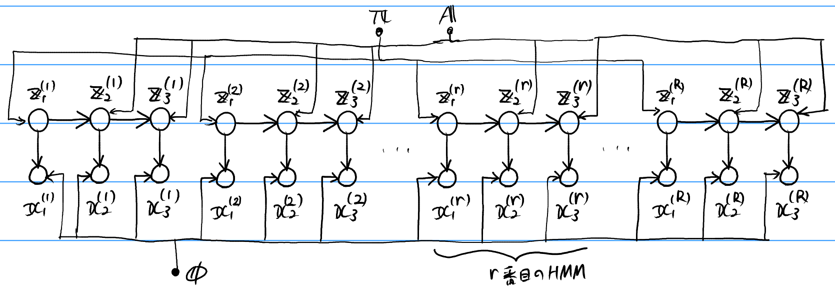
13.12

R個の独立したHMMのモデル図



r番目のHMMに対して EMアルゴリズムの式 (13.11)〜(13.23) と

フォワードバックワードアルゴリズムの式 (13.24)〜(13.44) が成立する。

r番目のHMMに対する式の変数,関数には (r) という印をつけることとする。

$X^{(r)} = \{x_1^{(r)} \cdots x_N^{(r)}\}$, $Z^{(r)} = \{z_1^{(r)} \cdots z_N^{(r)}\}$, $X = \{X^{(1)} \cdots X^{(R)}\}$, $Z = \{Z^{(1)} \cdots Z^{(R)}\}$ とする

R個の独立したHMMの同時分布は

$$p(X, Z \mid \theta) = \prod_{r=1}^{R} p(X^{(r)}, Z^{(r)} \mid \theta)$$ ← $\{X^{(t)}, Z^{(t)}\} \perp\!\!\!\perp \{X^{(t)}, Z^{(t)}\} \mid \theta$ なので

Z の事後分布は

$$p(Z \mid X, \theta^{old}) = p(Z^{(1)} \mid Z^{(2)} \cdots Z^{(R)}, X, \theta^{old})\, p(Z^{(2)} \cdots Z^{(R)} \mid X, \theta^{old})$$

$$= p(Z^{(1)} \mid X^{(1)}, \theta^{old})\, p(Z^{(2)} \cdots Z^{(R)} \mid X, \theta^{old})$$

$p(Z^{(1)} \mid Z^{(2)} \cdots Z^{(R)}, X, \theta^{old})$
$= p(Z^{(1)} \mid X^{(1)}, \theta^{old})$ ← $Z^{(1)} \perp\!\!\!\perp \{Z^{(2)} \cdots Z^{(R)}, X^{(2)} \cdots X^{(R)}\} \mid \theta$ より

$$\vdots$$

$$= \prod_{r=1}^{R} p(Z^{(r)} \mid X^{(r)}, \theta^{old})$$

となる。

ここで

$$p(\mathbf{z}^{(r)}|\mathbf{X}^{(r)}) = p(\mathbf{z}_N^{(r)}|\mathbf{z}_{N-1}^{(r)}\cdots\mathbf{z}_1^{(r)},\mathbf{X}^{(r)})\,p(\mathbf{z}_{N-1}^{(r)}\cdots\mathbf{z}_1^{(r)}|\mathbf{X}^{(r)})$$ ← 乗法定理

$$= p(\mathbf{z}_N^{(r)}|\mathbf{z}_{N-1}^{(r)},\mathbf{X}^{(r)})\,p(\mathbf{z}_{N-1}^{(r)}\cdots\mathbf{z}_1^{(r)}|\mathbf{X}^{(r)})$$ ← $\mathbf{z}_N^{(r)} \perp\!\!\!\perp \{\mathbf{z}_{N-2}^{(r)}\cdots\mathbf{z}_1^{(r)}\} \mid \mathbf{z}_{N-1}^{(r)}$ より $p(\mathbf{z}_N^{(r)}|\mathbf{z}_{N-1}^{(r)}\cdots\mathbf{z}_1^{(r)},\mathbf{X}^{(r)}) = p(\mathbf{z}_N^{(r)}|\mathbf{z}_{N-1}^{(r)},\mathbf{X}^{(r)})$

