Machine Learning

4771

Instructor: Tony Jebara

Topic 20

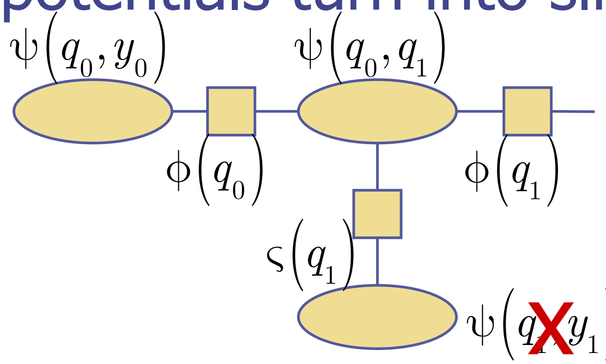
- HMMs with Evidence
- HMM Collect
- HMM Evaluate
- HMM Distribute
- HMM Decode
- HMM Parameter Learning via JTA & EM

HMMs: JTA with Evidence

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

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

- The potentials turn into slices:

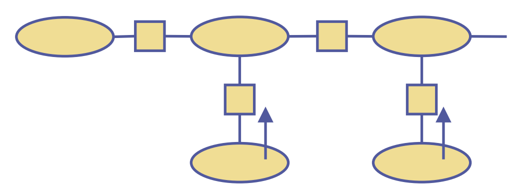


$$\varsigma^*\left(q_t\right) \cancel{=} \sum_{y_t} \psi\left(q_t, y_t\right)$$

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

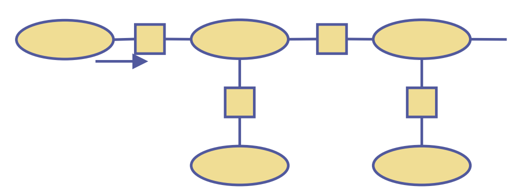
$$\longrightarrow \psi\left(q_1, \bar{y}_1\right)$$

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



- Collect all zeta separators bottom up:

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

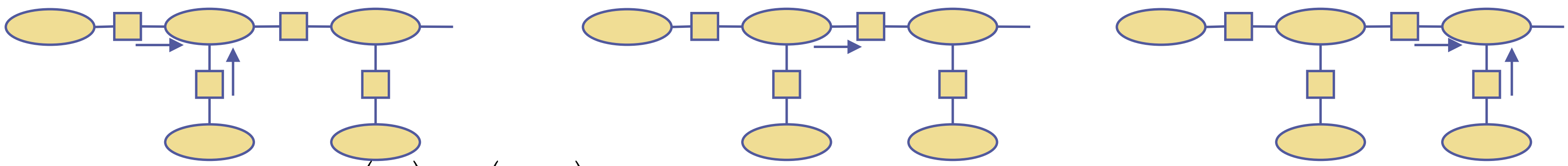


- Collect leftmost phi separator to the right:

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

HMMs: Collect with Evidence

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



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

$$\phi^*\left(q_{t+1}\right) = \sum\nolimits_{q_t}\psi^*\left(q_t,q_{t+1}\right) = \sum\nolimits_{q_t}\phi^*\left(q_t\right)p\left(\bar{y}_{t+1} \mid q_{t+1}\right)\alpha_{q_t,q_{t+1}}$$

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

$$\phi^*\left(q_1\right) = \sum\nolimits_{q_0}p\left(\bar{y}_0,q_0\right)p\left(\bar{y}_1 \mid q_1\right)p\left(q_1 \mid q_0\right) = p\left(\bar{y}_0,\bar{y}_1,q_1\right)$$

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

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

HMMs: Evaluate with Evidence

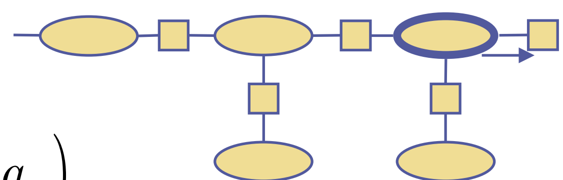
- Say we are solving the first HMM problem:

  1) Evaluate: given $y_0,\ldots,y_T$ & θ compute $p(y_0,\ldots,y_T|\theta)$
- If we want to compute the likelihood, we are already done!
- We really just need to do collect (not even distribute).
- From previous slide we had:

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

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

  its normalizer is p(EVIDENCE)!



hypothetical

  Or use hypothetical $\phi^*\left(q_T\right) = p\left(\bar{y}_0,\ldots,\bar{y}_T,q_T\right)$
- Can compute the likelihood just by marginalizing this phi

