第9章
混合モデルとEM

Pattern Recognition and Machine Learning

2015年度　夏合宿

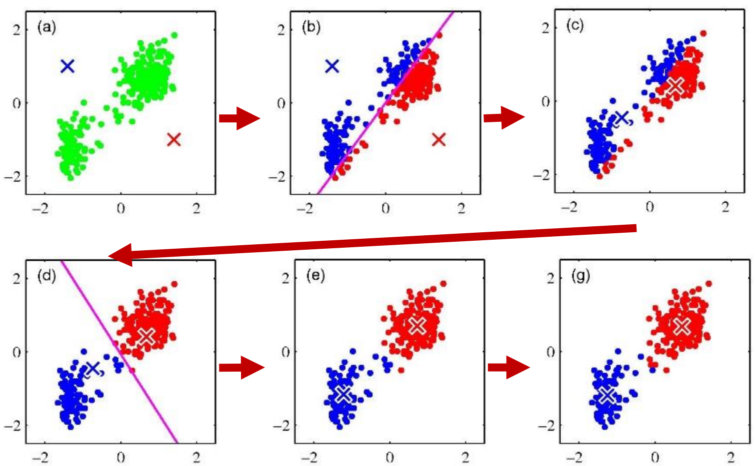
交通研B4　庄司惟

■目次

第9章（混合モデルとEM）

1. K-meansと混合ガウスEM

K-means クラスタリング



①点群が観測点（上図中では二次元ベクトル）

②クラスター＝データ点間距離が小さいグループ

③××がクラスター中心（プロトタイプと呼ばれるベクトルで表現）

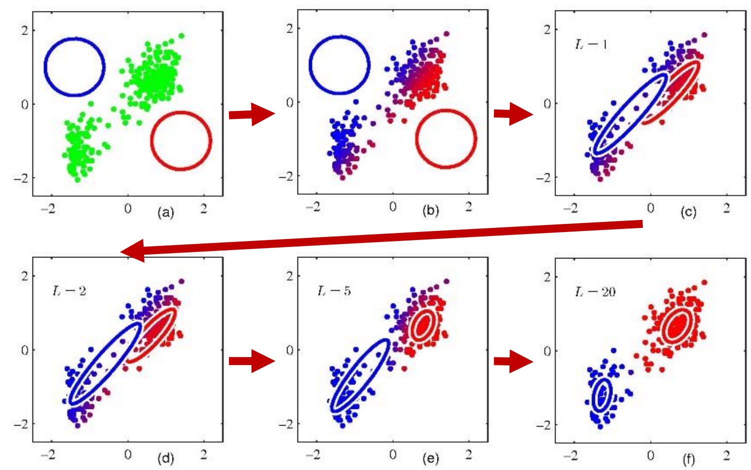
④各クラスターにデータ点を割り当て，割り当てられた点群からの総二乗距離が最小となるようにクラスター中心を移動し，再度データ点を割り当てなおす．これを繰り返す．

※観測点は必ず唯一つのクラスターに属する！！

緩和

極限

混合ガウス分布のEMアルゴリズム



①○○は，ガウス分布の標準偏差の等高線

②データ点の各分布に対する割り当ての度合い（＝負担率）を計算，負担率から分布形状（つまり，平均と分散）を調整，するとまた負担率が変化するので，これを繰り返す．

※観測点が各々の割合で各ガウス分布に属する！

Pattern Recognition and Machine Learning

2. K-meansクラスタリング

目的関数 $$J = \sum_{n=1}^{N}\sum_{k=1}^{K} \underline{r_{nk}}\,||\boldsymbol{x}_n - \underline{\boldsymbol{\mu}_k}||^2$$ の最小化

観測点

割り当ての
指示変数
(1 or 0)

クラスター中心

★目的関数を最小化する r_{nk} $\boldsymbol{\mu}_k$ を求める.