$$\vdots$$

$$= p(\mathbf{z}_N^{(r)}|\mathbf{z}_{N-1}^{(r)},\mathbf{X}^{(r)})\cdots p(\mathbf{z}_2^{(r)}|\mathbf{z}_1^{(r)},\mathbf{X}^{(r)})\,p(\mathbf{z}_1^{(r)}|\mathbf{X}^{(r)})$$

$$= \frac{p(\mathbf{z}_N^{(r)},\mathbf{z}_{N-1}^{(r)}|\mathbf{X}^{(r)})}{p(\mathbf{z}_{N-1}^{(r)}|\mathbf{X}^{(r)})}\cdots\frac{p(\mathbf{z}_2^{(r)},\mathbf{z}_1^{(r)}|\mathbf{X}^{(r)})}{p(\mathbf{z}_1^{(r)}|\mathbf{X}^{(r)})}\,p(\mathbf{z}_1^{(r)}|\mathbf{X}^{(r)})$$

$$= \frac{\xi^{(r)}(\mathbf{z}_{N-1}^{(r)},\mathbf{z}_N^{(r)})}{\gamma^{(r)}(\mathbf{z}_{N-1}^{(r)})}\cdots\frac{\xi^{(r)}(\mathbf{z}_1^{(r)},\mathbf{z}_2^{(r)})}{\gamma^{(r)}(\mathbf{z}_1^{(r)})}\,\gamma^{(r)}(\mathbf{z}_1^{(r)})$$

$$= \frac{\prod_{n=2}^{N}\xi^{(r)}(\mathbf{z}_{n-1}^{(r)},\mathbf{z}_n^{(r)})}{\prod_{n=2}^{N-1}\gamma^{(r)}(\mathbf{z}_n^{(r)})}$$

なるので

$$p(\mathbf{Z}|\mathbf{X},\theta^{old}) = \prod_{r=1}^{R} p(\mathbf{z}^{(r)}|\mathbf{X}^{(r)},\theta^{old}) = \prod_{r=1}^{R}\frac{\prod_{n=2}^{N}\xi^{(r)}(\mathbf{z}_{n-1}^{(r)},\mathbf{z}_n^{(r)})}{\prod_{n=2}^{N-1}\gamma^{(r)}(\mathbf{z}_n^{(r)})}$$

となる。

これよりEステップで求める$\mathbf{Z}$の事後分布$p(\mathbf{Z}|\mathbf{X},\theta^{old})$は、
$\xi^{(r)}(\mathbf{z}_{n-1}^{(r)},\mathbf{z}_n^{(r)})$, $\gamma^{(r)}(\mathbf{z}_n^{(r)})$により計算されることがわかる。
また$\xi^{(r)}(\mathbf{z}_{n-1}^{(r)},\mathbf{z}_n^{(r)})$, $\gamma^{(r)}(\mathbf{z}_n^{(r)})$は$\alpha^{(r)}(\mathbf{z}_n)$, $\beta^{(r)}(\mathbf{z}_n)$により計算される。(13.33), (13.43)
よって$\alpha^{(r)}(\mathbf{z}_n)$, $\beta^{(r)}(\mathbf{z}_n)$を再帰で求めれば$p(\mathbf{Z}|\mathbf{X},\theta^{old})$を計算できる。

実際のEステップでは$p(\mathbf{Z}|\mathbf{X},\theta^{old})$を明示的に求めることはない。
実際に求めるのは$p(\mathbf{Z}|\mathbf{X},\theta^{old})$を確率~~密~~な分布とする期待値$Q(\theta,\theta^{old})$である。
$Q(\theta,\theta^{old})$についても$\mathbf{x}_n^{(r)}$以外のノードの寄与は、$\gamma^{(r)}(\mathbf{z}_n^{(r)})$, $\xi^{(r)}(\mathbf{z}_{n-1}^{(r)},\mathbf{z}_n^{(r)})$で与えられることが分かる(以下に示している。)なので$\alpha^{(r)}(\mathbf{z}_n)$, $\beta^{(r)}(\mathbf{z}_n)$を求めれば$Q(\theta,\theta^{old})$も計算できる。

