統計的機械学習
第6回　混合ガウスモデル2

2013/07/02

表現工学科　尾形研究室　ゼミ
野田　邦昭

ガウス混合モデル

■ガウス混合モデル(Gaussian mixture model):

$$q(x;\theta) = \sum_{j=1}^{m} w_j \phi(x;\mu_j,\sigma_j)$$

$$w_j \geq 0, \sum_{j=1}^{m} w_j = 1$$

$$\theta = (w_1,\ldots,w_m,\mu_1^T,\ldots,\mu_m^T,\sigma_1,\ldots,\sigma_m)^T$$

$$\mu_j \in \Re^d, \sigma_j > 0$$

$$\phi(x;\mu,\sigma) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{(x-\mu)^T(x-\mu)}{2\sigma^2}\right)$$

m:混合数

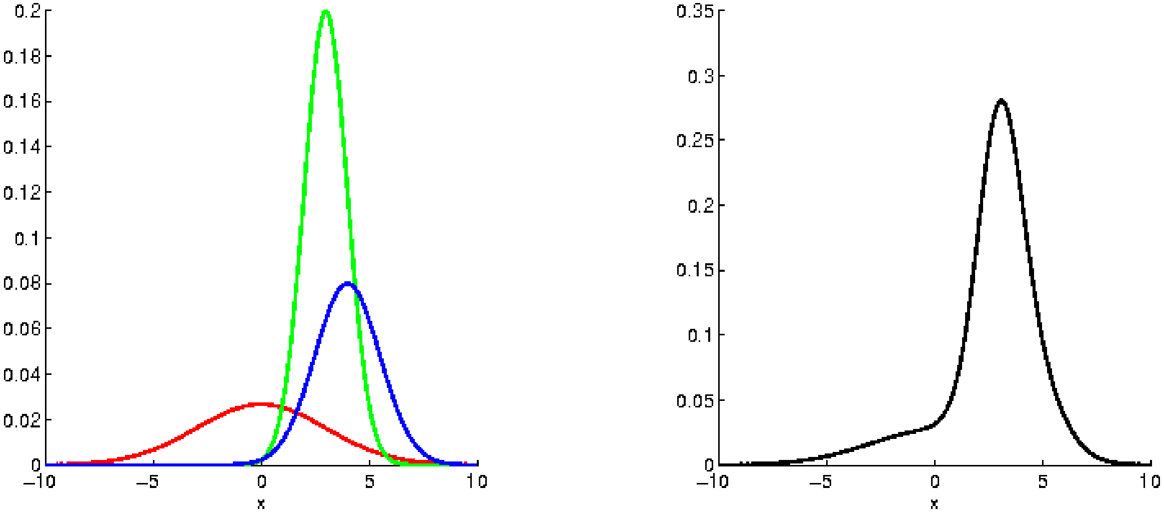
■注意: $q(x;\theta)$ は確率密度関数なので

$$\int_D q(x;\theta)dx = 1 \quad \forall x \in D, q(x;\theta) \geq 0$$

を満たす.

ガウス混合モデル

- 有限個のガウスモデルの線形結合：



- 通常のガウスモデルよりも複雑な確率密度関数を表現できる.
- ガウスカーネル密度推定より単純なので，訓練標本が比較的少ない場合でも推定精度が良い.

ガウス混合モデルの最尤推定

- 最尤推定：（対数）尤度（訓練標本 $\{x_i\}_{i=1}^n$ が生成される確率）を最大にするように θ を定める

$$\log L(\theta) = \sum_{i=1}^n \log q(x_i;\theta) = \sum_{i=1}^n \log \sum_{j=1}^m w_j \phi(x_i;\mu_j,\sigma_j)$$

- 但し，拘束条件 $w_j \geq 0, \sum_{j=1}^m w_j = 1$ を考慮しなければならない！

- $w_j = \dfrac{\exp(\gamma_j)}{\sum_{j'=1}^m \exp(\gamma_{j'})}$ とおき，$\gamma_j \in \Re$ を決定することにすれば，拘束条件は自動的に満たされる．

尤度方程式

- 最尤推定解の必要条件:

$$\frac{\partial \log L}{\partial \gamma_j} = 0, \quad \frac{\partial \log L}{\partial \mu_j} = 0, \quad \frac{\partial \log L}{\partial \sigma_j} = 0$$

