東京大学工学部 4年生 夏学期 [03-501130]

応用音響学
第4回 (5/ 1)

亀岡弘和

東京大学大学院情報理工学系研究科
システム情報学専攻
kameoka@hil.t.u-tokyo.ac.jp

講義スケジュール

前半(亀岡担当)

- 4/08: 第1回
- 4/15: 第2回
- 4/22: 第3回
- 4/29: 休日
- 5/01(木): 第4回
- 5/06: 休日
- 5/13: 第5回
- 5/20: 第6回
- 5/27: 第7回

後半(牧野担当)

- 6/03: 第8回
- 6/10: 第9回
- 6/17: 第10回
- 6/24: 第11回
- 7/01: 第12回
- 7/08: 第13回
- 7/15: 第14回

- 7/22: 学期末試験

講義資料と成績評価

- 講義資料
  - http://hil.t.u-tokyo.ac.jp/~kameoka/aa/
- 成績評価
  - 出席点
  - 学期末試験

本日の話題

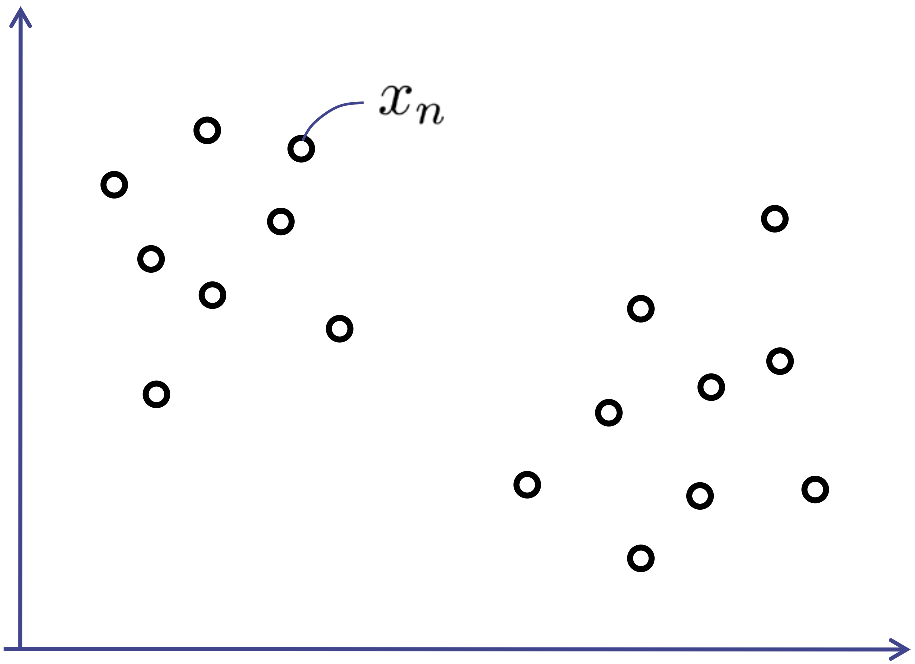
- 確率モデル（生成モデル）による
  パターン認識（クラスタリング）問題の定式化
  - k-meansクラスタリング
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係

本日の話題

- 確率モデル(生成モデル)による
  パターン認識(クラスタリング)問題の定式化
  - k-meansクラスタリングのおさらい
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係

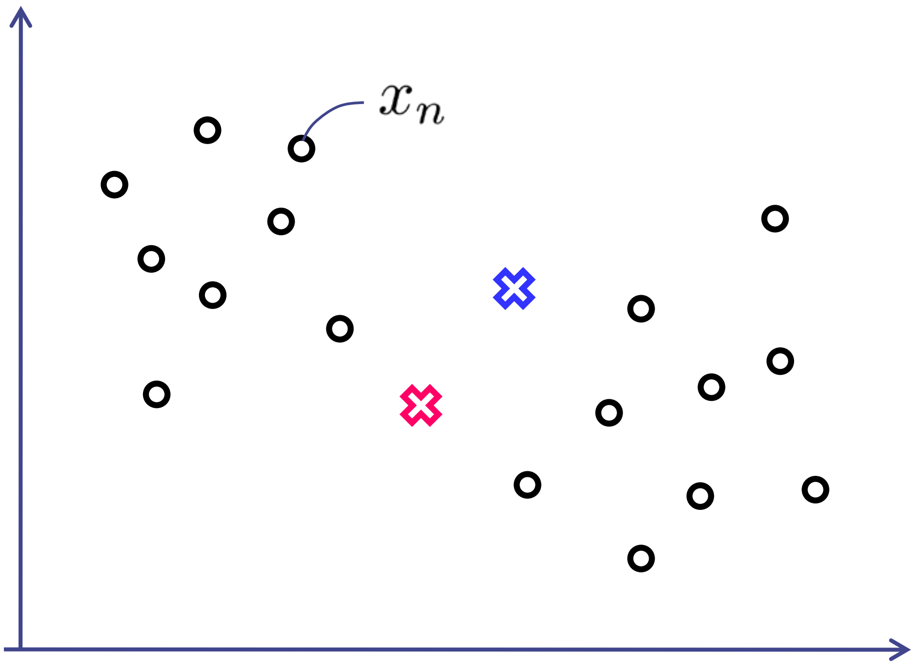
k-meansクラスタリング

- データ $X = \{x_1, \ldots, x_N\}$ の教師なしクラスタリングの一手法



k-meansクラスタリング

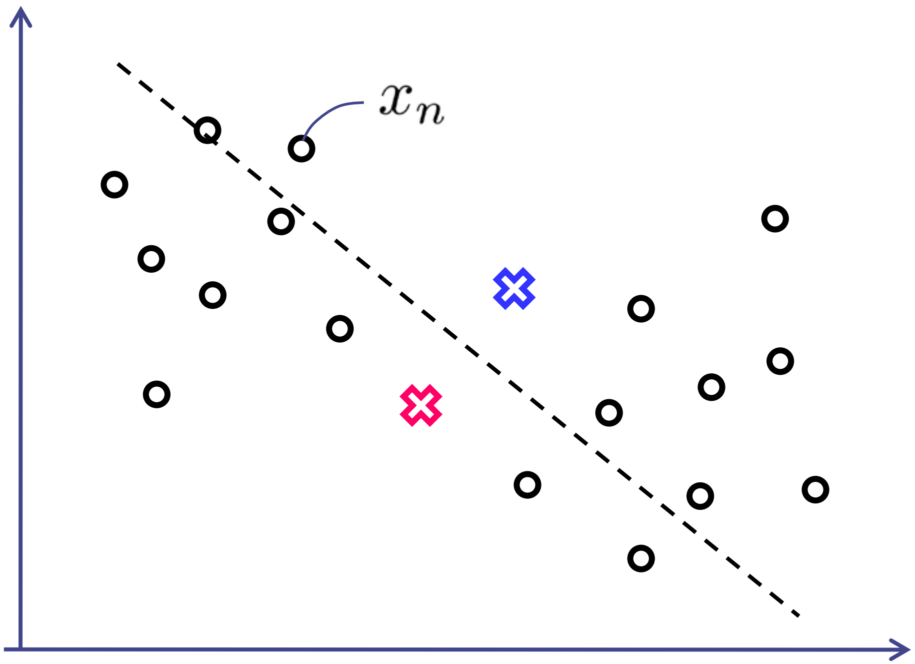
- データ $X = \{x_1, \ldots, x_N\}$ の教師なしクラスタリングの一手法