パラメータπ, Aは$\mathbf{x}_n^{(r)}$以外のノードにリンクされているので、
$\mathbf{x}_n^{(r)}$以外のノードの寄与が求められればπ, Aを決めることができる。

$\mathbf{Z}$の事後分布 $p(\mathbf{Z}|\mathbf{X},\theta^{old})$ より

$$Q(\theta,\theta^{old}) = \mathop{\mathbb{E}}_{p(\mathbf{Z}|\mathbf{X},\theta^{old})}[\ln p(\mathbf{X},\mathbf{Z}|\theta)] = \sum_{\mathbf{Z}} \prod_{r=1}^{R} p(\mathbf{Z}^{(r)}|\mathbf{X}^{(r)},\theta^{old}) \ln \prod_{n=1}^{R} p(\mathbf{X}^{(n)},\mathbf{Z}^{(n)}|\theta)$$

$$= \sum_{\mathbf{Z}^{(1)}} \cdots \sum_{\mathbf{Z}^{(R)}} \prod_{r=1}^{R} p(\mathbf{Z}^{(r)}|\mathbf{X}^{(r)},\theta^{old}) \sum_{n=1}^{R} \ln p(\mathbf{X}^{(n)},\mathbf{Z}^{(n)}|\theta)$$

$$= \sum_{\mathbf{Z}^{(1)}} \cdots \sum_{\mathbf{Z}^{(R-1)}} \prod_{r=1}^{R-1} p(\mathbf{Z}^{(r)}|\mathbf{X}^{(r)},\theta^{old}) \sum_{\mathbf{Z}^{(R)}} p(\mathbf{Z}^{(R)}|\mathbf{X}^{(R)},\theta^{old}) \sum_{n=1}^{R} \ln p(\mathbf{X}^{(n)},\mathbf{Z}^{(n)}|\theta)$$

$$= \sum_{\mathbf{Z}^{(1)}} \cdots \sum_{\mathbf{Z}^{(R-1)}} \prod_{r=1}^{R-1} p(\mathbf{Z}^{(r)}|\mathbf{X}^{(r)},\theta^{old}) \sum_{\mathbf{Z}^{(R)}} p(\mathbf{Z}^{(R)}|\mathbf{X}^{(R)},\theta^{old}) \left\{ \sum_{n=1}^{R-1} \ln p(\mathbf{X}^{(n)},\mathbf{Z}^{(n)}|\theta) + \ln p(\mathbf{X}^{(R)},\mathbf{Z}^{(R)}|\theta) \right\}$$

$$= \sum_{\mathbf{Z}^{(1)}} \cdots \sum_{\mathbf{Z}^{(R-1)}} \prod_{r=1}^{R-1} p(\mathbf{Z}^{(r)}|\mathbf{X}^{(r)},\theta^{old}) \left\{ \sum_{n=1}^{R-1} \ln p(\mathbf{X}^{(n)},\mathbf{Z}^{(n)}|\theta) + \sum_{\mathbf{Z}^{(R)}} p(\mathbf{Z}^{(R)}|\mathbf{X}^{(R)},\theta^{old}) \ln p(\mathbf{X}^{(R)},\mathbf{Z}^{(R)}|\theta) \right\}$$

$$= \sum_{\mathbf{Z}^{(1)}} \cdots \sum_{\mathbf{Z}^{(R-1)}} \prod_{r=1}^{R-1} p(\mathbf{Z}^{(r)}|\mathbf{X}^{(r)},\theta^{old}) \left\{ \sum_{n=1}^{R-1} \ln p(\mathbf{X}^{(n)},\mathbf{Z}^{(n)}|\theta) + Q^{(R)}(\theta,\theta^{old}) \right\}$$

$$= \left\{ \sum_{\mathbf{Z}^{(1)}} \cdots \sum_{\mathbf{Z}^{(R-1)}} \prod_{r=1}^{R-1} p(\mathbf{Z}^{(r)}|\mathbf{X}^{(r)},\theta^{old}) \sum_{n=1}^{R-1} \ln p(\mathbf{X}^{(n)},\mathbf{Z}^{(n)}|\theta) \right\} + Q^{(R)}(\theta,\theta^{old})$$

