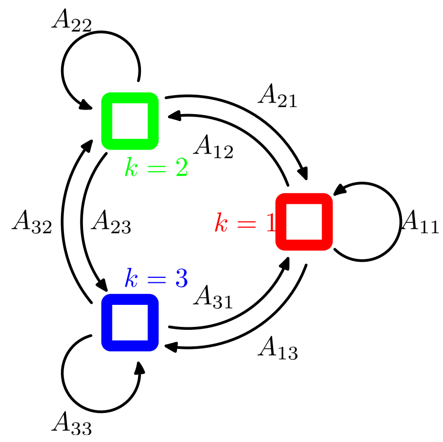
Announcements

Extra resources for EM and HMM posted on lecture page.

(Pages from Bishop, HMM tutorial by Rabiner)

Transition matrix representation

(Not a graphical model)



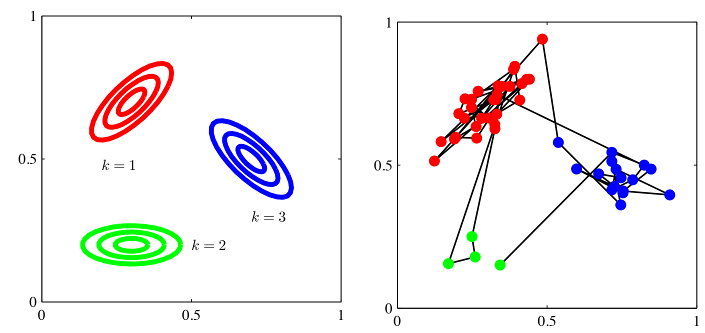
Data distribution from an HMM

An HMM is specified by: $\theta = \{\pi, A, \phi\}$

$$p(X,Z|\theta) = p(z_1|\pi)\left[\prod_{n=2}^{N} p(z_n|z_{n-1}, A)\right]\prod_{m=1}^{N} p(x_m|z_m, \phi)$$

(complete data, i.e., we can generate from this).

Data distribution from an HMM



Transition probability to another state is 5%

Classic HMM computational problems

Given data, what is the HMM (**learning**).

Given an HMM, what is the **distribution over the state** variables. Also, how likely are the observations, given the model.

Given an HMM, what is the most likely **state sequence** for some data?

Learning the HMM

If we know the states, we can compute the parameters.

If we know the parameters, we can compute the states (if we know how to solve the second problem).

General EM algorithm

1. Choose initial values for $\theta^{(s=1)}$
   (can also do assignments, but then jump to M step).
2. E step: Evalute $p\left(Z \middle| X, \theta^{(s)}\right)$
3. M step: Evalute $\theta^{(s+1)} = \arg\max_{\theta} \left\{ \mathrm{Q}\left(\theta^{(s+1)}, \theta^{(s)}\right) \right\}$

   where $\mathrm{Q}\left(\theta^{(s+1)}, \theta^{(s)}\right) = \sum_{Z} p\left(Z \middle| X, \theta^{(s)}\right) \log\left(p\left(X, Z \middle| \theta^{(s+1)}\right)\right)$
4. Check for convergence; If not done, goto 2.

★ At each step, our objective function is increases unless it is at a local maximum. It is important to check this is

Q function for EM for HMM

$$\log\left(p\left(X, Z \middle| \theta\right)\right) =$$

$$\sum_{k=1}^{K} z_{1k} \log(\pi) + \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} z_{n-1,j} z_{n,k} \log\left(p\left(z_n \middle| z_{n-1}, A\right)\right) + \sum_{n=1}^{N} \sum_{k}^{K} z_{nk} \log\left(p\left(x_m \middle| z_m, \phi\right)\right)$$