(Step1)
各観測点について，そこからの距離が最短であるクラスター中心を選ぶ.
指示変数 r_{nk} の値を更新.

(Step2)
r_{nk} を固定して目的関数を最小化する．単純な二次関数なので一階条件を解いて $\boldsymbol{\mu}_k$ の値を更新.

$$\boldsymbol{\mu}_k = \underline{\frac{\sum_{n=1}^{N} r_{nk}\boldsymbol{x}_n}{\sum_{n=1}^{N} r_{nk}}}$$

k番目のクラスターに割り当てられた全ての観測点の平均値（∴K-means !!）

混合ガウス分布とは（復習？）

$$p(\mathbf{x}) = \sum_{k=1}^{K} \pi_k N(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

素直なアプローチ（潜在変数用いない）を考察する.
パラメータ推定のために最大化する対数尤度関数は以下の通り.

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

対数の中にK個の要素に関する和が存在し，尤度関数の一階条件を解こうとしても，陽な解が得られない！！（K-meansとの決定的な違い）

とりあえず，素直なアプローチとして，K-means同様 π_k を固定し，パラメータ $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$ に関して一階条件を求めようとしてみよう．（無理っぽいけど）

3. 混合ガウス分布

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

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\boldsymbol{\Sigma}|^{1/2}} exp\left\{-\frac{1}{2}\underline{(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}\right\}$$

D：観測点ベクトルの次数

Σが対称行列→二次形式

$\boldsymbol{\mu}_k$ に関する一階条件は

$$\sum_{n=1}^{N}\frac{\pi_k\mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K}\pi_j\mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}\left\{\frac{\partial}{\partial\boldsymbol{\mu}_k}\left(-\frac{1}{2}(\boldsymbol{x}_n-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu}_k)\right)\right\}$$

付録1：二次形式の微分 参照

$$\underline{=\sum_{n=1}^{N}\frac{\pi_k\mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K}\pi_j\mathcal{N}(\boldsymbol{x}_n|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}\{-\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_n-\boldsymbol{\mu}_k)\}=0}$$

これ以上整理できない.

（∵ガウス分布の密度関数がそのまま残っている，nに関する和の形が残っている．）

3.2 二次形式のベクトル微分

$n \times n$ 正方対称行列 $\mathbf{A}$ に対する (各 a_{ij} は定数) 二次形式$\boldsymbol{x}'\mathbf{A}\boldsymbol{x}$のベクトル微分

$$
\frac{\partial \boldsymbol{x}'\mathbf{A}\boldsymbol{x}}{\partial \boldsymbol{x}} = \begin{pmatrix} \dfrac{\partial \sum\sum a_{ij}x_i x_j}{\partial x_1} \\ \dfrac{\partial \sum\sum a_{ij}x_i x_j}{\partial x_2} \\ \vdots \\ \dfrac{\partial \sum\sum a_{ij}x_i x_j}{\partial x_n} \end{pmatrix} = \begin{pmatrix} 2\sum a_{1j}x_j \\ 2\sum a_{2j}x_j \\ \vdots \\ 2\sum a_{nj}x_j \end{pmatrix}
$$

$$
= 2\mathbf{A}\boldsymbol{x}
$$

3. 混合ガウス分布

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

$$\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}}\frac{1}{|\boldsymbol{\Sigma}|^{1/2}} exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

D : 観測点ベクトルの次数

Σ_k に関する一階条件は

$$0 = \sum_{n=1}^{N}\frac{\pi_k}{\sum_{j=1}^{K}\pi_j N(\mathbf{x}_n|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}\left\{\frac{1}{(2\pi)^{D/2}}\exp\left\{-\frac{1}{2}(\mathbf{x}_n-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n-\boldsymbol{\mu}_k)\right\}\right\}\left\{\frac{-1}{2}\frac{1}{|\boldsymbol{\Sigma}_k|^{3/2}}\left(\frac{\partial}{\partial\boldsymbol{\Sigma}_k}|\boldsymbol{\Sigma}_k|\right)-\frac{1}{2}\frac{\partial}{\partial\boldsymbol{\Sigma}_k}(\mathbf{x}_n-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n-\boldsymbol{\mu}_k\ldots$$