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

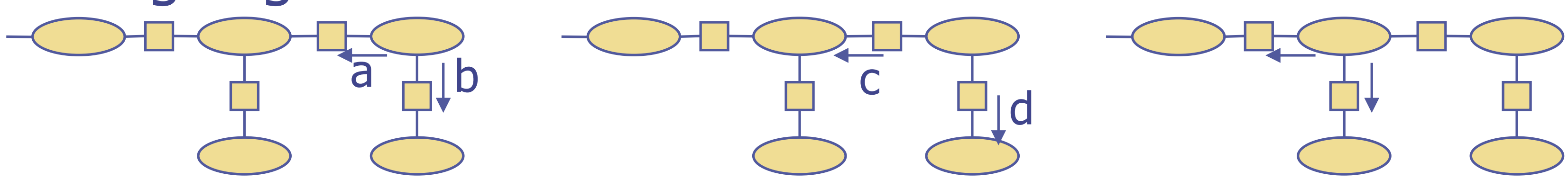
- So, adding up the entries in last ϕ^* gives us the likelihood

HMMs: Distribute with Evidence

- Back to collecting… say just finished collecting to the root with our last update formula:

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

- 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

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

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

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

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

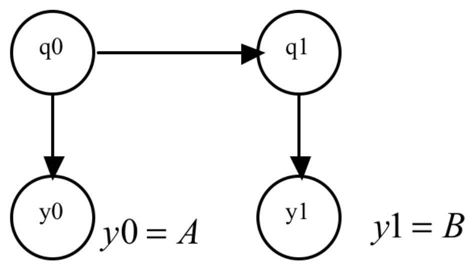
HMM Example

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, evaluate the likelihood of this data p(y) given your HMM and its parameters. Also 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 $p(q_0)$, the state transition matrix given by $p(q_t| q_{t-1})$, and the emission matrix $p(y_t| q_t)$, respectively.

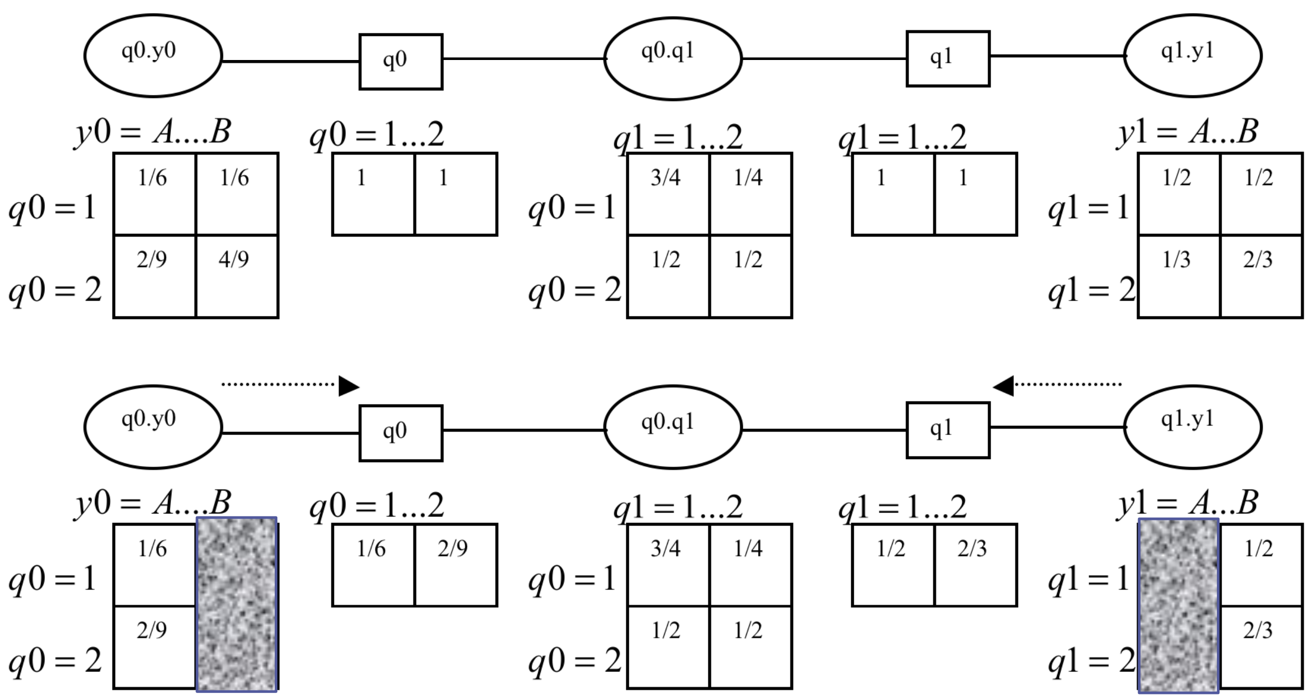
$$\pi = p\left(q_0\right) = \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{bmatrix} 1/3 & 2/3 \end{bmatrix} \end{array}$$

