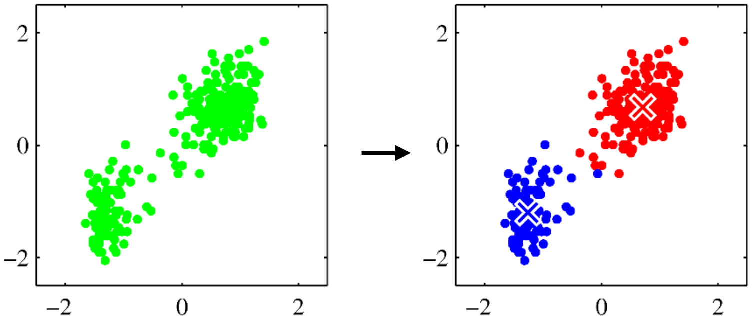
機械学習 Machine Learning

混合モデルとEM Mixture Models and EM

福嶋 誠 Makoto Fukushima

情報科学部
School of Informatics and Data Science

K-meansクラスタリング K-means clustering



2値指示変数
Binary indicator variables

$$r_{nk} \in \{0,1\} \qquad k = 1, \ldots, K$$

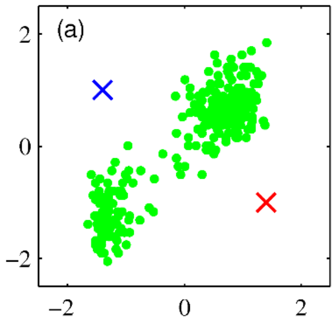
$$\mathbf{x}_n \rightarrow k \quad \Rightarrow \quad \begin{array}{ll} r_{nk} = 1 & \\ r_{nj} = 0 & j \neq k \end{array}$$

誤差関数 Error function

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2$$

K-meansクラスタリング K-means clustering

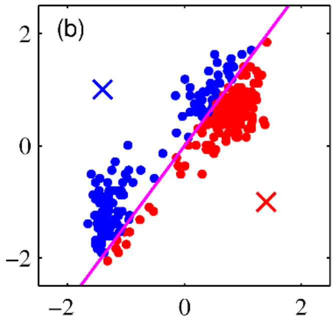
はじめに $\{\boldsymbol{\mu}_k\}$ の初期値を選ぶ.
First we choose some initial values for the $\{\boldsymbol{\mu}_k\}$.



<u>Eステップ E step</u>

次に，最初のフェーズで $\{\boldsymbol{\mu}_k\}$を固定しつつ $\{r_{nk}\}$ についてJ を最小化する.
Then in the first step, we minimize J with respect to the $\{r_{nk}\}$, keeping the $\{\boldsymbol{\mu}_k\}$ fixed.

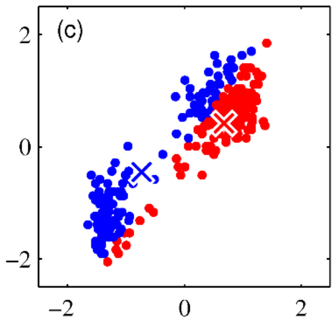
$$r_{nk} = \begin{cases} 1, & \text{if } k = \arg\min_j \|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2 \text{ のとき} \\ 0, & \text{otherwise. それ以外} \end{cases}$$



K-meansクラスタリング K-means clustering

Mステップ M step

第二フェーズでは, $\{r_{nk}\}$ を固定しつつ $\{\boldsymbol{\mu}_k\}$ について J を最小化する.
In the second step, we minimize J with respect to the $\{\boldsymbol{\mu}_k\}$, keeping $\{r_{nk}\}$ fixed.