$$\boldsymbol{\Sigma}_k^{-1} = \left(\boldsymbol{\Sigma}_k^{-1}\right)^T \qquad \frac{\partial}{\partial\boldsymbol{\Sigma}_k}|\boldsymbol{\Sigma}_k| = |\boldsymbol{\Sigma}_k|\left(\boldsymbol{\Sigma}_k^{-1}\right)^T \qquad \left\{\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n-\boldsymbol{\mu}_k)(\mathbf{x}_n-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}\right\}^T$$

$$\boldsymbol{\Sigma}_k = \frac{1}{\sum_{n=1}^{N}\frac{\pi_k N(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K}\pi_j N(\mathbf{x}_n|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}}\sum_{n=1}^{N}\frac{\pi_k N(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K}\pi_j N(\mathbf{x}_n|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}(\mathbf{x}_n-\boldsymbol{\mu}_k)(\mathbf{x}_n-\boldsymbol{\mu}_k)^T$$

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

π_k に関する一階条件は，

制約条件（パラメタπの総和が1)を考慮したラグランジュ関数を微分して，

$$0 = \sum_{n=1}^{N} \frac{N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j N(\mathbf{x}_n \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} + \lambda \longrightarrow 0 = \sum_{k=1}^{K}\left\{\sum_{n=1}^{N} \frac{\pi_k N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j N(\mathbf{x}_n \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} + \lambda \pi_k\right\}$$

両辺にπkを掛け，
全体をkに関して総和

$\sum_{k=1}^{K} \pi_k = 1$ より， $\lambda = -N$

$$0 = \sum_{n=1}^{N} \frac{\pi_k N(\mathbf{x}_n \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_{j=1}^{K} \pi_j N(\mathbf{x}_n \mid \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} - N \pi_k$$

まとめると・・・

$$0=\sum_{n=1}^{N}\frac{\pi_k N(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\underline{\sum_{j=1}^{K}\pi_j N(\mathbf{x}_n|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}}\{-\boldsymbol{\Sigma}_k^{-1}(\mathbf{x}_n-\boldsymbol{\mu}_k)\}$$

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

$$\boldsymbol{\Sigma}_k=\frac{1}{\sum_{n=1}^{N}\frac{\pi_k N(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\underline{\sum_{j=1}^{K}\pi_j N(\mathbf{x}_n|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}}}\sum_{n=1}^{N}\frac{\pi_k N(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\underline{\sum_{j=1}^{K}\pi_j N(\mathbf{x}_n|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}}(\mathbf{x}_n-\boldsymbol{\mu}_k)(\mathbf{x}_n-\boldsymbol{\mu}_k)^T$$

$$0=\sum_{n=1}^{N}\frac{\pi_k N(\mathbf{x}_n|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\underline{\sum_{j=1}^{K}\pi_j N(\mathbf{x}_n|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j)}}-N\pi_k$$

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

まとめると・・・

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

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

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} \gamma(z_{nk})$$

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

実際には複雑,

しかしγ（負担率）でまとめてしまうと各パラメタが陽な形で得られる.
⇒単純な繰り返し手続きで計算すればよいという示唆を得る.

繰り返し計算とは

①各パラメタの値に基づき，γ（負担率）を計算
②それを定数として代入し，各パラメタを求める
①②を繰り返す！

Eステップ

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

Mステップ

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

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

$$\pi_k = \frac{1}{N} \sum_{n=1}^{N} \gamma(z_{nk})$$

対数尤度計算
収束判定

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

Pattern Recognition and Machine Learning

4. EMアルゴリズムとは？

<u>Expectation-Maximization algorithm</u>

観測不可能な潜在変数に確率モデルが依存する場合にパラメータを推定する手法.