より，最尤推定解は以下を満たす（証明は宿題）

$$\hat{w}_j = \frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_{i,j}$$

$$\hat{\sigma}_j = \sqrt{\frac{1}{d}\frac{\sum_{i=1}^{n} \hat{\eta}_{i,j}(x_i - \hat{\mu}_j)^T (x_i - \hat{\mu}_j)}{\sum_{i'=1}^{n} \hat{\eta}_{i',j}}}$$

$$\hat{\mu}_j = \frac{\sum_{i=1}^{n} \hat{\eta}_{i,j} x_i}{\sum_{i'=1}^{n} \hat{\eta}_{i',j}}$$

$$\hat{\eta}_{i,j} = \frac{\hat{w}_j \phi(x_i; \hat{\mu}_j, \hat{\sigma}_j)}{\sum_{j'=1}^{m} \hat{w}_{j'} \phi(x_i; \hat{\mu}_{j'}, \hat{\sigma}_{j'})}$$

$d: x$の次元

- しかし，この連立方程式は簡単に解けない．

EMアルゴリズム(expectation-maximization algorithm)

- 適当な初期値から開始($t=0$): $\left\{\hat{w}_j^{(t)}, \hat{\mu}_j^{(t)}, \hat{\sigma}_j^{(t)}\right\}_{j=1}^{m}$
- Eステップ: $\left\{\hat{w}_j^{(t)}, \hat{\mu}_j^{(t)}, \hat{\sigma}_j^{(t)}\right\}_{j=1}^{m}$ から $\left\{\hat{\eta}_{i,j}^{(t)}\right\}_{i=1,j=1}^{n,m}$ を計算

$$\hat{\eta}_{i,j}^{(t)} = \frac{\hat{w}_j^{(t)}\phi(x_i;\hat{\mu}_j^{(t)},\hat{\sigma}_j^{(t)})}{\sum_{j'=1}^{m}\hat{w}_{j'}^{(t)}\phi(x_i;\hat{\mu}_{j'}^{(t)},\hat{\sigma}_{j'}^{(t)})}$$

- Mステップ: $\left\{\hat{\eta}_{i,j}^{(t)}\right\}_{i=1,j=1}^{n,m}$ から $\left\{\hat{w}_j^{(t+1)}, \hat{\mu}_j^{(t+1)}, \hat{\sigma}_j^{(t+1)}\right\}_{j=1}^{m}$ を計算