By analogy with the GMM

$$Q\left(\theta^{(s+1)}, \theta^{(s)}\right) = \sum_{z} p\left(Z \middle| \theta^{(s)}\right) \log\left(X, Z \middle| \theta^{(s+1)}\right)$$

$$= \sum_{k=1}^{K} \gamma\left(z_{1k}\right) \log(\pi) + \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi\left(z_{n-1,j} z_{n,k}\right) \log\left(p\left(z_n \middle| z_{n-1}, A\right)\right)$$

$$+ \sum_{n=1}^{N} \sum_{k}^{K} \gamma\left(z_{nk}\right) \log\left(p\left(x_m \middle| z_m, \phi\right)\right)$$

EM for HMM

Doing the maximization using Lagrange multipliers

$$\pi_k = \frac{\gamma(z_{1k})}{\sum_{k'} \gamma(z_{1k'})}$$ (Much like the mixture model case)

$$A_{jk} = \frac{\sum_{n=2} \zeta(z_{n-1,j}, z_{nk})}{\sum_{k'} \sum_{n=2} \zeta(z_{n-1,j}, z_{nk'})}$$

EM for HMM

The maximization of $p(x_n|\phi)$ is exactly the same as the mixture model.

For example, if we have Gaussian emmision probabilities, then

$$\mu_k = \frac{\sum_n x_n \gamma(z_{nk})}{\sum_n \gamma(z_{nk})}$$

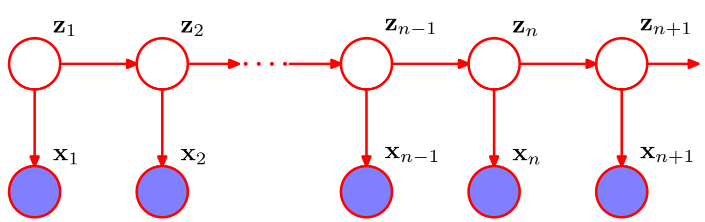
E step for EM for HMM

Computing the E step is a bit more involved.

Recall that in the mixture case it was easy because we only needed to consider the relative likelihood that each cluster independently explain the observations.

However, here the sequence also must play a role.

Graphical model for the E step



Note that our task is to compute marginal probabilities

Computing marginals in an HMM

Various names, flavors, notations, ...

Forward-Backward algorithm

Alpha-beta algorithm

Sum-product for HMM

(Bishop also says "Baum Welch" but that is a synonym for the EM algorithm as whole).

Alpha-beta algorithm

$$\begin{aligned}\gamma(z_n) &= p(z_n|X)\\ &= \frac{p(X|z_n)p(z_n)}{p(X)}\\ &= \frac{p(x_1,....,x_n|z_n)p(x_{n+1},....,x_N|z_n)p(z_n)}{p(X)}\\ &= \frac{p(x_1,....,x_n,z_n)p(x_{n+1},....,x_N|z_n)}{p(X)}\\ &= \frac{\alpha(z_n)\beta(z_n)}{p(X)}\end{aligned}$$

Where we define

$$\alpha(z_n) = p(x_1,....,x_n,z_n)$$

$$\beta(z_n) = p(x_{n+1},....,x_N|z_n)$$

Alpha-beta algorithm

$$\gamma(z_n) = \frac{\alpha(z_n)\beta(z_n)}{p(X)}$$

$$\alpha(z_n) = p(x_1,....,x_n,z_n)$$

$$\beta(z_n) = p(x_{n+1},....,x_N|z_n)$$

Note that in EM, $p(X) = p\left(X|\theta^{(s)}\right)$

Alpha-beta algorithm

$$\begin{aligned}\alpha(z_n) &= p(x_1,....,x_n,z_n)\\ &= p(x_1,....,x_n|z_n)p(z_n)\\ &= p(x_n|z_n)p(x_1,....,x_{n-1}|z_n)p(z_n)\\ &= p(x_n|z_n)p(x_1,....,x_{n-1},z_n)\\ &= p(x_n|z_n)\sum_{z_{n-1}} p(x_1,....,x_{n-1},z_{n-1},z_n)\\ &= p(x_n|z_n)\sum_{z_{n-1}} p(x_1,....,x_{n-1},z_n|z_{n-1})p(z_{n-1})\\ &= p(x_n|z_n)\sum_{z_{n-1}} p(x_1,....,x_{n-1}|z_{n-1})p(z_n|z_{n-1})p(z_{n-1})\\ &= p(x_n|z_n)\sum_{z_{n-1}} p(x_1,....,x_{n-1},z_{n-1})p(z_n|z_{n-1})\\ &= p(x_n|z_n)\sum_{z_{n-1}} \alpha(z_{n-1})p(z_n|z_{n-1})\end{aligned}$$