(Step 0)
クラスタ代表値を初期設定

k-meansクラスタリング

- データ $X = \{x_1, \ldots, x_N\}$ の教師なしクラスタリングの一手法

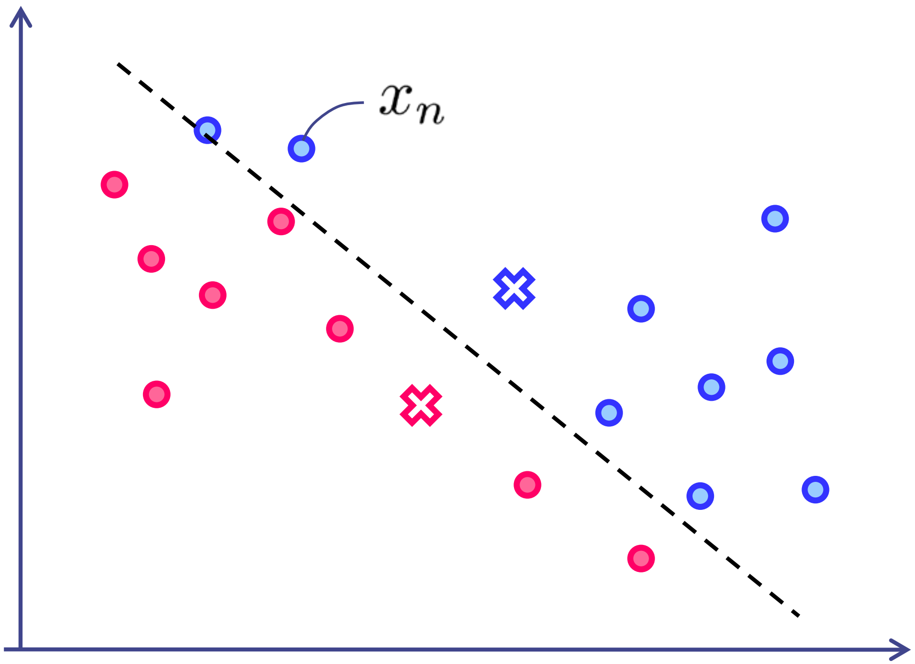


(Step 0)
クラスタ代表値を初期設定

(Step 1)
クラスタ境界を決定

k-meansクラスタリング

- データ $X = \{x_1, \ldots, x_N\}$ の教師なしクラスタリングの一手法

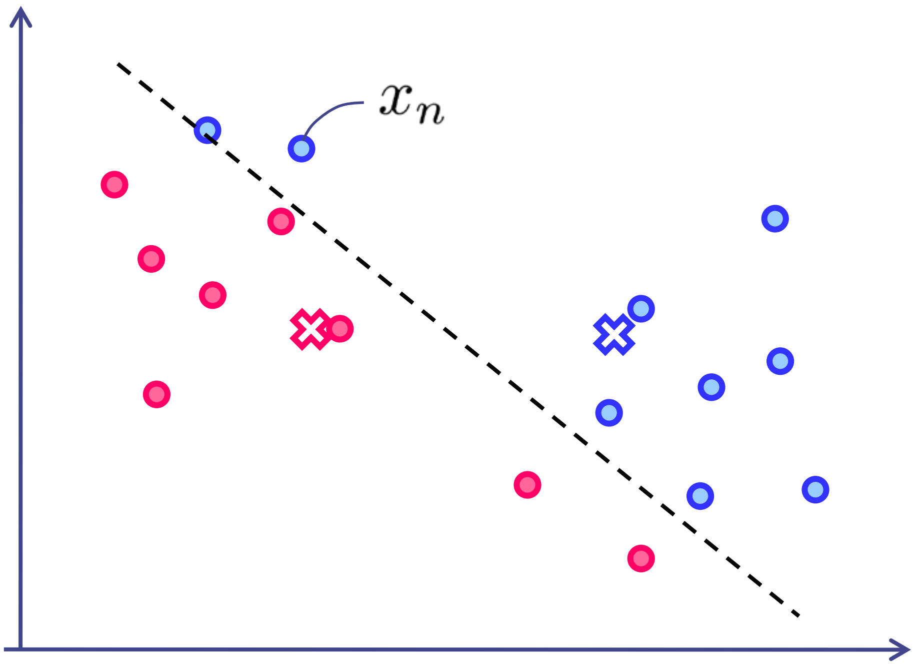


(Step 0)
クラスタ代表値を初期設定

(Step 1)
クラスタ境界を決定

k-meansクラスタリング

- データ $X = \{x_1, \ldots, x_N\}$ の教師なしクラスタリングの一手法



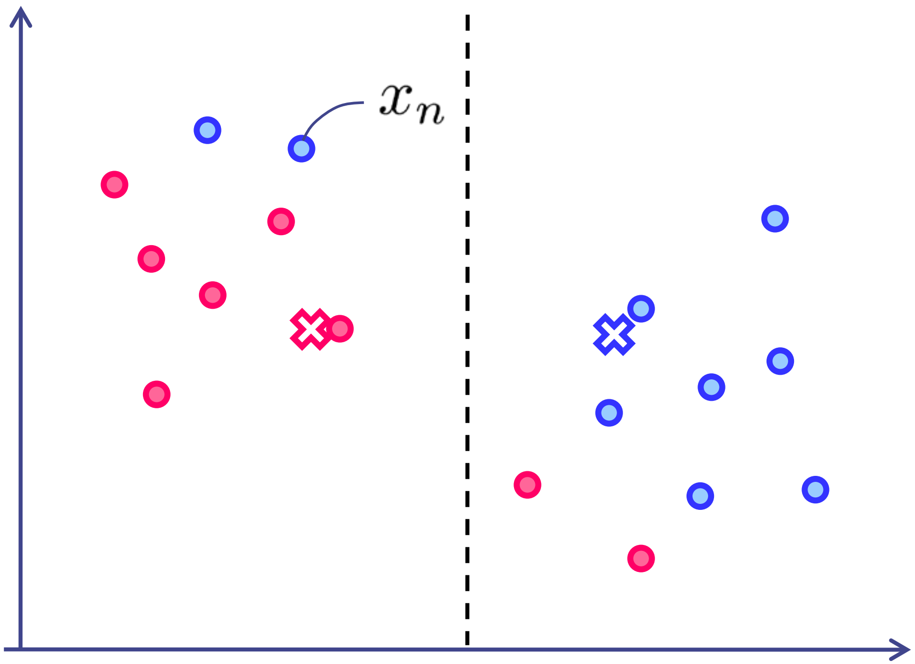
(Step 0)
クラスタ代表値を初期設定

(Step 1)
クラスタ境界を決定

(Step 2)
クラスタ代表値を決定

k-meansクラスタリング

- データ $X = \{x_1, \ldots, x_N\}$ の教師なしクラスタリングの一手法



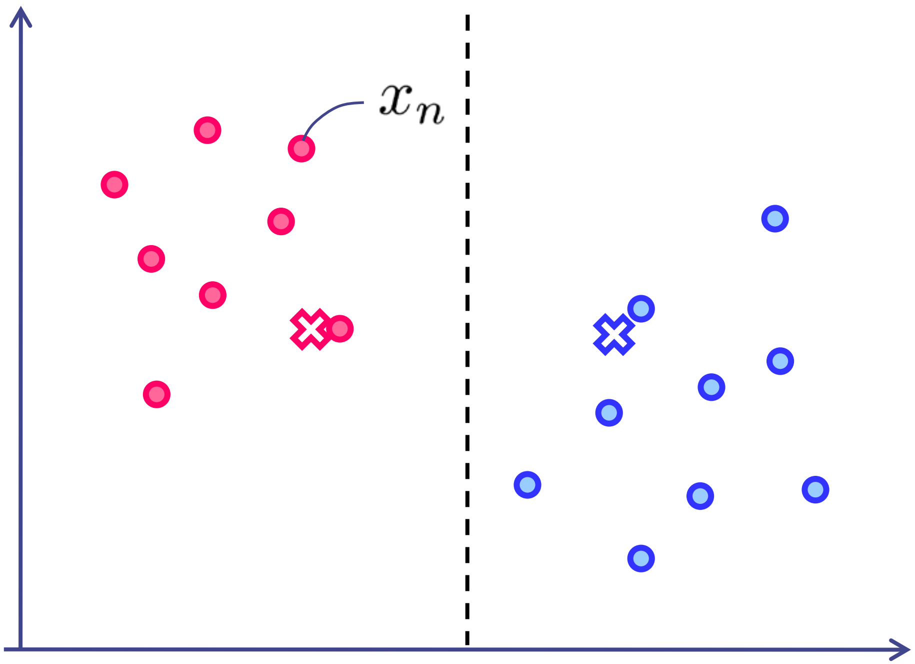
(Step 0)
クラスタ代表値を初期設定

(Step 1)
クラスタ境界を決定

(Step 2)
クラスタ代表値を決定

k-meansクラスタリング

- データ $X = \{x_1, \ldots, x_N\}$ の教師なしクラスタリングの一手法



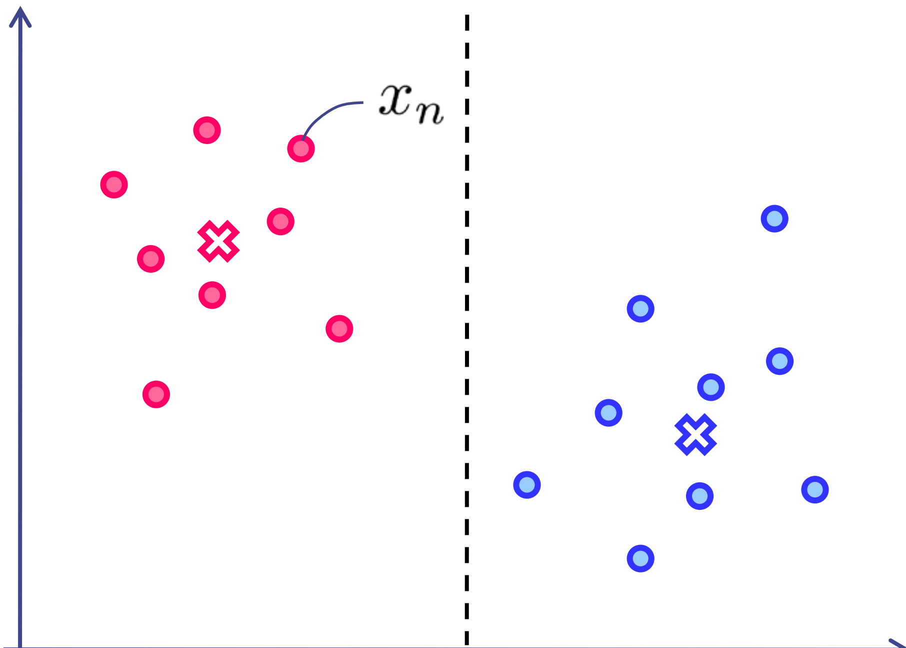
(Step 0)
クラスタ代表値を初期設定

(Step 1)
クラスタ境界を決定

(Step 2)
クラスタ代表値を決定

k-meansクラスタリング

- データ $X = \{x_1, \ldots, x_N\}$ の教師なしクラスタリングの一手法



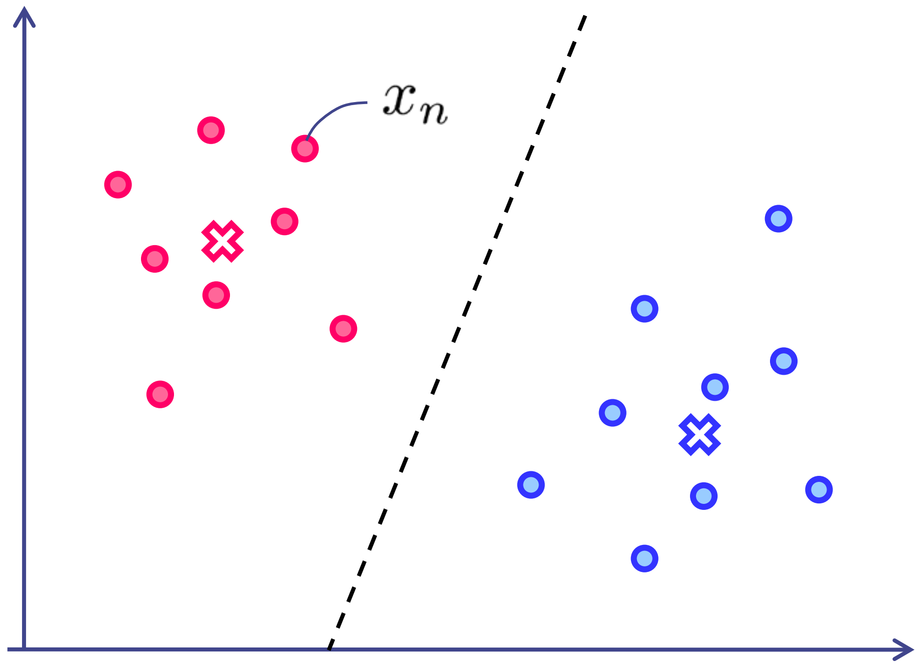