基本的に，最尤法に基づく．期待値最大化法とも.

EステップとMステップを交互に繰り返す.

<u>Eステップ（期待値ステップ）</u>
現在推定されている潜在変数の分布に基づいて，モデル尤度の期待値を計算する.

<u>Mステップ（最大化ステップ）</u>
Eステップで得られた尤度の期待値を最大化するようなパラメータを求める．これにより次のEステップで用いられる潜在変数分布を決めることができる.

→よくわからない…まずは潜在変数を理解する.

5. 混合ガウスの潜在変数

K次元の2値確率変数zを導入する.

これは1-of-K，つまりK通りの内どれか一つだけ1,他は0.

Zk=1なる事象の事前確率

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

※πの定義より,$0 \le \pi_k \le 1, \sum_{k=1}^{K} \pi_k = 1$であり，確率の値の定義を満たす.

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

$$p(\mathbf{x}) = \sum_{z} p(\mathbf{z}) p(\mathbf{x} \mid \mathbf{z}) = \sum_{k=1}^{K} \pi_k N(\mathbf{x} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

$$p(z_k = 1) = \pi_k \qquad p(\mathbf{x} | z_k = 1) = N(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

$$p(\mathbf{x}) = \sum_{z} p(\mathbf{z}) p(\mathbf{x} | \mathbf{z}) = \sum_{k=1}^{K} \pi_k N(\mathbf{x} | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

※$p(\mathbf{x}) p(\mathbf{z} | \mathbf{x}) = p(\mathbf{z}) p(\mathbf{x} | \mathbf{z})$

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

Xを観測したときの対応する事後確率

＝混合要素kがxの観測を説明する度合を表す「負担率」と解釈できる．

5. 混合ガウスの潜在変数

以上，混合ガウス分布での示唆より，
EMアルゴリズムは，
「潜在変数の事後分布」に着目してつくられた簡易計算法なのでは？
と考えられる．

混合ガウス分布から拡張し，EMアルゴリズムの一般的な利用を見てみる．

4. 一般のEMアルゴリズム

観測変数Xと潜在変数Zの同時分布p(X,Z|θ)が与えられていて，パラメータθで支配されているとする．アルゴリズムの目的は尤度関数p(X|θ)をθに関して最大化することである．

1: パラメタの初期値θ(old)を選ぶ．

2(Eステップ): 事後確率p(Z|X,θ(old))を計算する．

3(Mステップ): 上で求めた事後確率を所与として尤度関数を最大化するθ(new)を求める．

4: 尤度関数またはパラメタ値が収束条件を満たしていない場合θ(old)にθ(new)を代入し2へ！

3(Mステップ): 上で求めた事後確率を所与として尤度関数を最大化するθ(new)を求める．

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

$$Q(\theta, \theta^{old}) = \underbrace{\sum_{z} p(\mathbf{Z} \mid \mathbf{X}, \theta^{old}) \ln p(\mathbf{X}, \mathbf{Z} \mid \theta)}_{\text{完全データ対数尤度の期待値}}$$

⇒なぜ期待値を最大化してるのか？

4. 一般のEMアルゴリズム

$$p(\mathbf{X}\mid\theta)=\sum_{Z} p(\mathbf{X},\mathbf{Z}\mid\theta)$$

これが最適化困難であるときどうするか．

確率の連鎖律より
$$\ln p(\mathbf{X},\mathbf{Z}\mid\theta)=\ln p(\mathbf{Z}\mid\mathbf{X},\theta)+\ln p(\mathbf{X}\mid\theta)$$