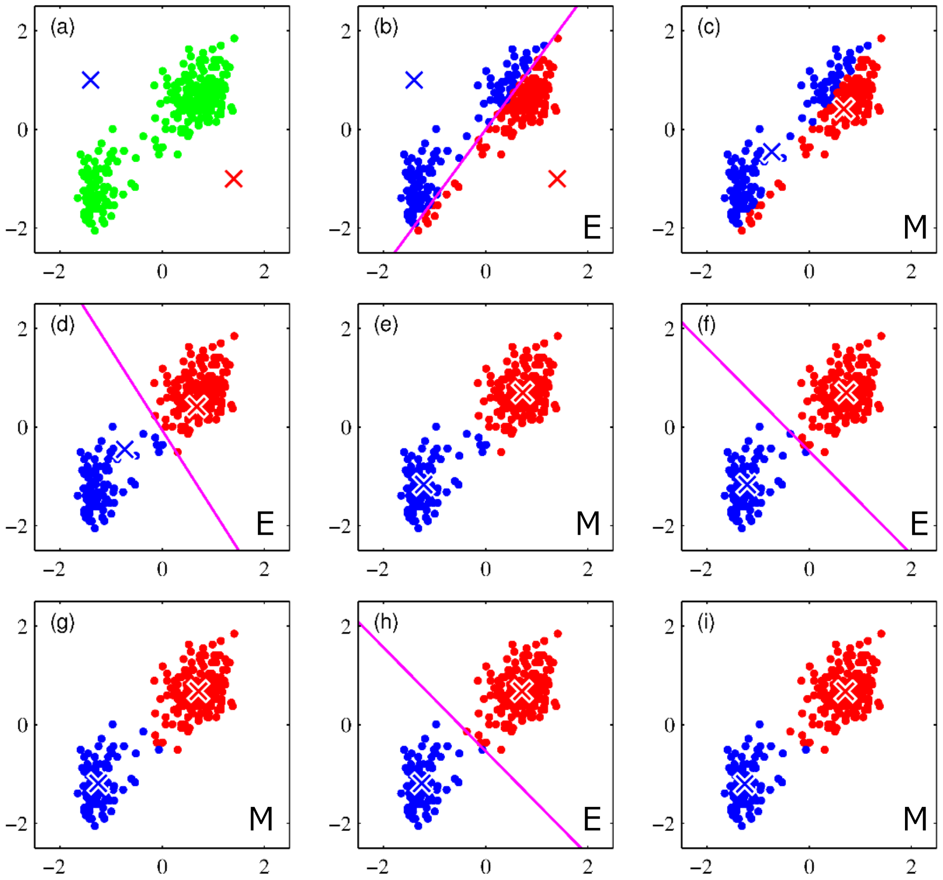


$$2\sum_{n=1}^{N} r_{nk}(\mathbf{x}_n - \boldsymbol{\mu}_k) = 0$$

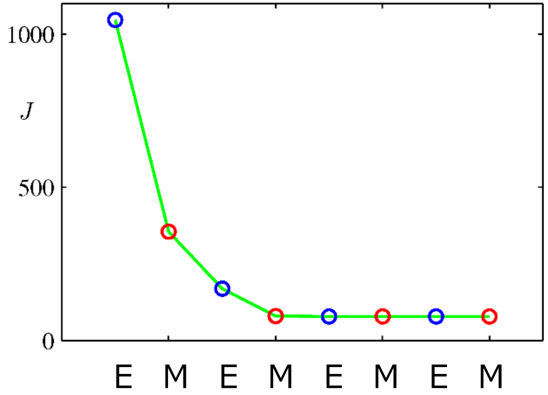
$$\Rightarrow \quad \boldsymbol{\mu}_k = \frac{\sum_n r_{nk}\mathbf{x}_n}{\sum_n r_{nk}}$$

このような二段階最適化を収束するまで繰り返す.
This two-step optimization is then repeated until convergence.

K-meansクラスタリング *K*-means clustering



誤差関数 J のプロット
Plot of the error function J



K-meansクラスタリング K-means clustering

Input: Initial prototype vectors $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$

Data set $\mathbf{x}_1, \ldots, \mathbf{x}_N$

Output: Final prototype vectors $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K$

```
{r_nk ← 0} // Initially set all assignments to zero
repeat
    {r_nk^(old)} ← {r_nk}
    // Update assignments
    for N ∈ {1, ..., N} do
        k ← arg min_j ||x_n − μ_j||^2
        r_nk ← 1
        r_nj ← 0,   j ∈ {1, ..., K}, j ≠ k
    end for
    // Update prototype vectors
    for k ∈ {1, ..., K} do
        μ_k ← Σ_n r_nk x_n / Σ_n r_nk
    end for
until {r_nk} = {r_nk^(old)} // Assignments unchanged
return μ_1, ..., μ_K, {r_nk}
```

$\{r_{nk} \leftarrow 0\}$ // Initially set all assignments to zero

repeat

$\{r_{nk}^{(\text{old})}\} \leftarrow \{r_{nk}\}$

// Update assignments

for $N \in \{1, \ldots, N\}$ **do**

$k \leftarrow \arg\min_j \|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2$

$r_{nk} \leftarrow 1$

$r_{nj} \leftarrow 0, \quad j \in \{1, \ldots, K\},\ j \neq k$

end for

// Update prototype vectors

for $k \in \{1, \ldots, K\}$ **do**

$\boldsymbol{\mu}_k \leftarrow \sum_n r_{nk}\mathbf{x}_n / \sum_n r_{nk}$

end for

until $\{r_{nk}\} = \{r_{nk}^{(\text{old})}\}$ // Assignments unchanged

return $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K, \{r_{nk}\}$

K-meansクラスタリング K-means clustering

逐次的な更新 Sequential update

$$\boldsymbol{\mu}_k^{\text{new}} = \boldsymbol{\mu}_k^{\text{old}} + \frac{1}{N_k}(\mathbf{x}_n - \boldsymbol{\mu}_k^{\text{old}})$$

N_k： $\boldsymbol{\mu}_k$ の更新に用いたデータ点の数
the number of data points that have been used to update $\boldsymbol{\mu}_k$

K-meansクラスタリング K-means clustering
画像分割 Image segmentation

43,248 bits | 86,472 bits | 173,040 bits | 1,036,800 bits

混合ガウス分布 Mixtures of Gaussians

混合ガウス分布 Gaussian mixture distribution

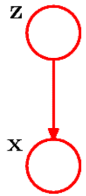
$$p(\mathbf{x}) = \boxed{\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \qquad 0 \leqslant \pi_k \leqslant 1 \qquad \sum_{k=1}^{K} \pi_k = 1$$

2値確率変数 Binary random variable $\mathbf{z}$ $\quad z_k \in \{0, 1\} \quad \sum_k z_k = 1$

$$p(z_k = 1) = \pi_k \qquad p(\mathbf{z}) = \prod_{k=1}^{K} \pi_k^{z_k}$$

条件付き分布 Conditional distribution

$$p(\mathbf{x}|z_k = 1) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \qquad p(\mathbf{x}|\mathbf{z}) = \prod_{k=1}^{K} \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k}$$