(Step 0)
クラスタ代表値を初期設定

(Step 1)
クラスタ境界を決定

(Step 2)
クラスタ代表値を決定

k-meansクラスタリング

- データ $X = \{x_1, \ldots, x_N\}$ の教師なしクラスタリングの一手法



(Step 0)
クラスタ代表値を初期設定

(Step 1)
クラスタ境界を決定

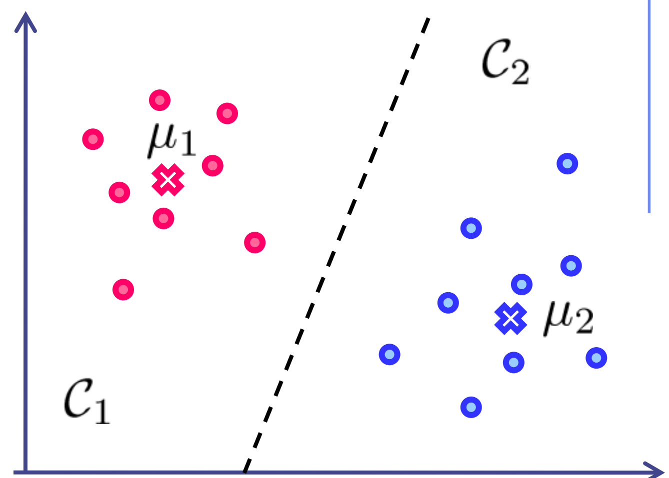
(Step 2)
クラスタ代表値を決定

k-meansクラスタリング

- アルゴリズムの収束性
  - クラスタリングの目的関数

$$J(Z,\mu) \equiv \sum_k \sum_n \mathbf{1}[z_n = k] \underbrace{\|x_n - \mu_k\|_2^2}_{\text{散らばり}}$$

$$\mathbf{1}[z_n = k] = \begin{cases} 1 & (z_n = k) \\ 0 & (z_n \neq k) \end{cases}$$



- $\{\mu_1, \ldots, \mu_K\}$: クラスタ代表値
- $\{z_1, \ldots, z_N\}$: 各データの帰属クラスタ

  - 反復アルゴリズム
    - **(Step 1)** $\hat{Z} \leftarrow \operatorname*{argmin}_Z J(Z,\mu)$ ⇒明らかに $\hat{z}_n = \operatorname*{argmin}_j \|x_n - \mu_j\|_2^2$
    - **(Step 2)** $\hat{\mu} \leftarrow \operatorname*{argmin}_\mu J(Z,\mu)$ ⇒ $\partial J(Z,\mu)/\partial \mu_j = 0$ を解くと

$$\hat{\mu}_j = \underbrace{\frac{\sum_n \mathbf{1}[z_n = j] x_n}{\sum_n \mathbf{1}[z_n = j]}}_{}$$ クラスタ重心