Alpha-beta algorithm

$$\alpha(z_n) = p(x_n|z_n)\sum_{z_{n-1}}\alpha(z_{n-1})p(z_n|z_{n-1})$$

This is a recursive evaluation of alpha. So we can compute all of them easily if we know the first one, $\alpha(z_1)$.

$$\begin{aligned}\alpha(z_1) &= p(x_1, z_1)\\ &= p(x_1|z_1)\,p(z_1) \qquad \text{(K dimensional vector)}\end{aligned}$$

$$\alpha(z_{1k}) = \pi_k p(x_1|\phi_k)$$

Alpha-beta algorithm

Similarly, we can derive a recurrence relation for beta

$$\beta(z_n) = \sum_{z_{n+1}}\beta(z_{n+1})p(x_{n+1}|z_{n+1})\,p(z_{n+1}|z_n)$$

Alpha-beta algorithm

The details for $\beta(z_n) = \sum_{z_{n+1}}\beta(z_{n+1})p(x_{n+1}|z_{n+1})\,p(z_{n+1}|z_n)$

$$\begin{aligned}\beta(\mathbf{z}_n) &= p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n)\\ &= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N,\mathbf{z}_{n+1}|\mathbf{z}_n)\\ &= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_n,\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n)\\ &= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+1},\ldots,\mathbf{x}_N|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n)\\ &= \sum_{\mathbf{z}_{n+1}} p(\mathbf{x}_{n+2},\ldots,\mathbf{x}_N|\mathbf{z}_{n+1})p(\mathbf{x}_{n+1}|\mathbf{z}_{n+1})p(\mathbf{z}_{n+1}|\mathbf{z}_n).\end{aligned}$$

Alpha-beta algorithm

Our recurrence relation for beta

$$\beta(z_n) = \sum_{z_{n+1}}\beta(z_{n+1})p(x_{n+1}|z_{n+1})\,p(z_{n+1}|z_n)$$

We can compute the betas if we know the last one.

$$\begin{aligned}p(z_N|X) &= \frac{\alpha(z_N)\beta(z_N)}{p(X)}\\ &= \frac{p(X,z_N)\beta(z_N)}{p(X)}\end{aligned}$$

So $\beta(z_N) = 1$

Alpha-beta algorithm

Given the alphas and betas, we can compute all the quantities we need for the E step.

$$\gamma(z_n) = \frac{\alpha(z_n)\beta(z_n)}{p(X)}$$

Note that $\sum_{z_n} \gamma(z_n) = 1$

so $\sum_{z_n} \frac{\alpha(z_n)\beta(z_n)}{p(X)} = 1$

and $p(X) = \sum_{z_n} \alpha(z_n)\beta(z_n)$

We do not need $p(X)$ for EM, but it is the likelihood which we want to monitor ($p(X) = p(X|\theta^{(s)})$).

Alpha-beta algorithm

Given the alphas and betas, we can compute all the quantities we need for the E step.

$$\begin{aligned}
\xi(\mathbf{z}_{n-1}, \mathbf{z}_n) &= p(\mathbf{z}_{n-1}, \mathbf{z}_n|\mathbf{X}) \\
&= \frac{p(\mathbf{X}|\mathbf{z}_{n-1}, \mathbf{z}_n)p(\mathbf{z}_{n-1}, \mathbf{z}_n)}{p(\mathbf{X})} \\
&= \frac{p(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1}|\mathbf{z}_{n-1})p(\mathbf{x}_n|\mathbf{z}_n)p(\mathbf{x}_{n+1}, \ldots, \mathbf{x}_N|\mathbf{z}_n)p(\mathbf{z}_n|\mathbf{z}_{n-1})p(\mathbf{z}_{n-1})}{p(\mathbf{X})} \\
&= \frac{\alpha(\mathbf{z}_{n-1})p(\mathbf{x}_n|\mathbf{z}_n)p(\mathbf{z}_n|\mathbf{z}_{n-1})\beta(\mathbf{z}_n)}{p(\mathbf{X})} \qquad (13.43)
\end{aligned}$$