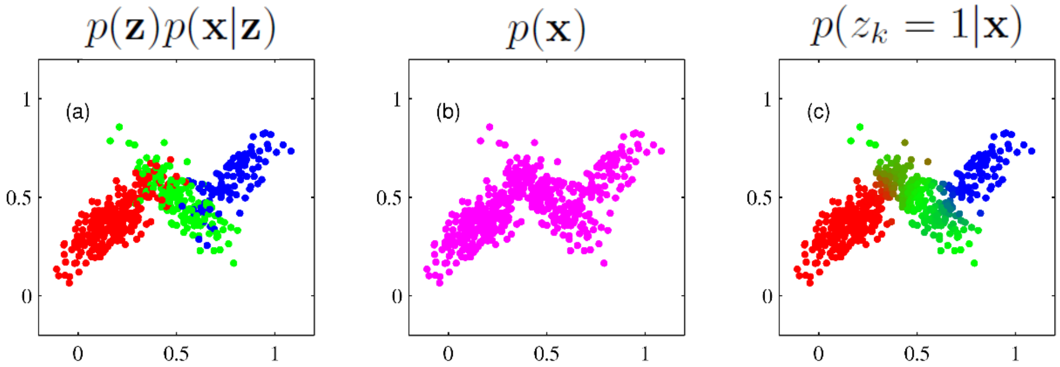
周辺分布 Marginal distribution

$$p(\mathbf{x}) = \sum_{\mathbf{z}} p(\mathbf{z}) p(\mathbf{x}|\mathbf{z}) = \sum_{\mathbf{z}} \prod_{k=1}^{K} (\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k))^{z_k} = \sum_{j=1}^{K} \prod_{k=1}^{K} (\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k))^{I_{kj}} = \boxed{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

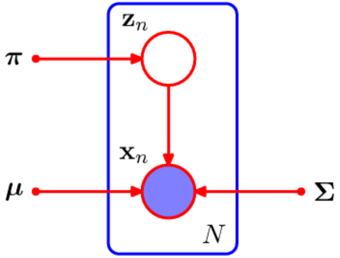
混合ガウス分布 Mixtures of Gaussians

負担率 Responsibility

$$\gamma(z_k) \equiv p(z_k = 1|\mathbf{x}) = \frac{p(z_k = 1)p(\mathbf{x}|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(\mathbf{x}|z_j = 1)} = \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

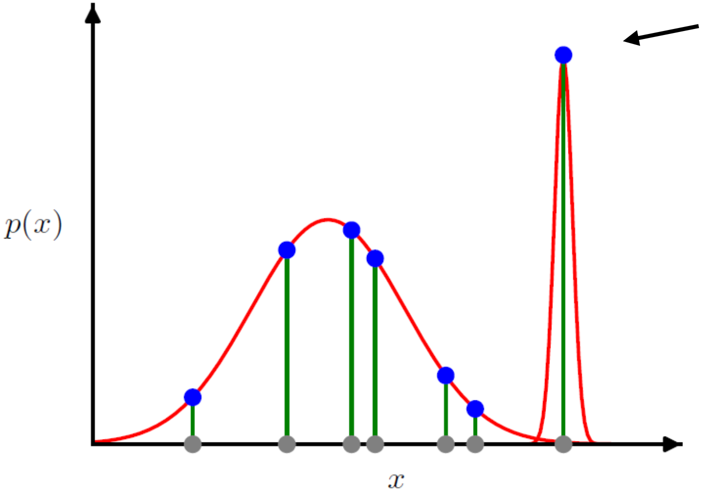


混合ガウス分布 Mixtures of Gaussians
尤度関数 Likelihood function



対数尤度関数 Log likelihood function

$$\ln p(\mathbf{X}|\boldsymbol{\pi},\boldsymbol{\mu},\boldsymbol{\Sigma}) = \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k) \right\}$$



1つのガウス分布の要素が
特定のデータ点でつぶれるとき,
When one of the Gaussian components
collapses onto a specific data point,

$$\boldsymbol{\mu}_j = \mathbf{x}_n \qquad \mathcal{N}(\mathbf{x}_n|\mathbf{x}_n,\sigma_j^2\mathbf{I}) = \frac{1}{(2\pi)^{1/2}}\frac{1}{\sigma_j^D}$$
$$\boldsymbol{\Sigma}_k = \sigma_k^2\mathbf{I}$$

$\sigma_j \to 0$ で尤度が無限大に発散してしまう.
The likelihood goes to infinity when $\sigma_j \to 0$.

混合ガウス分布 Mixtures of Gaussians
最尤推定 Maximum likelihood

EMアルゴリズム EM (expectation-maximization) algorithm　Dempster et al. (1977)

混合ガウスモデルの文脈における比較的大雑把な取り扱いによる説明
An explanation with a relatively informal treatment in the context of a Gaussian mixture model

$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ の $\boldsymbol{\mu}_k$ に関する導関数を0とおくと
By setting the derivatives of $\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with respect to $\boldsymbol{\mu}_k$ to zero,

$$0 = \sum_{n=1}^{N} \underbrace{\frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}}_{\gamma(z_{nk})} \boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n - \boldsymbol{\mu}_k)$$

$\Rightarrow$ $\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n$　ただし where $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$

混合ガウス分布 Mixtures of Gaussians
最尤推定 Maximum likelihood

EMアルゴリズム EM (expectation-maximization) algorithm Dempster et al. (1977)

混合ガウスモデルの文脈における比較的大雑把な取り扱いによる説明
An explanation with a relatively informal treatment in the context of a Gaussian mixture model

$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ の $\boldsymbol{\Sigma}_k$ に関する導関数を0とおくと
By setting the derivatives of $\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$ with respect to $\boldsymbol{\Sigma}_k$ to zero,

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}}$$

ただし where $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$

混合ガウス分布 Mixtures of Gaussians
最尤推定 Maximum likelihood

EMアルゴリズム EM (expectation-maximization) algorithm　Dempster et al. (1977)