目的関数は当然単調減少する
→収束性が保証される

本日の話題

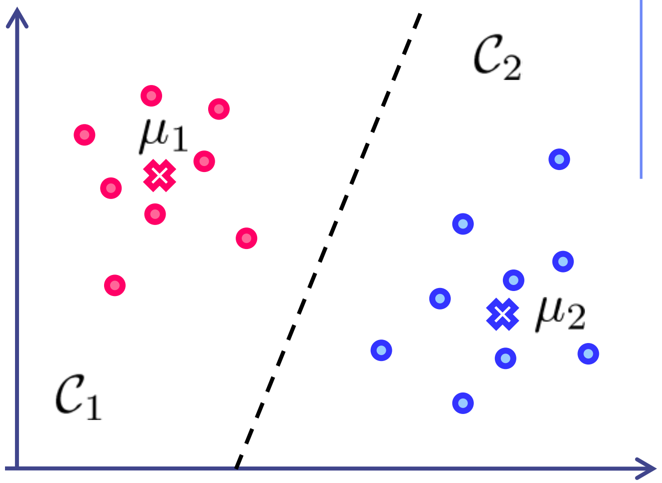
- 確率モデル（生成モデル）による
  パターン認識（クラスタリング）問題の定式化
  - k-meansクラスタリング
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係

本日の話題

- 確率モデル（生成モデル）による
  パターン認識（クラスタリング）問題の定式化
  - k-meansクラスタリング
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係

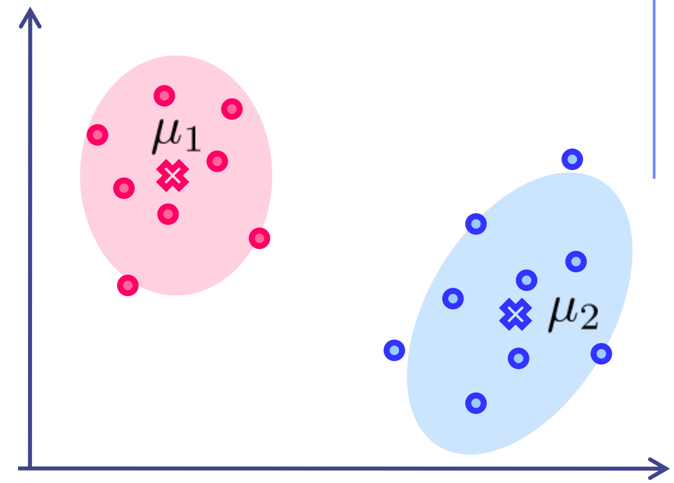
k-meansクラスタリングの生成モデルとしての解釈

- K個のクラスタ（塊）からなるデータの生成プロセスを確率モデルで表現できないか？
  - クラスタリング問題はその逆プロセス



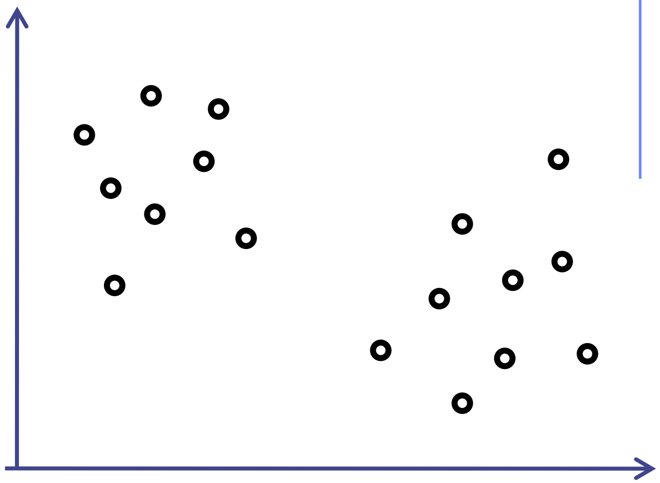
k-meansクラスタリングの生成モデルとしての解釈

- K個のクラスタ(塊)からなるデータの生成プロセスを確率モデルで表現できないか？
  - クラスタリング問題はその逆プロセス
- 生成プロセスの例
  - K個の正規分布をランダムに生成
  - サンプルごとに
    - 正規分布のクラス番号をランダムに選択
    - 選択されたクラス番号の正規分布に従ってサンプル値を生成
  - 生成された全サンプルが観測データ！



k-meansクラスタリングの生成モデルとしての解釈

- K個のクラスタ（塊）からなるデータの生成プロセスを確率モデルで表現できないか？
  - クラスタリング問題はその逆プロセス
- 生成プロセスの例
  - K個の正規分布をランダムに生成
  - サンプルごとに
    - 正規分布のクラス番号をランダムに選択
    - 選択されたクラス番号の正規分布に従ってサンプル値を生成
  - 生成された全サンプルが観測データ！