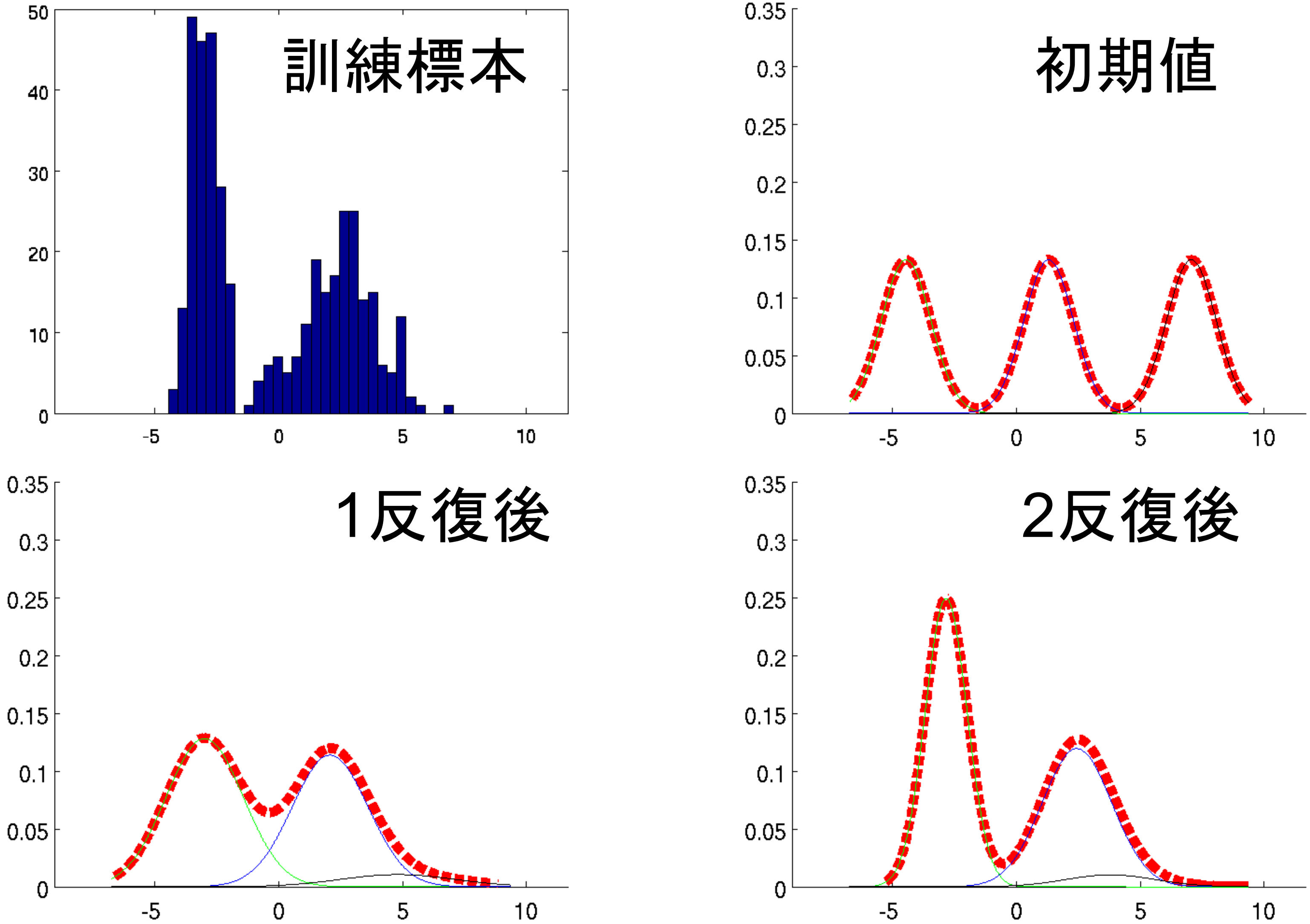
$$\hat{w}_j^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_{i,j}^{(t)}$$

$$\hat{\mu}_j^{(t+1)} = \frac{\sum_{i=1}^{n}\hat{\eta}_{i,j}^{(t)}x_i}{\sum_{i'=1}^{n}\hat{\eta}_{i',j}^{(t)}}$$

$$\hat{\sigma}_j^{(t+1)} = \sqrt{\frac{1}{d}\frac{\sum_{i=1}^{n}\hat{\eta}_{i,j}^{(t)}(x_i-\hat{\mu}_j^{(t)})^T(x_i-\hat{\mu}_j^{(t)})}{\sum_{i'=1}^{n}\hat{\eta}_{i',j}^{(t)}}}$$