混合ガウスモデルの文脈における比較的大雑把な取り扱いによる説明
An explanation with a relatively informal treatment in the context of a Gaussian mixture model

$\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right)$ の π_k に関する導関数を0とおくと　$\boxed{\sum_{k=1}^{K} \pi_k = 1}$

By setting the derivatives of $\ln p(\mathbf{X}|\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) + \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right)$ with respect to π_k to zero,

$$0 = \sum_{n=1}^{N} \frac{\mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} + \lambda$$

⇨ $\lambda = -N \quad \pi_k = \dfrac{N_k}{N}$　ただし where $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$

混合ガウス分布 Mixtures of Gaussians
最尤推定 Maximum likelihood

Eステップ E step

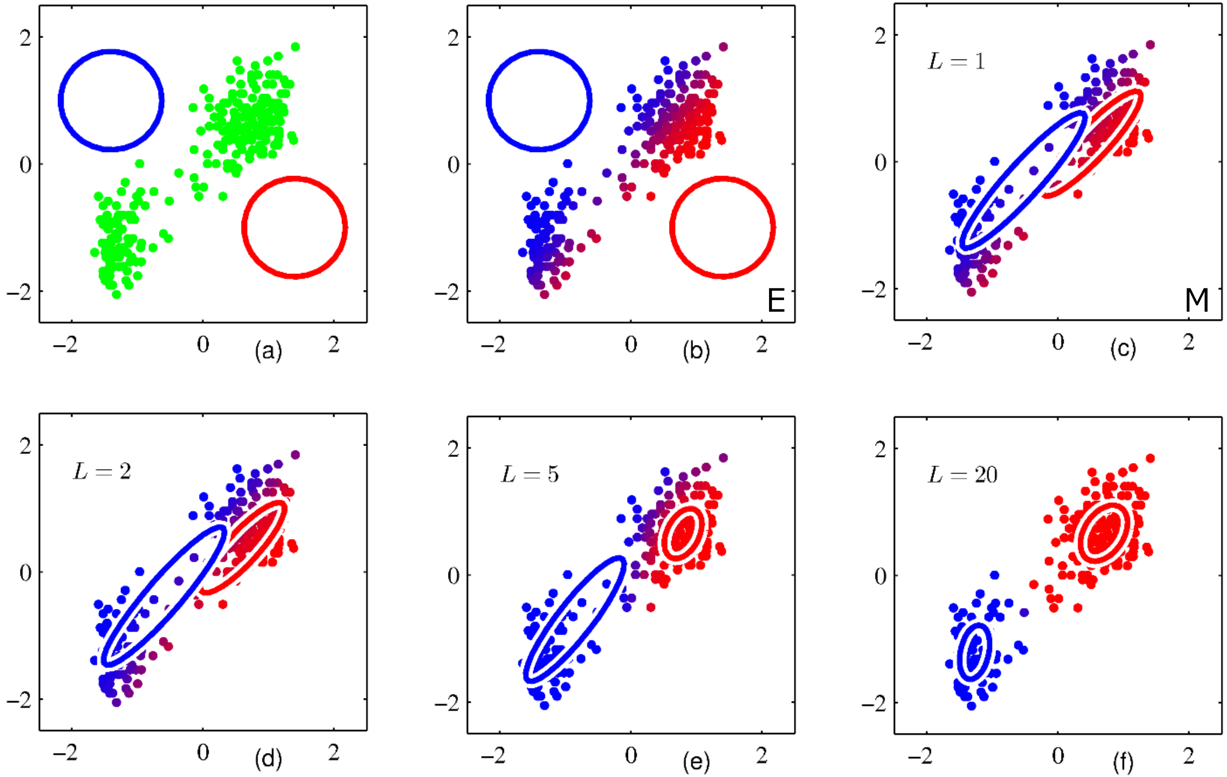
$$\gamma(z_k) \equiv p(z_k = 1|\mathbf{x}) = \frac{p(z_k = 1)p(\mathbf{x}|z_k = 1)}{\sum_{j=1}^{K} p(z_j = 1)p(\mathbf{x}|z_j = 1)} = \frac{\pi_k \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

Mステップ M step

$$N_k = \sum_{n=1}^{N} \gamma(z_{nk}) \qquad \boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{n=1}^{N} \gamma(z_{nk})\mathbf{x}_n \qquad \pi_k = \frac{N_k}{N}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k}\sum_{n=1}^{N} \gamma(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}}$$

混合ガウス分布 Mixtures of Gaussians
最尤推定 Maximum likelihood



混合ガウス分布 Mixtures of Gaussians
最尤推定 Maximum likelihood

Input: Initial model parameters $\{\boldsymbol{\mu}_k\}, \{\boldsymbol{\Sigma}_k\}, \{\pi_k\}$

Data set $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$

Output: Final model parameters $\{\boldsymbol{\mu}_k\}, \{\boldsymbol{\Sigma}_k\}, \{\pi_k\}$

---

repeat

`// E step`

for $n \in \{1, \ldots, N\}$ **do**

for $k \in \{1, \ldots, K\}$ **do**

$$\gamma(z_{nk}) \leftarrow \frac{\pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)}$$

end for

end for

`// M step`

for $k \in \{1, \ldots, K\}$ **do**

$$N_k \leftarrow \sum_{n=1}^{N} \gamma(z_{nk})$$

$$\boldsymbol{\mu}_k \leftarrow \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n$$

$$\boldsymbol{\Sigma}_k \leftarrow \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}}$$

$$\pi_k \leftarrow \frac{N_k}{N}$$

end for

`// Log likelihood`

$$\mathcal{L} \leftarrow \sum_{n=1}^{N} \ln \left\{ \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}_n | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}$$