- データの確率的な生成プロセスの仮定

⇔データの確率モデル化

本日の話題

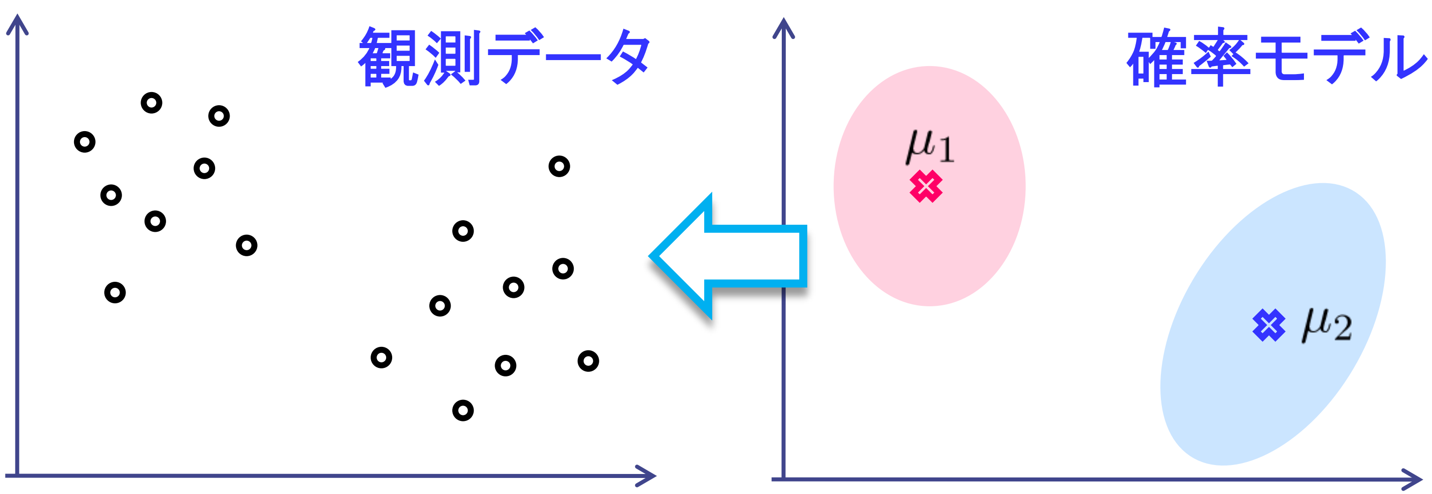
- 確率モデル(生成モデル)による
  パターン認識(クラスタリング)問題の定式化
  - k-meansクラスタリング
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係

本日の話題

- 確率モデル（生成モデル）による
  パターン認識（クラスタリング）問題の定式化
  - k-meansクラスタリング
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係

最尤推定

- データから、尤度関数が最大となる確率モデルのパラメータを推定するための一般的な枠組
  - 尤度関数とは？
    - データを X，パラメータを Θ とすると，尤度関数は $p(X|\Theta)$ のこと
    - 「データ X の生成源として，パラメータ Θ の確率モデルがどの程度尤もらしいか」を意味した規準
  - 最尤推定
    - 最も尤もらしい、データの生成源を推定すること
  - クラスタリング問題の例では･･･



本日の話題

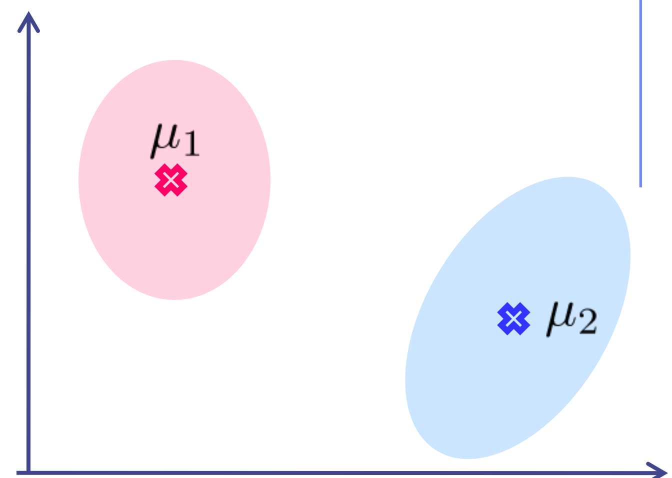
- 確率モデル(生成モデル)による
  パターン認識(クラスタリング)問題の定式化
  - k-meansクラスタリング
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係

本日の話題

- 確率モデル（生成モデル）による
  パターン認識（クラスタリング）問題の定式化
  - k-meansクラスタリング
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係

混合正規分布モデル (Gaussian Mixture Model; GMM)

- 混合正規分布モデルによるデータの生成プロセス
  - K個の正規分布をランダムに生成
  - サンプルごとに
    - 正規分布のクラス番号をランダムに選択
    - 選択されたクラス番号の正規分布に従ってサンプル値を生成
  - 生成された全サンプルが観測データ



- 生成モデル

for $k = 1, \ldots, K$

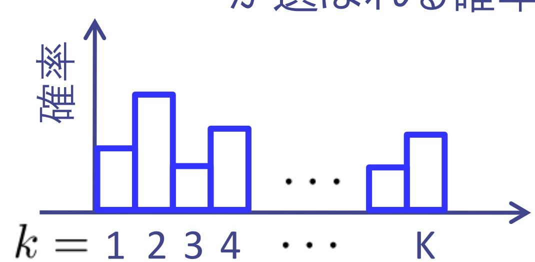
$$\{\mu_k, \sigma_k, \pi_k\} \sim H$$

for $n = 1, \ldots, N$

$$z_n \sim \mathrm{Categorical}(\pi_1, \ldots, \pi_K)$$

$$x_n | z_n \sim \mathcal{N}(\mu_{z_n}, \Sigma_{z_n})$$

π_k：k番目の正規分布が選ばれる確率



本日の話題

- 確率モデル（生成モデル）による
  パターン認識（クラスタリング）問題の定式化
  - k-meansクラスタリング
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係

本日の話題

- 確率モデル（生成モデル）による
  パターン認識（クラスタリング）問題の定式化
  - k-meansクラスタリング
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係

EMアルゴリズム

- GMMパラメータの最尤推定
  - GMMのパラメータ：$\Theta = \{\mu_k, \sigma_k, \pi_k\}_{k=1,\dots,K}$
  - 求めたいのは $\hat{\Theta}_{\mathrm{ML}} = \underset{\Theta}{\operatorname{argmax}}\, p(X|\Theta) = \underset{\Theta}{\operatorname{argmax}} \log p(X|\Theta)$

