M step for HMM

We assume the E step computed distributions for

The degree each state explains each data point (analogous to GMM responsibilities).

$$\gamma(z_n) = p\left(z_n \middle| X, \theta^{(s)}\right)$$

The degree that the combination of a state, and a previous one explain two data points.

$$\xi(z_{n-1}, z_n) = p\left(z_{n-1}, z_n \middle| X, \theta^{(s)}\right)$$

"xi"

EM for HMM (sketch)

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

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

We define

$$\gamma(z_n) = p\left(z_n \middle| X, \theta^{(s)}\right)$$

$$\xi(z_{n-1}, z_n) = p\left(z_{n-1}, z_n \middle| X, \theta^{(s)}\right)$$

"xi"

EM for HMM (sketch)

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

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

By analogy with the GMM

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

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

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

EM for HMM (sketch)

Doing the maximization using Lagrange multipliers gives us

$$\pi_k = \frac{\gamma(z_{1k})}{\sum_{k'} \gamma(z_{1k'})}$$

Much like the GMM. Taking the partial derivative for π_k kills second and third terms.

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

EM for HMM (sketch)

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

For example, if we have Gaussian emmisions, then

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

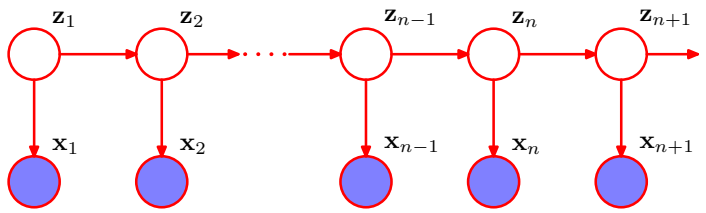
E step for EM for HMM

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

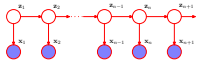
Alpha-beta algorithm

Sum-product for HMM

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

Alpha-beta algorithm

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

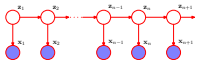


Where we define

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

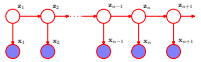
$\beta(z_n) = p(x_{n+1},....,x_N|z_n)$

Expressing alpha recursively



$$\alpha(z_n) = p(x_1,....,x_n,z_n)$$
$$= p(x_1,....,x_n|z_n)\,p(z_n) \quad \text{(definition of "|")}$$
$$= p(x_n|z_n)\,p(x_1,....,x_{n-1}|z_n)\,p(z_n) \quad \text{(conditional independence)}$$
$$= p(x_n|z_n)\,p(x_1,....,x_{n-1},z_n) \quad \text{(definition of "|")}$$
$$= p(x_n|z_n)\sum_{z_{n-1}} p(x_1,....,x_{n-1},z_{n-1},z_n) \quad \text{(marginal)}$$
$$= p(x_n|z_n)\sum_{z_{n-1}} p(x_1,....,x_{n-1},z_n|z_{n-1})p(z_{n-1}) \quad \text{(definition of "|")}$$
$$= p(x_n|z_n)\sum_{z_{n-1}} p(x_1,....,x_{n-1}|z_{n-1})p(z_n|z_{n-1})\,p(z_{n-1}) \quad \text{(conditional independence)}$$
$$= p(x_n|z_n)\sum_{z_{n-1}} p(x_1,....,x_{n-1},z_{n-1})p(z_n|z_{n-1}) \quad \text{(definition of "|")}$$
$$= p(x_n|z_n)\sum_{z_{n-1}} \alpha(z_{n-1})p(z_n|z_{n-1}) \quad \text{(definition of } \alpha(z_n)\text{)}$$

Expressing alpha recursively



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

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

$$\alpha(z_1) = p(x_1,z_1) \quad \text{(we defined } \alpha(z_n) = p(x_1,....,x_n,z_n)\text{)}$$
$$= p(z_1)\,p(x_1|z_1) \quad \text{(this is a K dimensional vector for fixed } x_1\text{)}$$

$$\alpha(z_1)_k = \pi_k\, p(x_1|\phi_k)$$

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.

$$\begin{aligned}
p(z_N|X) &= \frac{\alpha(z_N)\beta(z_N)}{p(X)} \\
&= \frac{p(X, z_N)\beta(z_N)}{p(X)} \qquad \text{(we defined } \alpha(z_n) = p(x_1, \ldots, x_n, z_n)\text{)} \\
&= p(z_N|X)\beta(z_N)
\end{aligned}$$

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.

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

We know 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.

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

(in Bishop)