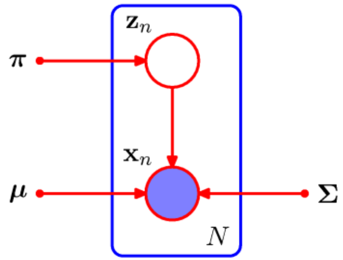
until convergence

return $\{\boldsymbol{\mu}_k\}, \{\boldsymbol{\Sigma}_k\}, \{\pi_k\}$

EMアルゴリズム Expectation–maximization algorithm

対数尤度関数 Log likelihood function

$$\ln p(\mathbf{X}|\boldsymbol{\theta}) = \ln \left\{ \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) \right\}$$



完全データ対数尤度関数 The complete-data log likelihood

$$\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$$

$\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$ の $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\mathrm{old}})$ に関する期待値
Expected value of $\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$ under $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\mathrm{old}})$

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\mathrm{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$$

EMアルゴリズム Expectation–maximization algorithm

Eステップ E step

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\mathrm{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$$

Mステップ M step

$$\boldsymbol{\theta}^{\mathrm{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\mathrm{old}})$$

EMアルゴリズム Expectation–maximization algorithm

Input: Joint distribution $p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$

Initial parameters $\boldsymbol{\theta}^{\text{old}}$

Data set $\mathbf{x}_1, \ldots, \mathbf{x}_N$

Output: Final parameters $\boldsymbol{\theta}$

---

repeat

$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) \leftarrow \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$ `// E step`

$\boldsymbol{\theta}^{\text{new}} \leftarrow \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ `// M step`

$\mathcal{L} \leftarrow p(\mathbf{X}|\boldsymbol{\theta}^{\text{new}})$ `// Evaluate log likelihood`

$\boldsymbol{\theta}^{\text{old}} \leftarrow \boldsymbol{\theta}^{\text{new}}$ `// Update the parameters`

until convergence

return $\boldsymbol{\theta}^{\text{new}}$

EMアルゴリズム Expectation–maximization algorithm
混合ガウス分布 Gaussian mixtures

完全データ対数尤度関数の期待値
Expected value of the complete-data log likelihood

$$\mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})] = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \left\{ \ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}$$

$$\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) = \sum_{n=1}^{N} \sum_{k=1}^{K} z_{nk} \left\{ \ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right\}$$

$$\mathbb{E}[z_{nk}] = \frac{\sum_{\mathbf{z}_n} z_{nk} \prod_{k'} \left[\pi_{k'} \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_{k'}, \boldsymbol{\Sigma}_{k'}) \right]^{z_{nk'}}}{\sum_{\mathbf{z}_n} \prod_{j} \left[\pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j) \right]^{z_{nj}}} = \frac{\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} = \gamma(z_{nk})$$

$$p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi}) \propto \prod_{n=1}^{N} \prod_{k=1}^{K} \left[\pi_k \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right]^{z_{nk}}$$

<u>Eステップ E step</u>

EMアルゴリズム Expectation–maximization algorithm 混合ガウス分布 Gaussian mixtures

完全データ対数尤度関数の期待値
Expected value of the complete-data log likelihood

$$\mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})] = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \{\ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\}$$

上式を最大化するパラメータの更新式
Update equations of the parameters that maximize the above equation

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n \qquad \pi_k = \frac{N_k}{N}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}}$$

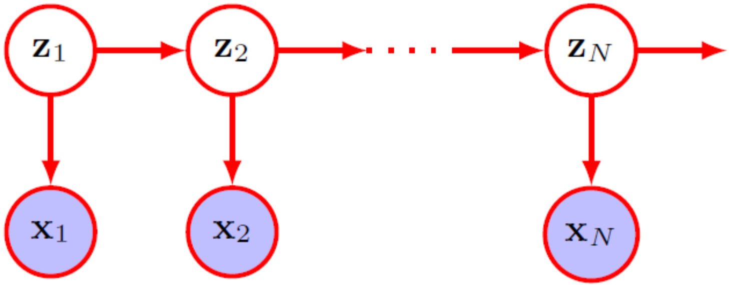
ただし
where $N_k = \sum_{n=1}^{N} \gamma(z_{nk})$

<u>Mステップ M step</u>

先に導いた混合ガウス分布に関するEMアルゴリズムと正確に一致.
This is precisely the EM algorithm for Gaussian mixtures as derived before.

EMアルゴリズム Expectation–maximization algorithm
混合ガウス分布 Gaussian mixtures

潜在変数をマルコフ連鎖によってつなげることで，混合モデルは隠れマルコフモデル（HMM）に拡張することができる.
The mixture model can be extended by connecting the latent variables in a Markov chain to give a hidden Markov model (HMM).