対数をとっても最大化の目標は変わらない

- 尤度関数の導出

生成モデル

$$
\begin{cases}
p(Z|\Theta) = \prod_n p(z_n|\Theta) = \prod_n \pi_{z_n} \\
P(X|Z,\Theta) = \prod_n p(x_n|z_n,\Theta) = \prod_n \mathcal{N}(x_n; \mu_{z_n}, \Sigma_{z_n})
\end{cases}
$$

$$
\mathcal{N}(x_n; \mu_k, \Sigma_k) \equiv \frac{1}{\sqrt{(2\pi)^J |\Sigma_k|}} \exp\left\{ -\frac{1}{2}(x_n - \mu_k)^{\mathsf{T}} \Sigma_k^{-1} (x_n - \mu_k) \right\}
$$

$$
\boxed{
\begin{aligned}
p(X|\Theta) &= \sum_Z p(X|Z,\Theta) p(Z|\Theta) = \prod_n \sum_{z_n=1}^{K} p(x_n|z_n,\Theta) p(z_n|\Theta) \\
&= \prod_n \sum_{z_n=1}^{K} \pi_{z_n} \mathcal{N}(x_n; \mu_{z_n}, \Sigma_{z_n})
\end{aligned}
}
$$

EMアルゴリズム

- $p(X|\Theta) = \prod_n \sum_{z_n=1}^{K} \pi_{z_n} \mathcal{N}(x_n; \mu_{z_n}, \Sigma_{z_n})$ を最大化する Θ を求めたい
  - しかし解析的には求められない！
  - ちなみに，Z が既知であれば(各データにクラスラベルが付与されていれば)，各正規分布パラメータの最尤推定は解析的に求められる
    ⇒平均，分散，データの各クラスへの帰属率を求めるだけ
  - 解析的な求解を困難にしているのは

$$\log p(X|\Theta) = \sum_n \log \sum_{z_n=1}^{K} \pi_{z_n} \mathcal{N}(x_n; \mu_{z_n}, \Sigma_{z_n})$$

におけるlogの中のsummation

Z のように
尤度関数に姿を現さない
変数のことを「潜在変数」
または「隠れ変数」という

補助関数法

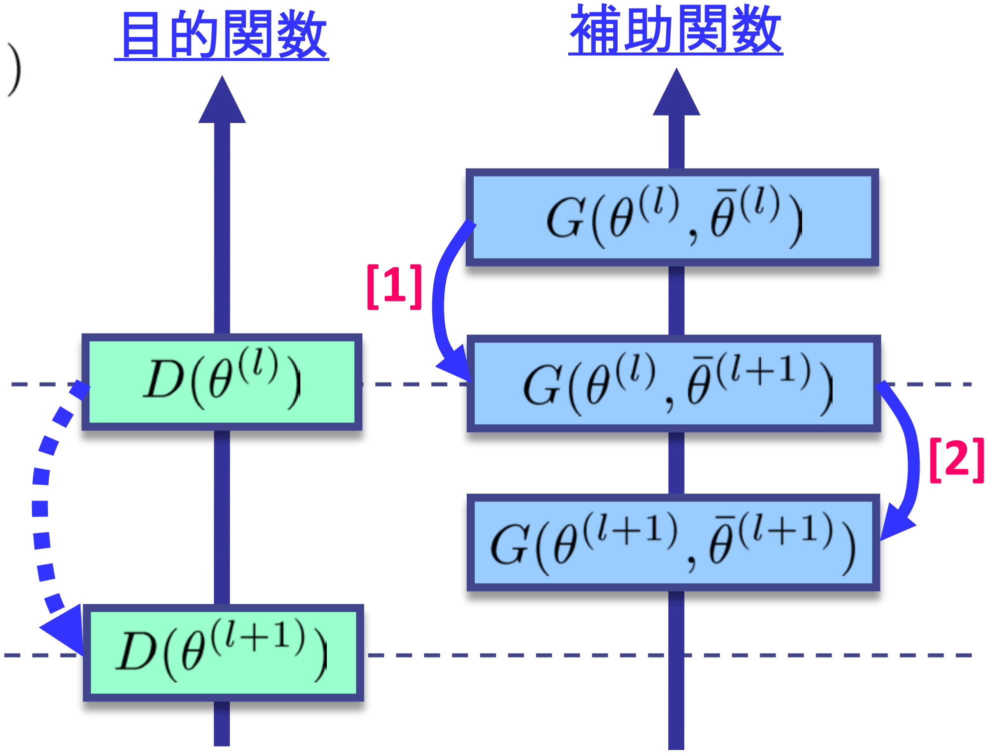
- 補助関数法
  - 目的関数 $D(\theta)$ を局所最小化するためのテクニック
  - $D(\theta) = \min_{\bar{\theta}} G(\theta, \bar{\theta})$ を満たす $G(\theta, \bar{\theta})$ を補助関数と定義
  - 反復アルゴリズム

[1] $\bar{\theta}^{(l+1)} = \underset{\bar{\theta}}{\operatorname{argmin}}\, G(\theta^{(l)}, \bar{\theta})$

[2] $\theta^{(l+1)} = \underset{\theta}{\operatorname{argmin}}\, G(\theta, \bar{\theta}^{(l+1)})$

  - 収束性

$$
\begin{aligned}
D(\theta^{(l)}) &= G(\theta^{(l)}, \bar{\theta}^{(l+1)}) \\
&\geq G(\theta^{(l+1)}, \bar{\theta}^{(l+1)}) \\
&\geq D(\theta^{(l+1)})
\end{aligned}
$$



Jensenの不等式

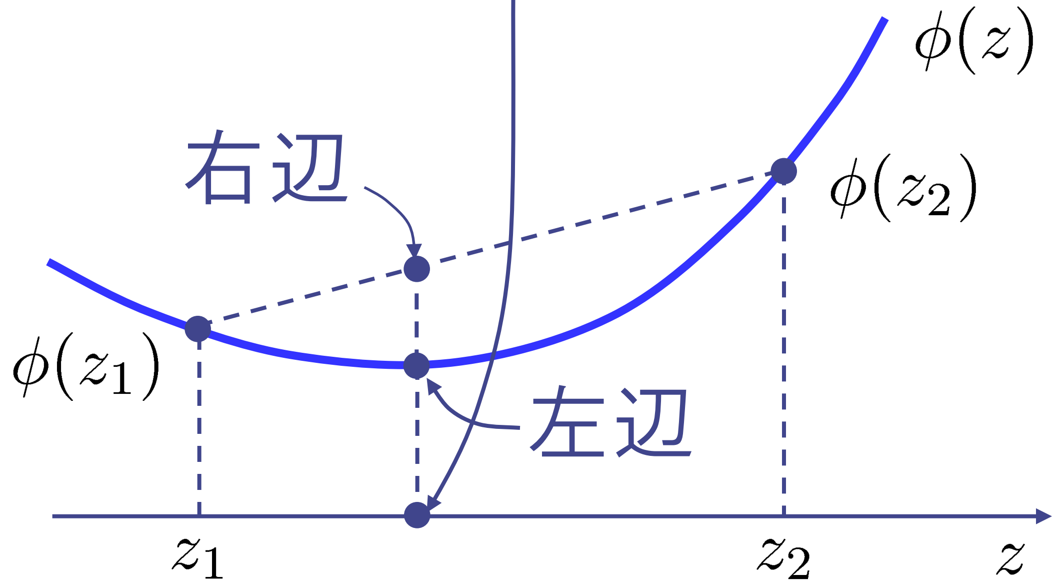
- Jensenの不等式
  - $\phi(\cdot)$：凸関数
  - $\lambda_i \geq 0, \ \sum_i \lambda_i = 1$

$$\Rightarrow \phi\left(\underline{\sum_i \lambda_i z_i}\right) \leq \sum_i \lambda_i \phi(z_i)$$