- $t=t+1$ し，収束するまでE,Mステップを繰り返す．

例（混合数3）


訓練標本
初期値
1反復後
2反復後
0 10 20 30 40 50
0 0.05 0.1 0.15 0.2 0.25 0.3 0.35
-5 0 5 10

EMアルゴリズムの解釈

■EMアルゴリズムによって尤度は単調非減少

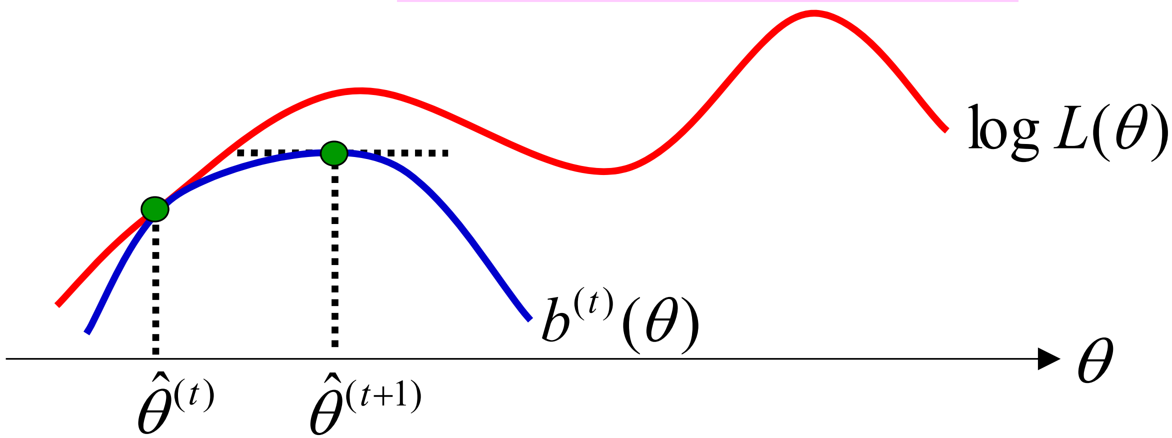
- Eステップ: $\hat{\theta}^{(t)}$を通る対数尤度の下界を求めることに対応