EMアルゴリズム Expectation–maximization algorithm
K-meansとの関連 Relation to *K*-means

K-means: データ点を各クラスターにハードに割り当てる.
A hard assignment of data points to clusters

$$p(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) = \frac{1}{(2\pi\epsilon)^{D/2}} \exp\left\{-\frac{1}{2\epsilon}\|\mathbf{x}-\boldsymbol{\mu}_k\|^2\right\} \qquad \gamma(z_{nk}) = \frac{\pi_k \exp\left\{-\|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2/2\epsilon\right\}}{\sum_j \pi_j \exp\left\{-\|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2/2\epsilon\right\}}$$

$$\epsilon \to 0 \quad \Rightarrow \quad \left\{ \begin{array}{l} \gamma(z_{nk}) \to r_{nk} = \begin{cases} 1, & \text{if } k = \arg\min_j \|\mathbf{x}_n - \boldsymbol{\mu}_j\|^2 \text{ のとき} \\ 0, & \text{otherwise. それ以外} \end{cases} \\ \boldsymbol{\mu}_k = \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})\mathbf{x}_n \to \boldsymbol{\mu}_k = \frac{\sum_n r_{nk}\mathbf{x}_n}{\sum_n r_{nk}} \\ \mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})] \to -\frac{1}{2}\underline{\sum_{n=1}^{N}\sum_{k=1}^{K} r_{nk}\|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2}_{\,= J} + \text{const} \end{array} \right.$$

エビデンスの下限 Evidence lower bound

さらに一般的な観点からのEMアルゴリズムを紹介する.
Introduce an even more general perspective on the EM algorithm.

尤度関数 Likelihood function $p(\mathbf{X}|\boldsymbol{\theta}) = \sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$

$$\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) = \ln p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}) + \ln p(\mathbf{X}|\boldsymbol{\theta})$$

$$\Rightarrow \quad \ln p(\mathbf{X}|\boldsymbol{\theta}) = \mathcal{L}(q, \boldsymbol{\theta}) + \mathrm{KL}(q\|p)$$

ただし
where

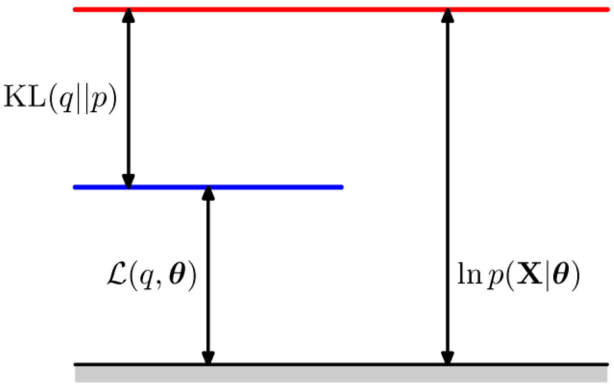
$$\mathcal{L}(q, \boldsymbol{\theta}) = \sum_{\mathbf{Z}} q(\mathbf{Z}) \ln \left\{ \frac{p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})}{q(\mathbf{Z})} \right\}$$

$$\mathrm{KL}(q\|p) = -\sum_{\mathbf{Z}} q(\mathbf{Z}) \ln \left\{ \frac{p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta})}{q(\mathbf{Z})} \right\}$$

$\mathrm{KL}(q\|p) \geqslant 0$

$\Rightarrow \quad \mathcal{L}(q, \boldsymbol{\theta}) \leqslant \ln p(\mathbf{X}|\boldsymbol{\theta})$

下限 Lower bound



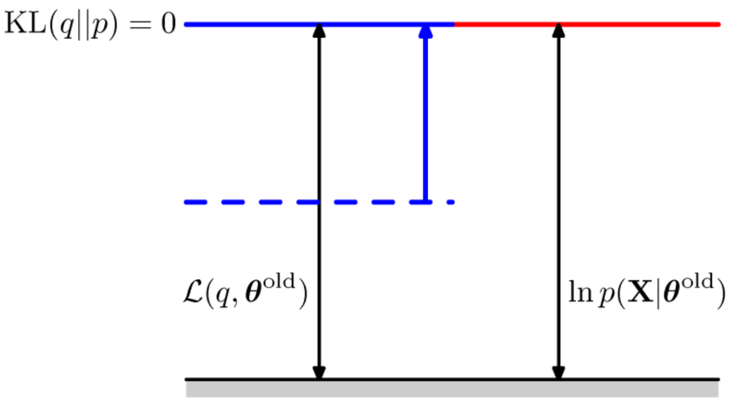
エビデンスの下限 Evidence lower bound
EM再訪 EM revisited

Eステップ E step

$\mathcal{L}(q, \boldsymbol{\theta}^{\text{old}})$ を $q(\mathbf{Z})$ について最大化する（$\boldsymbol{\theta}^{\text{old}}$ は固定したまま）．
$\mathcal{L}(q, \boldsymbol{\theta}^{\text{old}})$ is maximized with respect to $q(\mathbf{Z})$ while holding $\boldsymbol{\theta}^{\text{old}}$ fixed.

$q(\mathbf{Z})$ が $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$ と等しいとき（KL距離が消えるとき）
When $q(\mathbf{Z})$ is equal to $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$ (when the KL divergence vanishes)

$\mathcal{L}(q, \boldsymbol{\theta}^{\text{old}})$ は最大となる.
$\mathcal{L}(q, \boldsymbol{\theta}^{\text{old}})$ is maximized.