例えば,

$\phi(z) = -\log z$ の場合：

$$-\log\left(\sum_i \lambda_i z_i\right) \leq -\sum_i \lambda_i \log z_i$$



EMアルゴリズム

- 対数尤度 $\log p(X|\Theta) = \sum_n \log \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n; \mu_k, \Sigma_k)$ の下限関数

$$
\begin{aligned}
&\log \sum_{k=1}^{K} \pi_k \mathcal{N}(x_n; \mu_k, \Sigma_k) \\
&= \log \sum_{k=1}^{K} \lambda_{k,n} \frac{\pi_k \mathcal{N}(x_n; \mu_k, \Sigma_k)}{\lambda_{k,n}} \\
&\geq \sum_{k=1}^{K} \lambda_{k,n} \log \frac{\pi_k \mathcal{N}(x_n; \mu_k, \Sigma_k)}{\lambda_{k,n}}
\end{aligned}
$$

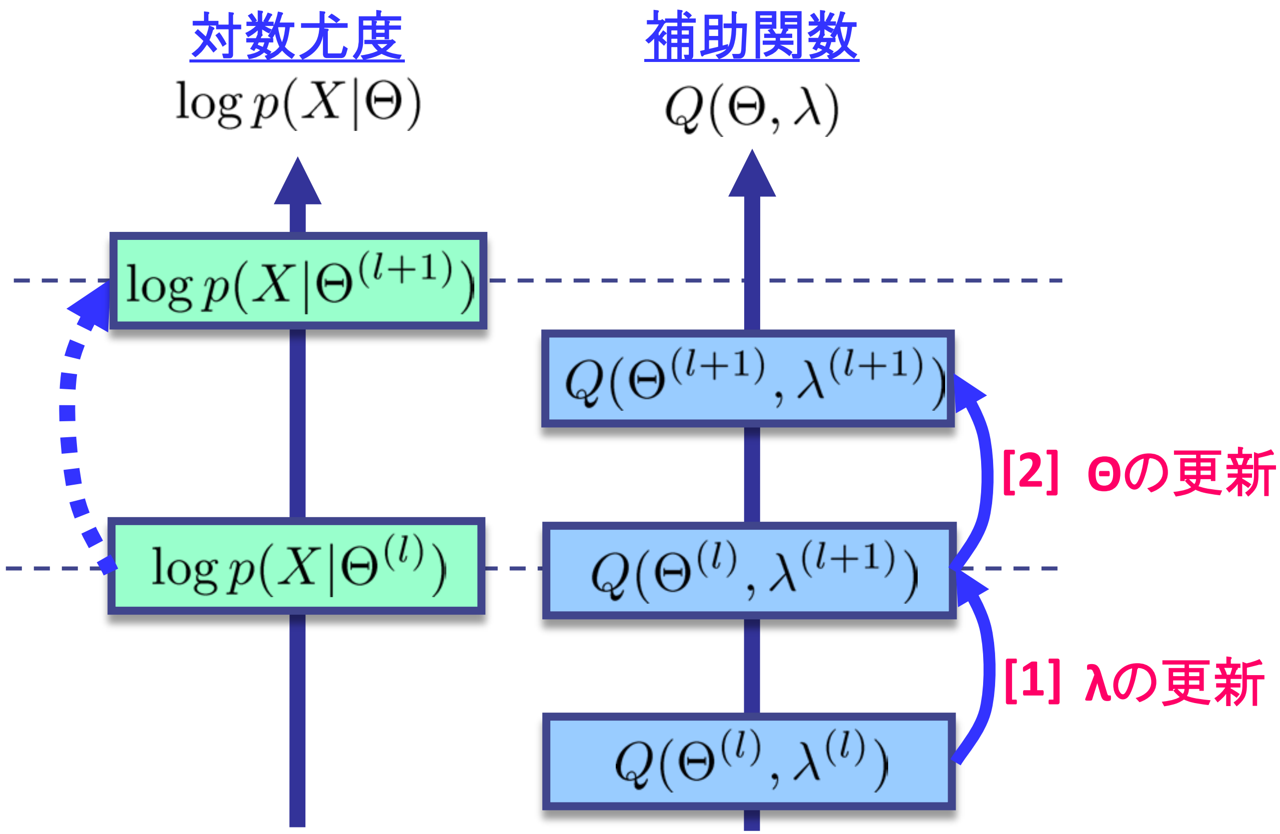
← 対数関数が凹関数であることに注意してJensenの不等式を適用

$Q(\Theta, \lambda) \equiv \sum_n \sum_{k=1}^{K} \lambda_{k,n} \log \frac{\pi_k \mathcal{N}(x_n; \mu_k, \Sigma_k)}{\lambda_{k,n}}$ は $\log p(X|\Theta)$ の