$$\forall \theta, \ \log L(\theta) \geq b^{(t)}(\theta)$$

$$\log L(\hat{\theta}^{(t)}) = b^{(t)}(\hat{\theta}^{(t)})$$

- Mステップ: 下界を最大化するパラメータ値を求めることに対応

$$\hat{\theta}^{(t+1)} = \arg\max_{\theta} b^{(t)}(\theta)$$

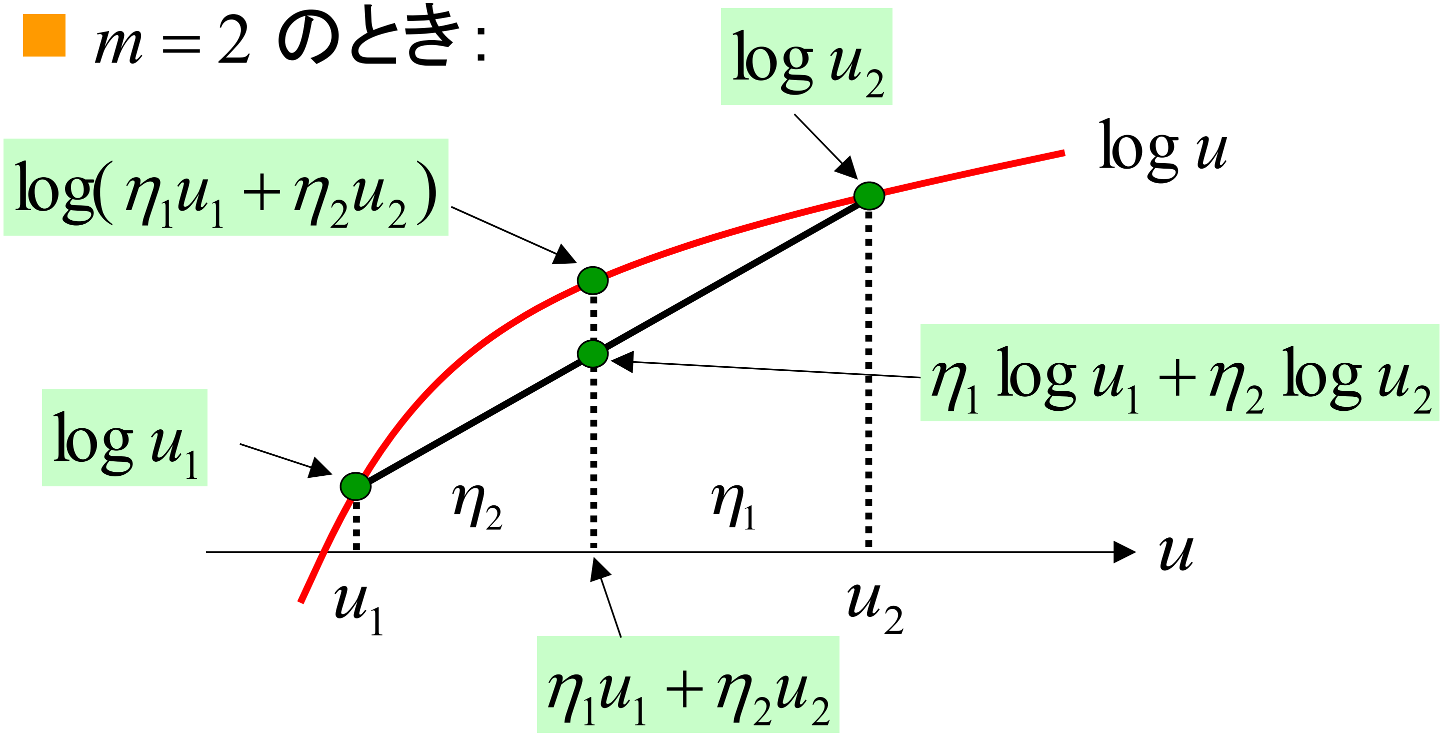


Eステップの導出

- ジェンセンの不等式(Jensen's inequality):

$$\log \sum_{j=1}^{m} \eta_j u_j \geq \sum_{j=1}^{m} \eta_j \log u_j \qquad \eta_j \geq 0, \sum_{j=1}^{m} \eta_j = 1$$

- $m = 2$ のとき:



Eステップの導出(続き)

- 対数尤度: $$\log L(\theta) = \sum_{i=1}^{n} \log \sum_{j=1}^{m} w_j \phi(x_i; \mu_j, \sigma_j)$$

- ダミーの変数 $\hat{\eta}_{i,j}^{(t)}$ を入れる:

$$\log L(\theta) = \sum_{i=1}^{n} \log \sum_{j=1}^{m} \hat{\eta}_{i,j}^{(t)} \underbrace{\frac{w_j \phi(x_i; \mu_j, \sigma_j)}{\hat{\eta}_{i,j}^{(t)}}}_{\text{前頁の } u_j \text{ とみなす}}$$

$$\hat{\eta}_{i,j}^{(t)} = \frac{\hat{w}_j^{(t)} \phi(x_i; \hat{\mu}_j^{(t)}, \hat{\sigma}_j^{(t)})}{\sum_{j'=1}^{m} \hat{w}_{j'}^{(t)} \phi(x_i; \hat{\mu}_{j'}^{(t)}, \hat{\sigma}_{j'}^{(t)})}$$

$$\hat{\eta}_{i,j}^{(t)} \geq 0, \sum_{j=1}^{m} \hat{\eta}_{i,j}^{(t)} = 1$$

Eステップの導出（続き）

$$\log L(\theta)=\sum_{i=1}^{n}\log\sum_{j=1}^{m}\hat{\eta}_{i,j}^{(t)}\frac{w_j\phi(x_i;\mu_j,\sigma_j)}{\hat{\eta}_{i,j}^{(t)}}$$

- ジェンセンの不等式より対数尤度の下界を得る：

$$\log L(\theta)\ge b^{(t)}(\theta)$$

$$b^{(t)}(\theta)=\sum_{i=1}^{n}\sum_{j=1}^{m}\hat{\eta}_{i,j}^{(t)}\log\frac{w_j\phi(x_i;\mu_j,\sigma_j)}{\hat{\eta}_{i,j}^{(t)}}$$

- $\log L(\hat{\theta}^{(t)})=b^{(t)}(\hat{\theta}^{(t)})$ を満たす：

$$b^{(t)}(\hat{\theta}^{(t)})=\sum_{i=1}^{n}\sum_{j=1}^{m}\hat{\eta}_{i,j}^{(t)}\log\frac{\hat{w}_j^{(t)}\phi(x_i;\hat{\mu}_j^{(t)},\hat{\sigma}_j^{(t)})}{\hat{\eta}_{i,j}^{(t)}}$$

$$\hat{\eta}_{i,j}^{(t)}=\frac{\hat{w}_j^{(t)}\phi(x_i;\hat{\mu}_j^{(t)},\hat{\sigma}_j^{(t)})}{\sum_{j'=1}^{m}\hat{w}_{j'}^{(t)}\phi(x_i;\hat{\mu}_{j'}^{(t)},\hat{\sigma}_{j'}^{(t)})}$$

$$=\sum_{i=1}^{n}\left(\sum_{j=1}^{m}\hat{\eta}_{i,j}^{(t)}\right)\log\sum_{j'=1}^{m}\hat{w}_{j'}^{(t)}\phi(x_i;\hat{\mu}_{j'}^{(t)},\hat{\sigma}_{j'}^{(t)})$$

$$\sum_{j=1}^{m}\hat{\eta}_{i,j}^{(t)}=1$$

$$=\sum_{i=1}^{n}\log\sum_{j=1}^{m}\hat{w}_j^{(t)}\phi(x_i;\hat{\mu}_j^{(t)},\hat{\sigma}_j^{(t)})=\log L(\hat{\theta}^{(t)})$$