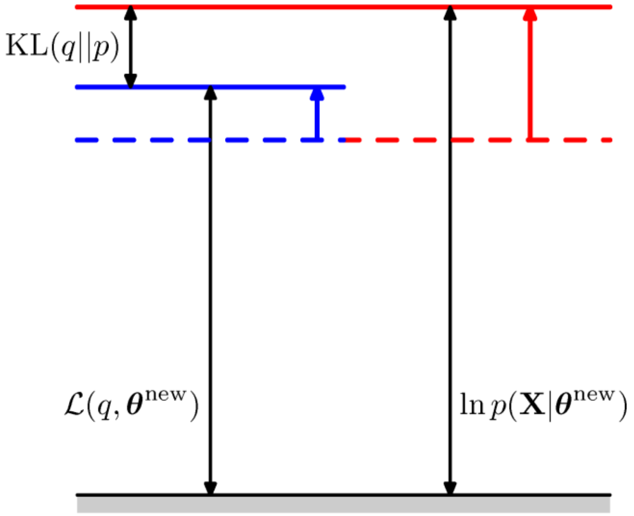


エビデンスの下限 Evidence lower bound
EM再訪 EM revisited

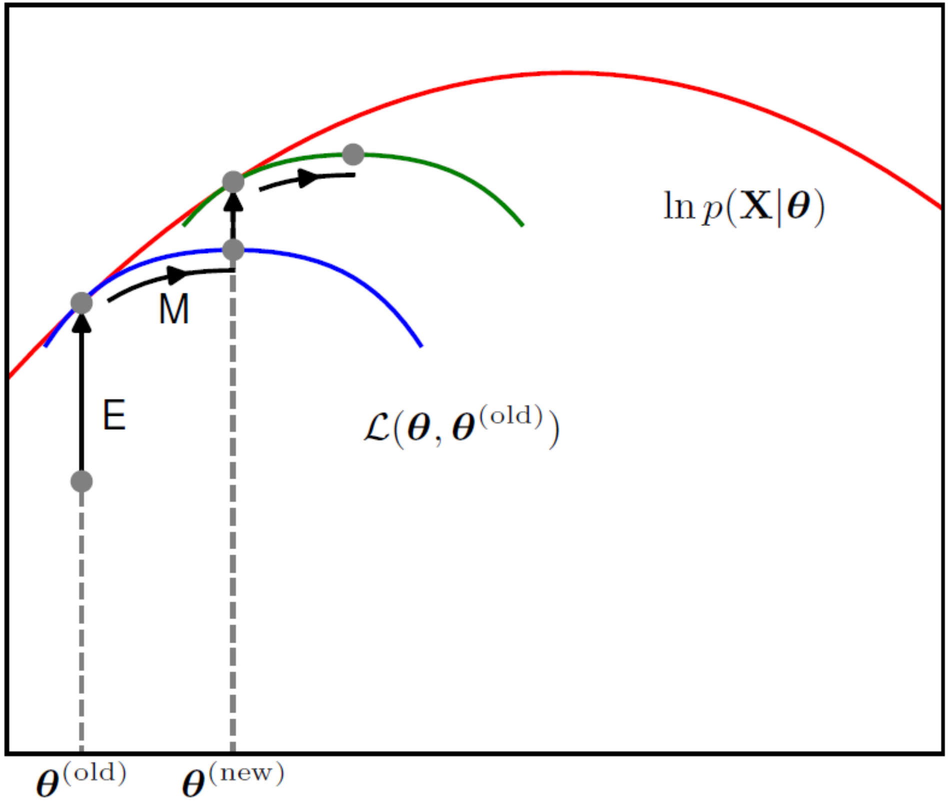
Mステップ M step

$\mathcal{L}(q,\boldsymbol{\theta})$ を $\boldsymbol{\theta}$ について最大化し（$q(\mathbf{Z})$ は固定したまま），$\boldsymbol{\theta}^{\text{new}}$ を得る.
$\mathcal{L}(q,\boldsymbol{\theta})$ is maximized with respect to $\boldsymbol{\theta}$ to give $\boldsymbol{\theta}^{\text{new}}$ while holding $q(\mathbf{Z})$ fixed.

$$
\begin{aligned}
\mathcal{L}(q,\boldsymbol{\theta}) &= \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X},\mathbf{Z}|\boldsymbol{\theta}) \\
&\quad - \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{Z}|\mathbf{X},\boldsymbol{\theta}^{\text{old}}) \\
&= \mathcal{Q}(\boldsymbol{\theta},\boldsymbol{\theta}^{\text{old}}) + \text{const}
\end{aligned}
$$



エビデンスの下限 Evidence lower bound
EM再訪 EM revisited



エビデンスの下限 Evidence lower bound
i.i.d.データ Independent and identically distributed data

i.i.d.データ集合を仮定している場合,
For the case of an i.i.d. dataset,

$$\mathbf{X}:\ \{\mathbf{x}_n\} \qquad n = 1, \ldots, N$$
$$\mathbf{Z}:\ \{\mathbf{z}_n\}$$

$$p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}) = \frac{p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})}{\sum_{\mathbf{Z}} p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})} = \frac{\prod_{n=1}^{N} p(\mathbf{x}_n, \mathbf{z}_n|\boldsymbol{\theta})}{\sum_{\mathbf{Z}} \prod_{n=1}^{N} p(\mathbf{x}_n, \mathbf{z}_n|\boldsymbol{\theta})} = \prod_{n=1}^{N} p(\mathbf{z}_n|\mathbf{x}_n, \boldsymbol{\theta})$$

エビデンスの下限 Evidence lower bound
パラメータ事前分布 Parameter priors

パラメータ事前分布 $p(\boldsymbol{\theta})$ を導入したモデルに対しては,
For models in which we have introduce a prior $p(\boldsymbol{\theta})$ over the parameters,

$$\ln p(\boldsymbol{\theta}|\mathbf{X}) = \ln p(\boldsymbol{\theta}, \mathbf{X}) - \ln p(\mathbf{X})$$

$$\boxed{\ln p(\mathbf{X}|\boldsymbol{\theta}) = \mathcal{L}(q, \boldsymbol{\theta}) + \mathrm{KL}(q\|p)}$$

$$\begin{aligned}\ln p(\boldsymbol{\theta}|\mathbf{X}) &= \boxed{\mathcal{L}(q, \boldsymbol{\theta}) + \mathrm{KL}(q\|p)} + \ln p(\boldsymbol{\theta}) - \ln p(\mathbf{X}) \\ &\geqslant \mathcal{L}(q, \boldsymbol{\theta}) + \ln p(\boldsymbol{\theta}) - \ln p(\mathbf{X})\end{aligned}$$

エビデンスの下限 Evidence lower bound
一般化EM Generalized EM

一般化EMアルゴリズムは手に負えないMステップの計算に対応する.
The generalized EM algorithm addresses the problem of an intractable M step.