補助関数！

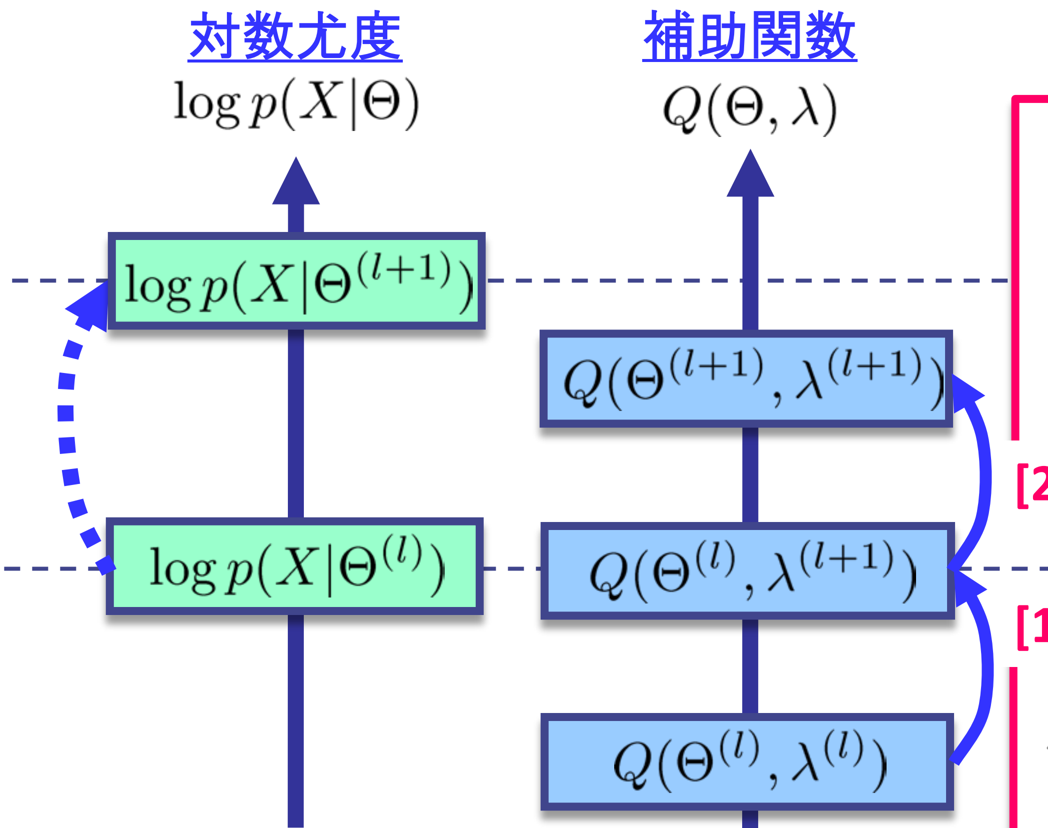
EMアルゴリズム

- 対数尤度と補助関数の関係：$\min_{\lambda} Q(\Theta, \lambda) = \log p(X|\Theta)$
- 反復アルゴリズムによる局所最大化



EMアルゴリズム

- 対数尤度と補助関数の関係： $\min_{\lambda} Q(\Theta, \lambda) = \log p(X|\Theta)$
- 反復アルゴリズムによる局所最大化



$$\Theta^{(l+1)} = \operatorname*{argmax}_{\Theta} Q(\Theta, \lambda^{(l+1)})$$

$$\mu_k^{(l+1)} = \frac{\sum_n \lambda_{k,n}^{(l+1)} x_{k,n}}{\sum_n \lambda_{k,n}^{(l+1)}} \quad \cdots$$

[2] Θの更新

[1] λの更新

$$\lambda^{(l+1)} = \operatorname*{argmax}_{\lambda} Q(\Theta^{(l)}, \lambda)$$

$$\lambda_{k,n}^{(l+1)} = \frac{\pi_k \mathcal{N}(x_n; \mu_k^{(l)}, \Sigma_k^{(l)})}{\sum_m \pi_m \mathcal{N}(x_n; \mu_m^{(l)}, \Sigma_m^{(l)})}$$

なぜEMアルゴリズムと呼ばれるか

■λの更新式

$$
\begin{aligned}
\lambda_{k,n}^{(l+1)} &= \frac{\pi_k \mathcal{N}(x_n; \mu_k^{(l)}, \Sigma_k^{(l)})}{\sum_m \pi_m \mathcal{N}(x_n; \mu_m^{(l)}, \Sigma_m^{(l)})} \\
&= \frac{p(x_n | z_n = k, \Theta) p(z_n = k | \Theta)}{\sum_m p(x_n | z_n = m, \Theta) p(z_n = m | \Theta)} \\
&= \frac{p(x_n | z_n = k, \Theta) p(z_n = k | \Theta)}{p(x_n | \Theta)} \\
&= p(z_n = m | x_n, \Theta)
\end{aligned}
$$

データ x_n がクラス m の正規分布から
生成された(事後)確率

■補助関数

$p(x_n, z_n = k|\Theta)$

観測データと潜在変数のペア
⇒「完全データ」と呼ぶ

$$
Q(\Theta, \lambda) \equiv \sum_n \sum_{k=1}^{K} \lambda_{k,n} \log \frac{\pi_k \mathcal{N}(x_n; \mu_k, \Sigma_k)}{\lambda_{k,n}}
$$

よって、$Q(\Theta, \lambda)$ は「完全データに対する対数尤度」の期待値に相当

本日の話題

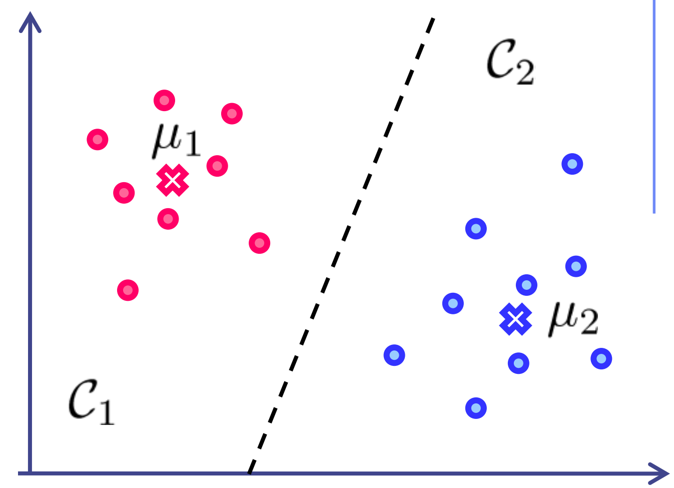
- 確率モデル(生成モデル)による
  パターン認識(クラスタリング)問題の定式化
  - k-meansクラスタリング
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係

本日の話題

- 確率モデル（生成モデル）による
  パターン認識（クラスタリング）問題の定式化
  - k-meansクラスタリング
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係

k-meansクラスタリングとの関係

- k-meansアルゴリズム
  - (Step 1) $z_n \leftarrow \underset{j}{\operatorname{argmin}} \|x_n - \mu_j\|_2^2$
  - (Step 2) $\mu_j \leftarrow \dfrac{\sum_n \mathbf{1}[z_n = j]x_n}{\sum_n \mathbf{1}[z_n = j]}$



- EMアルゴリズムによるGMMパラメータの最尤推定
  - (E-Step) $\lambda_{k,n} \leftarrow \dfrac{\pi_k \mathcal{N}(x_n; \mu_k, \Sigma_k)}{\sum_m \pi_m \mathcal{N}(x_n; \mu_m, \Sigma_m)}$
  - (M-step) $\mu_k \leftarrow \dfrac{\sum_n \lambda_{k,n} x_{k,n}}{\sum_n \lambda_{k,n}} \quad \cdots$

- 後者が前者と等価となる条件：$\Sigma_k \to 0$
  - Eステップにおけるデータのクラス分けが排他的になる

本日の講義のまとめ

- 確率モデル（生成モデル）による
  パターン認識（クラスタリング）問題の定式化
  - k-meansクラスタリング
  - k-meansクラスタリングの生成モデルとしての解釈
  - 最尤推定
  - 混合正規分布モデル (Gaussian Mixture Model; GMM)
  - EMアルゴリズム
    - 補助関数とJensen不等式
  - EMアルゴリズムによるGMMパラメータの最尤推定とk-meansアルゴリズムの関係