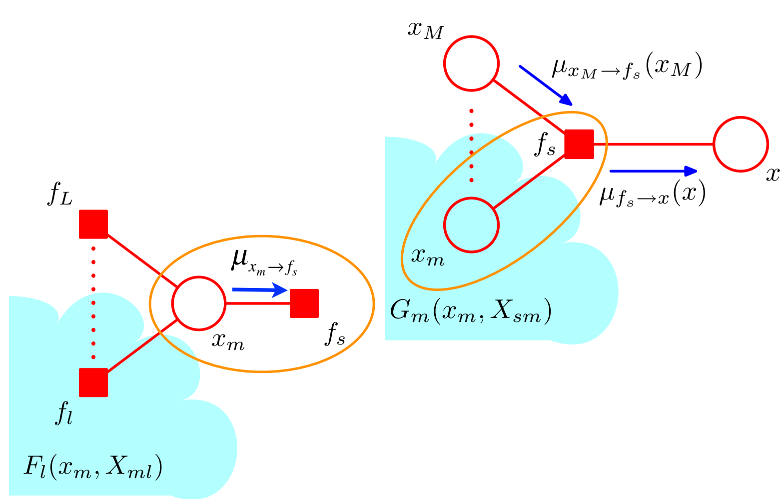
Computing marginals, version two

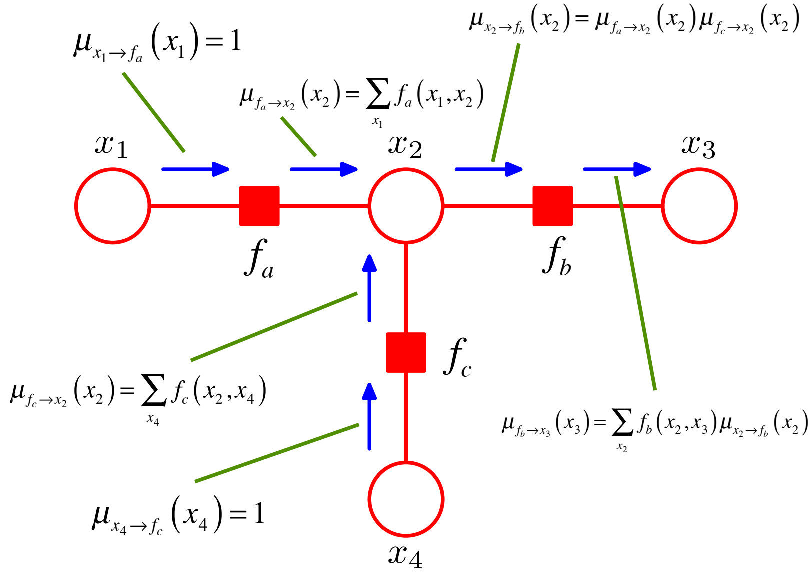
We can apply sum-product to our E step graph.

Recall the recursive nature of the inference problem when we studied the sum-product algorithm.

We could have done recursion (like we just did now), but message passing is more general.