$$\sum_{\mathbf{Z}^{(1)}} \cdots \sum_{\mathbf{Z}^{(R-1)}} \prod_{r=1}^{R-1} p(\mathbf{Z}^{(r)}|\mathbf{X}^{(r)},\theta^{old}) Q^{(R)}(\theta,\theta^{old})$$

$$= \underbrace{\left(\sum_{\mathbf{Z}^{(1)}} \cdots \sum_{\mathbf{Z}^{(R-1)}} \prod_{r=1}^{R-1} p(\mathbf{Z}^{(r)}|\mathbf{X}^{(r)},\theta^{old}) \right)}_{=1} Q^{(R)}(\theta,\theta^{old})$$

$$= Q^{(R)}(\theta,\theta^{old})$$

$$= \sum_{r=1}^{R} Q^{(r)}(\theta,\theta^{old})$$

となる。

よって (13.17) より

$$Q(\theta,\theta^{old}) = \sum_{r=1}^{R} \left\{ \sum_{k=1}^{K} \gamma^{(r)}(z_{1k}^{(r)}) \ln \pi_k + \sum_{n=2}^{N} \sum_{j=1}^{K} \sum_{k=1}^{K} \xi^{(r)}(z_{n-1,j}^{(r)}, z_{nk}^{(r)}) \ln A_{jk} \right.$$

$$\left. + \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma^{(r)}(z_{nk}^{(r)}) \ln p(x_n^{(r)}|\phi_k) \right\}$$

となる。

π, Aについて Qを最適化する

ラグランジアンは

制約 $\sum_{k=1}^{K} \pi_k = 1$ 制約 $\sum_{k=1}^{K} A_{jk} = 1$

$$L = Q(\theta, \theta^{old}) + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right) + \sum_{j=1}^{K} \mu_j \left(\sum_{k=1}^{K} A_{jk} - 1\right)$$

π_kについての停留条件は

$$0 = \frac{\partial L}{\partial \pi_k} = \sum_{r=1}^{R} \gamma^{(r)}(z_{1k}^{(r)}) \frac{1}{\pi_k} + \lambda$$

$$\therefore \pi_k = -\frac{1}{\lambda} \sum_{r=1}^{R} \gamma^{(r)}(z_{1k}^{(r)})$$

ここで

$$1 = \sum_{k=1}^{K} \pi_k = -\frac{1}{\lambda} \sum_{r=1}^{R} \sum_{k=1}^{K} \gamma^{(r)}(z_{1k}^{(r)})$$

$$\therefore -\frac{1}{\lambda} = \frac{1}{\sum_{r=1}^{R} \sum_{k=1}^{K} \gamma^{(r)}(z_{1k}^{(r)})}$$

よって

$$\pi_k = \frac{\sum_{r=1}^{R} \gamma^{(r)}(z_{1k}^{(r)})}{\sum_{r=1}^{R} \sum_{k=1}^{K} \gamma^{(r)}(z_{1k}^{(r)})} \quad \cdots (13.124)$$

を得る。

A_{jk}についての停留条件は

$$0 = \frac{\partial L}{\partial A_{jk}} = \sum_{r=1}^{R} \sum_{n=2}^{N} \xi^{(r)}(z_{n-1j}^{(r)}, z_{nk}^{(r)}) \frac{1}{A_{jk}} + \mu_j$$

$$\therefore A_{jk} = -\frac{1}{\mu_j} \sum_{r=1}^{R} \sum_{n=2}^{N} \xi^{(r)}(z_{n-1j}^{(r)}, z_{nk}^{(r)})$$

ここで

$$1 = \sum_{k=1}^{K} A_{jk} = -\frac{1}{\mu_j} \sum_{r=1}^{R} \sum_{k=1}^{K} \sum_{n=2}^{N} \xi^{(r)}(z_{n-1j}^{(r)}, z_{nk}^{(r)})$$

$$\therefore -\frac{1}{\mu_j} = \frac{1}{\sum_{r=1}^{R} \sum_{k=1}^{K} \sum_{n=2}^{N} \xi^{(r)}(z_{n-1j}^{(r)}, z_{nk}^{(r)})}$$