Mステップの導出

$$b^{(t)}(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{m} \hat{\eta}_{i,j}^{(t)} \log w_j \phi(x_i; \mu_j, \sigma_j) - \sum_{i=1}^{n}\sum_{j=1}^{m} \hat{\eta}_{i,j}^{(t)} \log \hat{\eta}_{i,j}^{(t)}$$

- 下界を最大にするパラメータ値を求める:

$$\frac{\partial b^{(t)}}{\partial \gamma_j} = 0, \quad \frac{\partial b^{(t)}}{\partial \mu_j} = 0, \quad \frac{\partial b^{(t)}}{\partial \sigma_j} = 0$$

$$\hat{w}_j^{(t+1)} = \frac{1}{n}\sum_{i=1}^{n} \hat{\eta}_{i,j}^{(t)}$$

$$\hat{\mu}_j^{(t+1)} = \frac{\sum_{i=1}^{n} \hat{\eta}_{i,j}^{(t)} x_i}{\sum_{i'=1}^{n} \hat{\eta}_{i',j}^{(t)}}$$

$$\hat{\sigma}_j^{(t+1)} = \sqrt{\frac{1}{d} \frac{\sum_{i=1}^{n} \hat{\eta}_{i,j}^{(t)} (x_i - \hat{\mu}_j^{(t)})^T (x_i - \hat{\mu}_j^{(t)})}{\sum_{i'=1}^{n} \hat{\eta}_{i',j}^{(t)}}}$$

EMアルゴリズムの一般形

- 不完全データに対する最尤推定：
  - モデル $q(z;\theta)$ のパラメータ θ を最尤推定したい
  - 完全な訓練標本 $\{z_i \mid z_i = (x_i, y_i)\}_{i=1}^{n}$ のうち，その一部 $\{x_i\}_{i=1}^{n}$ しか観測できない．
- Eステップ：現在のパラメータ $\hat{\theta}^{(t)}$を用いて観測されない部分 $\{y_i\}_{i=1}^{n}$ を推定し，対数尤度の期待値(expectation)を計算

$$Q(\theta;\hat{\theta}^{(t)}) = \sum_{i=1}^{n} \int \hat{p}(y_i \mid x_i;\hat{\theta}^{(t)}) \log q(x_i, y_i;\theta) dy_i$$

- Mステップ： $Q(\theta;\hat{\theta}^{(t)})$ を最大化(maximization)するように θ を更新

$$\hat{\theta}^{(t+1)} = \arg\max_{\theta} Q(\theta;\hat{\theta}^{(t)})$$

ガウス混合モデルの場合

- $\hat{\eta}_{i,j}^{(t)}$は，標本 x_i が j番目の混合から出てくる確率と解釈できる．

$$\hat{\eta}_{i,j}^{(t)} = \frac{\hat{w}_j^{(t)}\phi(x_i;\hat{\mu}_j^{(t)},\hat{\sigma}_j^{(t)})}{\sum_{j'=1}^{m}\hat{w}_{j'}^{(t)}\phi(x_i;\hat{\mu}_{j'}^{(t)},\hat{\sigma}_{j'}^{(t)})}$$

- x_i が出てきた“本当”の混合の番号を y_i とする．
- (x_i, y_i) が分かるとき，完全データに対する対数尤度は

$$L(\theta) = \sum_{i=1}^{n} \log w_{y_i}\phi(x_i;\mu_{y_i},\sigma_{y_i})$$

- Eステップ：対数尤度の y_i に関する期待値を計算

$$Q(\theta;\hat{\theta}^{(t)}) = \sum_{i=1}^{n}\sum_{j=1}^{m}\eta_{i,j}^{(t)} \log w_j\phi(x_i;\mu_j,\sigma_j)$$