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

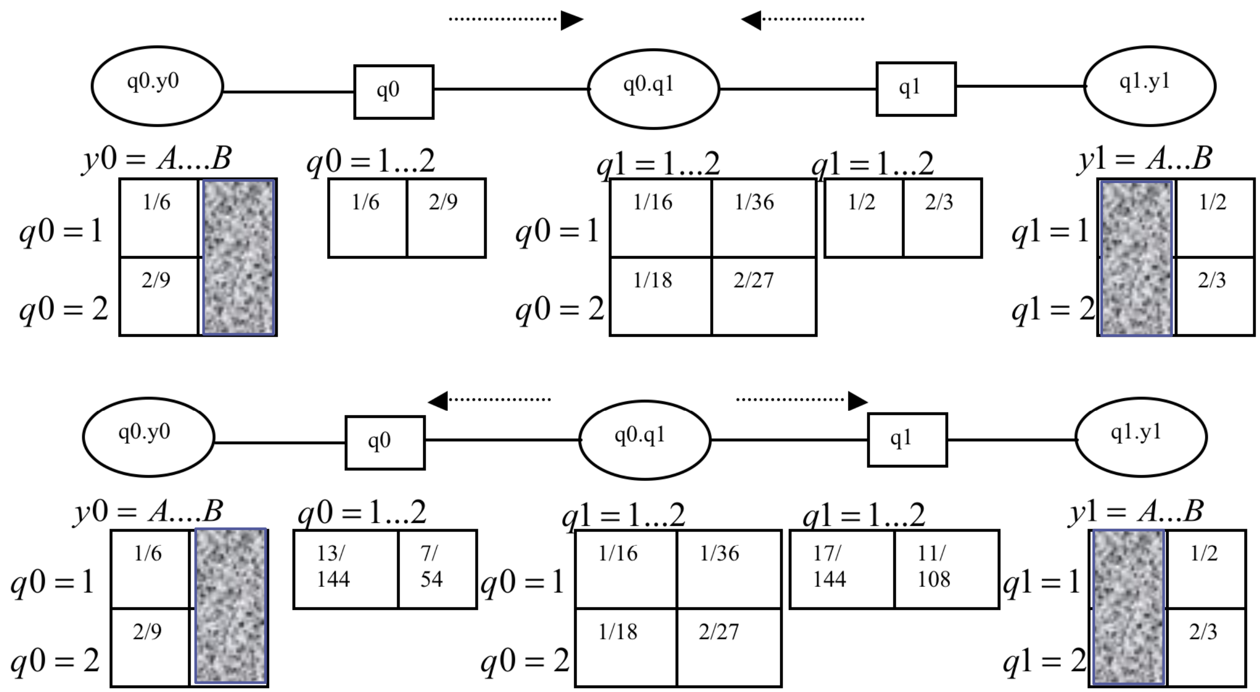
HMM Example



Initialized Junction Tree



HMM Example



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

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

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

HMMs: Marginals & MaxDecoding

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

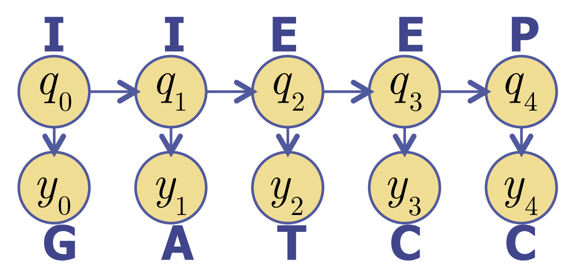
$$\phi^{**}\left(q_t\right) \propto p\left(q_t \mid \bar{y}_1,\ldots,\bar{y}_T\right) \qquad \varsigma^{**}\left(q_{t+1}\right) \propto p\left(q_{t+1} \mid \bar{y}_1,\ldots,\bar{y}_T\right)$$

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

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

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



HMMs: EM Learning

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

$$
\begin{aligned}
l(\theta) &= \log\left(p\left(q,y\right)\right)\\
&= \log\left(p\left(q_0\right)\prod_{t=1}^{T} p\left(q_t \mid q_{t-1}\right)\prod_{t=0}^{T} p\left(\bar{y}_t \mid q_t\right)\right)\\
&= \log p\left(q_0\right) + \sum_{t=1}^{T} \log p\left(q_t \mid q_{t-1}\right) + \sum_{t=0}^{T} \log p\left(\bar{y}_t \mid q_t\right)\\
&= \log \prod_{i=1}^{M}\left[\pi_i\right]^{q_0^i} + \sum_{t=1}^{T} \log \prod_{i=1}^{M}\prod_{j=1}^{M}\left[\alpha_{ij}\right]^{q_{t-1}^i q_t^j} + \sum_{t=0}^{T} \log \prod_{i=1}^{M}\prod_{j=1}^{N}\left[\eta_{ij}\right]^{q_t^i y_t^j}\\
&= \sum_{i=1}^{M} q_0^i \log \pi_i + \sum_{t=1}^{T}\sum_{i,j=1}^{M} q_{t-1}^i q_t^j \log \alpha_{ij} + \sum_{t=0}^{T}\sum_{i=1}^{M}\sum_{j=1}^{N} q_t^i y_t^j \log \eta_{ij}
\end{aligned}
$$

Introduce Lagrange & take derivatives →

$$
\sum_{i=1}^{M} \pi_i = 1 \qquad \sum_{j=1}^{M} \alpha_{ij} = 1 \qquad \sum_{j=1}^{N} \eta_{ij} = 1
$$

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

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)

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

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

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

- 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?

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)

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

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

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

- M-step is maximizing as before:

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

- 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)

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

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!