なので

$$A_{jk} = \frac{\sum_{r=1}^{R} \sum_{n=2}^{N} \xi^{(r)}(z_{n-1j}^{(r)}, z_{nk}^{(r)})}{\sum_{r=1}^{R} \sum_{k=1}^{K} \sum_{n=2}^{N} \xi^{(r)}(z_{n-1j}^{(r)}, z_{nk}^{(r)})} \quad \cdots (13.125)$$

を得る。

出力分布をガウス分布とする

$$p(\boldsymbol{x}_n^{(r)} \mid \phi_k) = \mathcal{N}(\boldsymbol{x}_n^{(r)} \mid \boldsymbol{\mu}_k, \Sigma_k)$$

このとき $Q(\theta, \theta^{old})$ は

$\phi_k = \{\boldsymbol{\mu}_k, \Sigma_k\}$

$$Q(\theta, \theta^{old}) = C + \sum_{r=1}^{R} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma^{(r)}(z_{nk}^{(r)}) \ln p(\boldsymbol{x}_n^{(r)} \mid \phi_k)$$

, C : $\boldsymbol{\mu}_k$ を含まない項

$$= C + \sum_{r=1}^{R} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma^{(r)}(z_{nk}^{(r)}) \ln \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\Sigma_k|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_n^{(r)} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\boldsymbol{x}_n^{(r)} - \boldsymbol{\mu}_k)\right\}$$

$$= C + \sum_{r=1}^{R} \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma^{(r)}(z_{nk}^{(r)}) \left(-\frac{1}{2}\right) (\boldsymbol{x}_n^{(r)} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\boldsymbol{x}_n^{(r)} - \boldsymbol{\mu}_k)$$

$\boldsymbol{\mu}_k$ についての停留条件は

$$\boldsymbol{0} = \frac{\partial Q}{\partial \boldsymbol{\mu}_k} = \sum_{r=1}^{R} \sum_{n=1}^{N} \gamma^{(r)}(z_{nk}^{(r)}) \left(-\frac{1}{2}\right) 2 (-1) \Sigma_k^{-1} (\boldsymbol{x}_n^{(r)} - \boldsymbol{\mu}_k) \leftarrow \left\{ \begin{array}{l} (\Sigma_k^{-1})^T = \Sigma_k^{-1} \\ \frac{\partial}{\partial \boldsymbol{x}} \boldsymbol{u}^T A \boldsymbol{v} = \frac{\partial \boldsymbol{u}}{\partial \boldsymbol{x}} A \boldsymbol{v} + \frac{\partial \boldsymbol{v}}{\partial \boldsymbol{x}} A^T \boldsymbol{u} \ (\text{denominator layout}) \end{array} \right.$$

演習13.7の解で導いた

$$= \sum_{r=1}^{R} \sum_{n=1}^{N} \gamma^{(r)}(z_{nk}^{(r)}) \Sigma_k^{-1} (\boldsymbol{x}_n^{(r)} - \boldsymbol{\mu}_k)$$

両辺に Σ_k を掛けて

$$\boldsymbol{0} = \sum_{r=1}^{R} \sum_{n=1}^{N} \gamma^{(r)}(z_{nk}^{(r)}) (\boldsymbol{x}_n^{(r)} - \boldsymbol{\mu}_k)$$

$$\therefore \sum_{r=1}^{R} \sum_{n=1}^{N} \gamma^{(r)}(z_{nk}^{(r)}) \boldsymbol{x}_n^{(r)} = \sum_{r=1}^{R} \sum_{n=1}^{N} \gamma^{(r)}(z_{nk}^{(r)}) \boldsymbol{\mu}_k$$

$$\therefore \boldsymbol{\mu}_k = \frac{\sum_{r=1}^{R} \sum_{n=1}^{N} \gamma^{(r)}(z_{nk}^{(r)}) \boldsymbol{x}_n^{(r)}}{\sum_{r=1}^{R} \sum_{n=1}^{N} \gamma^{(r)}(z_{nk}^{(r)})} \quad \cdots (13.126)$$

を得る。