<u>例 Examples</u>

- 勾配ベースの繰り返し最適化アルゴリズムをMステップで用いる.
  Use gradient-based iterative optimization algorithms during the M step.
- パラメータの組をいくつかのグループに分割し，Mステップでは，各グループに属するパラメータを固定しながら順番に最適化する.
  The parameters are partitioned into groups and the M step broken down into multiple steps each of which involves optimizing one of the groups with the reminder held fixed.

エビデンスの下限 Evidence lower bound
逐次EM Sequential EM

$$\mathbf{X}: \{\mathbf{x}_n\} \quad \mathbf{Z}: \{\mathbf{z}_n\} \qquad n = 1, \ldots, N$$

混合ガウス分布に対する逐次型のEMアルゴリズム
An incremental form of the EM algorithm for a mixture of Gaussians

<u>Eステップ E step</u>

負担率を1つのデータ点 $\mathbf{x}_m$ のみについて更新する.
Re-evaluate the responsibilities for one data point $\mathbf{x}_m$.

$$\gamma^{\text{old}}(z_{mk}) \rightarrow \gamma^{\text{new}}(z_{mk})$$

<u>Mステップ M step</u>

$$\boldsymbol{\mu}_k^{\text{new}} = \boldsymbol{\mu}_k^{\text{old}} + \left(\frac{\gamma^{\text{new}}(z_{mk}) - \gamma^{\text{old}}(z_{mk})}{N_k^{\text{new}}}\right)\left(\mathbf{x}_m - \boldsymbol{\mu}_k^{\text{old}}\right)$$

$$N_k^{\text{new}} = N_k^{\text{old}} + \gamma^{\text{new}}(z_{mk}) - \gamma^{\text{old}}(z_{mk})$$

$$\boldsymbol{\mu}_k^{\text{old}} = \frac{1}{N_k^{\text{old}}} \sum_n \gamma^{\text{old}}(z_{nk})\mathbf{x}_n$$

$$N_k^{\text{old}} = \sum_n \gamma^{\text{old}}(z_{nk})$$

エビデンスの下限 Evidence lower bound
逐次EM Sequential EM

$$N_k^{\text{old}} = \sum_n \gamma^{\text{old}}(z_{nk})$$

$$\Rightarrow \quad N_k^{\text{new}} = \sum_{n \neq m} \gamma^{\text{old}}(z_{nk}) + \gamma^{\text{new}}(z_{mk})$$

$$= N_k^{\text{old}} + \gamma^{\text{new}}(z_{mk}) - \gamma^{\text{old}}(z_{mk})$$

Mステップ M step

エビデンスの下限 Evidence lower bound
逐次EM Sequential EM

$$N_k^{\text{old}} = \sum_n \gamma^{\text{old}}(z_{nk}) \qquad \boldsymbol{\mu}_k^{\text{old}} = \frac{1}{N_k^{\text{old}}} \sum_n \gamma^{\text{old}}(z_{nk})\mathbf{x}_n$$

$$\Rightarrow \quad \boldsymbol{\mu}_k^{\text{new}} = \frac{1}{N_k^{\text{new}}}\left(\sum_{n\neq m} \gamma^{\text{old}}(z_{nk})\mathbf{x}_n + \gamma^{\text{new}}(z_{mk})\mathbf{x}_m\right)$$

$$= \frac{1}{N_k^{\text{new}}}\left(N_k^{\text{old}}\boldsymbol{\mu}_k^{\text{old}} - \gamma^{\text{old}}(z_{mk})\mathbf{x}_m + \gamma^{\text{new}}(z_{mk})\mathbf{x}_m\right)$$

$$= \frac{1}{N_k^{\text{new}}}\Big(\left(N_k^{\text{new}} - \gamma^{\text{new}}(z_{mk}) + \gamma^{\text{old}}(z_{mk})\right)\boldsymbol{\mu}_k^{\text{old}}$$

$$-\gamma^{\text{old}}(z_{mk})\mathbf{x}_m + \gamma^{\text{new}}(z_{mk})\mathbf{x}_m\Big)$$

$$= \boldsymbol{\mu}_k^{\text{old}} + \left(\frac{\gamma^{\text{new}}(z_{mk}) - \gamma^{\text{old}}(z_{mk})}{N_k^{\text{new}}}\right)(\mathbf{x}_m - \boldsymbol{\mu}_k^{\text{old}})$$

Mステップ M step

提出課題Ⅴ：混合モデルとEM
Assignment V: Mixture Models and EM

提出期限：**11月19日（火曜日）23:59:00 [日本標準時]**
Submission deadline: **November 19 (Tuesday) 23:59:00 [Japan Standard Time]**

提出課題は「一般」チャネルの「ファイル」にアップロードされます.
同チャネルに出現する通知のリンク先から解答を送信（提出）してください.

Assignments will be uploaded to "File" in the "General" channel. Send (submit) your answers via the link that will appear in the same channel.

- 全6回の課題への解答をもとに成績評価を行います.
  Your grades will be based on your answers to all six assignments.
- 解答送信後の解答再送信はできません.
  Once submitted, answers cannot be resubmitted.
- 提出期限は一切延長しません.
  The submission deadline will never be extended.