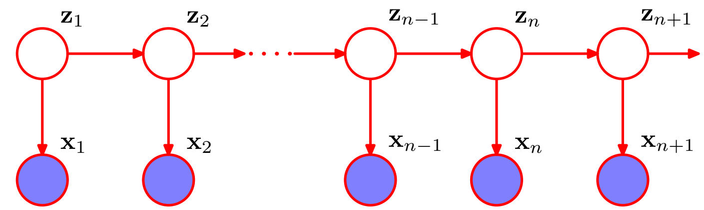
Recall sum-product



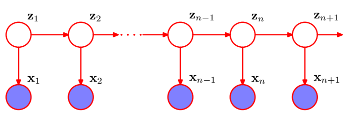


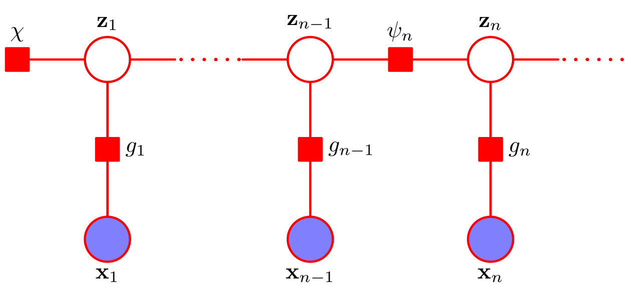
Computing marginals, version two

We can apply sum-product to our E step graph.

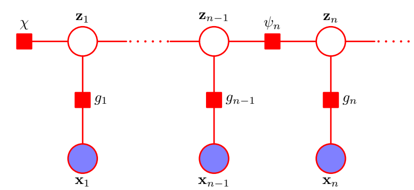


Factor graph

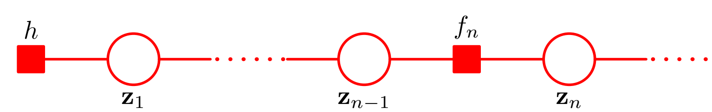




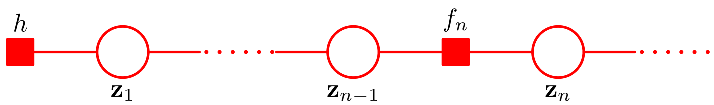
Simplified factor graph



Since we condition on all the x's, we can simplify the graph.



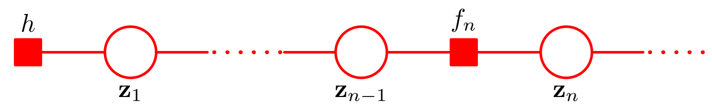
Sum-product for HMM



$$h = p(z_1)\,p(x_1|z_1)$$

$$f_n = p(z_n|z_{n-1})\,p(x_n|z_n)$$

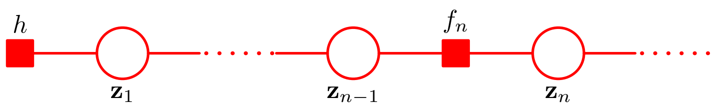
Sum-product for HMM



We can ignore the nodes because they just pass through the incoming message on the single in link.

$$\text{i.e.,}\quad \mu_{f_n \to z_n}(z_n) = \mu_{f_n \to f_{n+1}}(z_n)$$

Sum-product for HMM



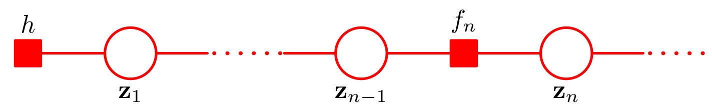
Factor node actions on left to right messages

$$\mu_{f_n \to f_{n+1}}(z_n) = \sum_{z_{n-1}} f_n(z_{n-1}, z_n)\,\mu_{f_{n-1} \to f_n}(z_{n-1})$$

The first message is

$$h = p(z_1)\,p(x_1|z_1) = \alpha(x_1, z_1)$$

Sum-product for HMM



If we identify $\mu_{f_n \to f_{n+1}}(z_n) = \alpha(z_n)$

$$\begin{aligned}
\alpha(z_n) &= \sum_{z_{n-1}} f_n(z_{n-1}, z_n)\,\alpha(z_{n-1}) \\
&= \sum_{z_{n-1}} p(z_n|z_{n-1})\,p(x_n|z_n)\,\alpha(z_{n-1}) \\
&= p(x_n|z_n) \sum_{z_{n-1}} p(z_n|z_{n-1})\,\alpha(z_{n-1})
\end{aligned}$$