代入

$$L(q,\theta)\equiv\sum_{Z} q(\mathbf{Z})\ln\left\{\frac{p(\mathbf{X},\mathbf{Z}\mid\theta)}{q(\mathbf{Z})}\right\}$$
と定義してこれを計算

$$=\sum_{Z} q(\mathbf{Z})\{\ln p(\mathbf{Z}\mid\mathbf{X},\theta)+\ln p(\mathbf{X}\mid\theta)-q(\mathbf{Z})\}$$

$$=\sum_{Z} q(\mathbf{Z})\ln\left\{\frac{\ln p(\mathbf{Z}\mid\mathbf{X},\theta)}{q(\mathbf{Z})}\right\}+\ln p(\mathbf{X}\mid\theta)\underbrace{\sum_{Z} q(\mathbf{Z})}_{=1}$$

$$=-KL(q\,\|\,p)+\ln p(\mathbf{X}\mid\theta)$$

$$KL(q \| p) = -\sum_{Z} q(\mathbf{Z}) \ln\left\{\frac{\ln p(\mathbf{Z} \mid \mathbf{X}, \theta)}{q(\mathbf{Z})}\right\}$$ に関して

これはq(Z)と事後分布p(Z|X,θ)の

カルバック-ライブラーダイバージェンス（KLダイバージェンス），

つまり相対エントロピーである．

∵関数-lnxが凸関数

$$一般型KL(q \| p) = -\int p(\mathbf{x}) \ln\left\{\frac{q(\mathbf{x})}{p(\mathbf{x})}\right\} d\mathbf{x} \geq -\ln \int q(\mathbf{x}) d\mathbf{x} = 0$$

$$※\int q(\mathbf{x}) d\mathbf{x} = 1(規格化条件)$$

Eステップ解説

以下を，q(Z)について最大化

$$L(q,\theta^{old}) \equiv -\underbrace{KL(q\|p)}_{\geq 0} + \ln \underbrace{p(\mathbf{X}|\theta^{old})}_{\text{q(Z)には依存しない}}$$

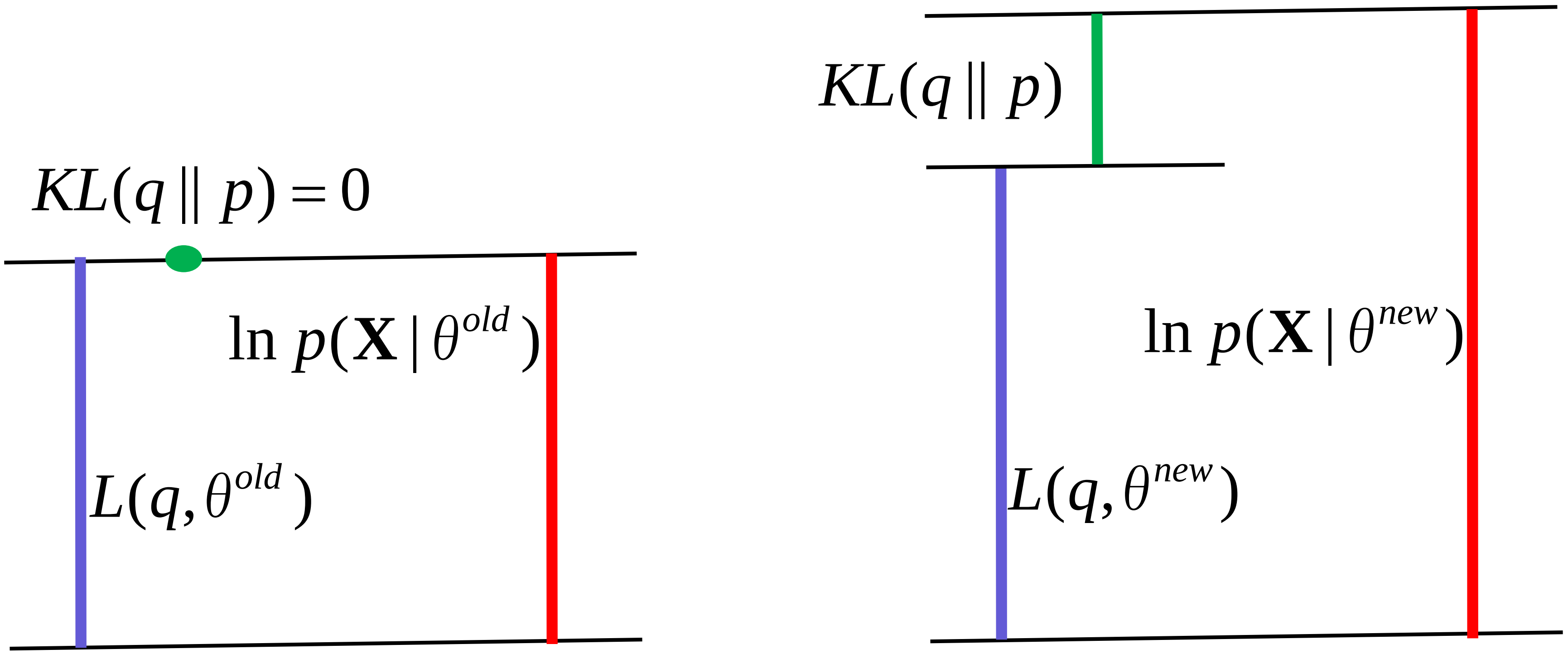
$$\longrightarrow \quad KL(q\|p)=0 \Rightarrow \underline{q(\mathbf{Z})=p(\mathbf{Z}|\mathbf{X},\theta^{old})}$$

