{t=0}^{T} E\left\{q_t^i\right\} y_t^k}$$

- What are E{}'s?

EM

HMMs: EM Learning

- But, we don't observe the q's, incomplete...

$$p\left(\bar{y} \mid \theta\right) = \sum_{q} p\left(q, \bar{y} \mid \theta\right) = \sum_{q_0} \cdots \sum_{q_T} p\left(q_0\right) \prod_{t=1}^{T} p\left(q_t \mid q_{t-1}\right) \prod_{t=0}^{T} p\left(\bar{y}_t \mid q_t\right)$$

- EM: Max expected complete likelihood given current p(q)

$$E\left\{l\left(\theta\right)\right\} = E_{p\left(q_0,\ldots,q_T \mid y\right)}\left\{\log p\left(q,y\right)\right\} = \sum_{q_0} \cdots \sum_{q_T} p\left(q \mid y\right) \log p\left(q,y\right)$$

$$= E\left\{\sum_{i=1}^{M} q_0^i \log \pi_i + \sum_{t=1}^{T} \sum_{i,j=1}^{M} q_{t-1}^i q_t^j \log \alpha_{ij} + \sum_{t=0}^{T} \sum_{i=1}^{M} \sum_{j=1}^{N} q_t^i y_t^j \log \eta_{ij}\right\}$$

$$= \sum_{i=1}^{M} E\left\{q_0^i\right\} \log \pi_i + \sum_{t=1}^{T} \sum_{i,j=1}^{M} E\left\{q_{t-1}^i q_t^j\right\} \log \alpha_{ij} + \sum_{t=0}^{T} \sum_{i=1}^{M} \sum_{j=1}^{N} E\left\{q_t^i\right\} y_t^j \log \eta_{ij}$$

- M-step is maximizing as before:

$$\hat{\pi}_i = E\left\{q_0^i\right\} \qquad \hat{\alpha}_{ij} = \frac{\sum_{t=0}^{T-1} E\left\{q_t^i q_{t+1}^j\right\}}{\sum_{k=1}^{M} \sum_{t=0}^{T-1} E\left\{q_t^i q_{t+1}^k\right\}} \qquad \hat{\eta}_{ij} = \frac{\sum_{t=0}^{T} E\left\{q_t^i\right\} y_t^j}{\sum_{k=1}^{N} \sum_{t=0}^{T} E\left\{q_t^i\right\} y_t^k}$$

- What are E{}'s? $E_{p(x)}\left\{x^i\right\} = \sum_x p\left(x\right) x^i = \sum_x p\left(x\right) \delta\left(x = x^i\right) = p\left(x^i\right)$

HMMs: EM Learning

- But, we don't observe the q's, incomplete...

$$p\left(\bar{y} \mid \theta\right) = \sum_{q} p\left(q, \bar{y} \mid \theta\right) = \sum_{q_0} \cdots \sum_{q_T} p\left(q_0\right) \prod_{t=1}^{T} p\left(q_t \mid q_{t-1}\right) \prod_{t=0}^{T} p\left(\bar{y}_t \mid q_t\right)$$

- EM: Max expected complete likelihood given current p(q)

$$E\left\{l\left(\theta\right)\right\} = E_{p\left(q_0,\ldots,q_T \mid y\right)}\left\{\log p\left(q,y\right)\right\} = \sum_{q_0} \cdots \sum_{q_T} p\left(q \mid y\right) \log p\left(q,y\right)$$

$$= E\left\{\sum_{i=1}^{M} q_0^i \log \pi_i + \sum_{t=1}^{T} \sum_{i,j=1}^{M} q_{t-1}^i q_t^j \log \alpha_{ij} + \sum_{t=0}^{T} \sum_{i=1}^{M} \sum_{j=1}^{N} q_t^i y_t^j \log \eta_{ij}\right\}$$

$$= \sum_{i=1}^{M} E\left\{q_0^i\right\} \log \pi_i + \sum_{t=1}^{T} \sum_{i,j=1}^{M} E\left\{q_{t-1}^i q_t^j\right\} \log \alpha_{ij} + \sum_{t=0}^{T} \sum_{i=1}^{M} \sum_{j=1}^{N} E\left\{q_t^i\right\} y_t^j \log \eta_{ij}$$

- M-step is maximizing as before:

?

$$\hat{\pi}_i = E\left\{q_0^i\right\} \qquad \hat{\alpha}_{ij} = \frac{\sum_{t=0}^{T-1} E\left\{q_t^i q_{t+1}^j\right\}}{\sum_{k=1}^{M} \sum_{t=0}^{T-1} E\left\{q_t^i q_{t+1}^k\right\}} \qquad \hat{\eta}_{ij} = \frac{\sum_{t=0}^{T} E\left\{q_t^i\right\} y_t^j}{\sum_{k=1}^{N} \sum_{t=0}^{T} E\left\{q_t^i\right\} y_t^k}$$

- What are E{}'s? $E_{p(x)}\left\{x^i\right\} = \sum_x p\left(x\right) x^i = \sum_x p\left(x\right) \delta\left(x = x^i\right) = p\left(x^i\right)$
- Our JTA ψ & ϕ marginals! (JTA is the E-Step for given θ)

$$E\left\{q_t^i q_{t+1}^j\right\} = p\left(q_t = i, q_{t+1} = j \mid \bar{y}\right) = \frac{\psi^{**}\left(q_t = i, q_{t+1} = j\right)}{\sum_{ij} \psi^{**}\left(q_t = i, q_{t+1} = j\right)} \qquad E\left\{q_t^i\right\} = p\left(q_t = i \mid \bar{y}\right) = \frac{\phi^{**}\left(q_t = i\right)}{\sum_i \phi^{**}\left(q_t = i\right)}$$

$\psi^{**}(q_0, q_1)$ $\phi(q_0)$ and $\phi(q_1)$

Thank you!

- So, to incomplete maximize likelihood with EM,
  - initialize parameters randomly,
  - Run Junction Tree Algorithm to get marginals
  - Use marginals over q's in the maximum likelihood step

- Please complete course evaluation on courseworks
- Good luck with finals week and happy holidays!