のとき最大である！

Mステップ解説

$$L(q,\theta) \equiv -KL(q \| p) + \ln p(\mathbf{X}|\theta)$$

q(Z)を固定し，θについて下界を最大化



⇒lnp(X|θ)の増加量はL(q,θ)の増加量よりも大きい

Mステップ解説

$$L(q,\theta) \equiv \sum_{Z} q(\mathbf{Z}) \ln \left\{ \frac{p(\mathbf{X},\mathbf{Z} \mid \theta)}{q(\mathbf{Z})} \right\} \qquad q(\mathbf{Z}) = p(\mathbf{Z} \mid \mathbf{X}, \theta^{old})$$

$$= \sum_{Z} p(\mathbf{Z} \mid \mathbf{X}, \theta^{old}) \ln \left\{ \frac{p(\mathbf{X},\mathbf{Z} \mid \theta)}{p(\mathbf{Z} \mid \mathbf{X}, \theta^{old})} \right\}$$

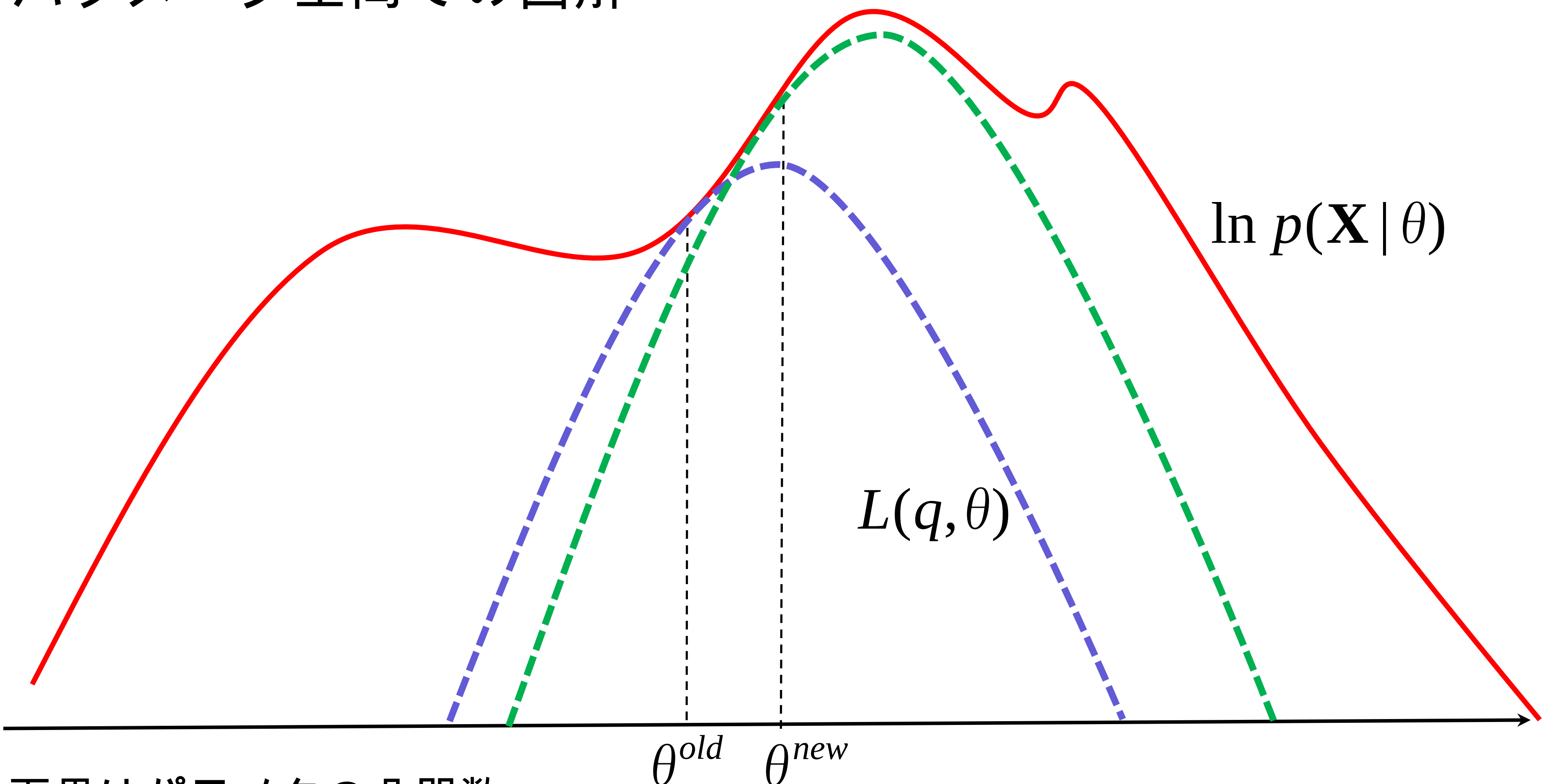
$$= \sum_{Z} p(\mathbf{Z} \mid \mathbf{X}, \theta^{old}) \ln p(\mathbf{X},\mathbf{Z} \mid \theta) - \sum_{Z} p(\mathbf{Z} \mid \mathbf{X}, \theta^{old}) \ln p(\mathbf{Z} \mid \mathbf{X}, \theta^{old})$$

$$= Q(\theta, \theta^{old}) + const.$$

L(q,θ)を最大化するとき，実際は「完全データ対数尤度の期待値」を最大化することになる．

変数θは対数の中のみに表れている→最適化容易

パラメータ空間での図解



※下界はパラメタの凸関数

めも

まとめると・・・

EMアルゴリズムはKmeansと比べるとほぼ終息するまでに必要な繰り返し数と繰り返し一回の計算量が多い．

⇒一般的に，混合ガウスモデルの適切な初期値を見つけるためにKmeansが使われたりする．（そのあとにEMを使う）

Q,どうKmeansとEMを対応させる？

A,共分散行列の初期値→クラスタのサンプル分散

混合係数の初期値→各クラスタに属するデータ点の割合

※対数尤度関数が多峰であることが多く，その場合に唯一解を与えるとは限らない．

観測データ集合Xから計算される尤度/対数尤度関数

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

完全データ集合（⇒各データ点がどの要素から生成したかを示すラベルを含むデータ集合）に関する

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

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