目次

混合ガウス分布

- 混合ガウス分布を導入する理由
  - 曖昧さを含んだクラスタリング (ソフト割り当て) を実現するため
  - 言い換えると、データに対して、各クラスタに属する確率が分かるようにするため

- 混合ガウス分布とは
  - 各ガウス分布の線形の重ね合わせ

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

  - 各ガウス分布 $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ は混合要素とよばれる
  - 各ガウス分布は個別に、平均 $\boldsymbol{\mu}_k$ と共分散 $\boldsymbol{\Sigma}_k$ のパラメータをもつ

混合ガウス分布

- パラメータ π_k を混合係数といい、以下の条件を満たす

$$\sum_k \pi_k = 1 \tag{2}$$

これは、$p(\boldsymbol{x})$ を $\boldsymbol{x}$ について積分すれば明らかである

$$\begin{aligned}
\int p(\boldsymbol{x})d\boldsymbol{x} &= 1 \\
\int \sum_k \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)d\boldsymbol{x} &= 1 \\
\sum_k \pi_k \int \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)d\boldsymbol{x} &= 1 \\
\sum_k \pi_k &= 1
\end{aligned}$$

混合ガウス分布

各ガウス分布 $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ は、正規化されている

$$\forall k \quad \int \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) d\boldsymbol{x} = 1$$

- $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \geq 0$ であるので、$p(\boldsymbol{x}) \geq 0$ となるための十分条件は、全ての k について、$\pi_k \geq 0$ が成立することである

$$\forall k \in \{1, \ldots, K\} \quad \pi_k \geq 0 \Rightarrow p(\boldsymbol{x}) \geq 0$$

- これと $\sum_k \pi_k = 1$ から、結局全ての π_k について以下が成り立つ

$$0 \leq \pi_k \leq 1 \tag{3}$$

混合ガウス分布

混合ガウス分布

$$p(\boldsymbol{x}) = \sum_k \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

$$\sum_k \pi_k = 1, \quad \forall k \quad 0 \leq \pi_k \leq 1$$

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

- 混合ガウス分布を決定づけるパラメータは、$\boldsymbol{\pi} \equiv \{\pi_1, \pi_2, \ldots, \pi_K\}$、$\boldsymbol{\mu} \equiv \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_K\}$、$\boldsymbol{\Sigma} \equiv \{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \boldsymbol{\Sigma}_K\}$

混合ガウス分布

- 混合ガウス分布を導入する理由 (再確認)
  - データに対して、各クラスタに属する確率が分かるようにするため

- 問題設定
  - $\boldsymbol{x}$ の N 個の観測点で構成されるデータ集合 $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ $(\boldsymbol{x}_i \in \mathbb{R}^D)$
  - データ集合 $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ を、K 個のクラスタに分割
  - K は、既知の定数であるとする

  - k 番目のクラスタが、平均 $\boldsymbol{\mu}_k$、共分散行列 $\boldsymbol{\Sigma}_k$ の正規分布 $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ で表現できるとする

  - 各クラスタの分布 $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ を、π_k で重み付けして足し合わせた混合分布が、データ全体を表す分布である

混合ガウス分布

- 最尤推定を試みる
  - 最尤推定によって、混合ガウス分布のパラメータ $\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$ が分かったとする
  - このとき次のようにすれば、クラスタリングが可能

  - 新たなデータ $\boldsymbol{x}$ が得られたとき、全ての $k(k = 1, \ldots, K)$ について $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ を計算する
  - これを最大にするような k が、データ $\boldsymbol{x}$ が属するクラスタである

混合ガウス分布

- 最尤推定を試みる
  - 結論から先に言うと、いきなり最尤推定を試すと失敗する

  - 最尤推定によって、混合ガウス分布 $p(\boldsymbol{x})$ のパラメータ $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}\}$ を求めてみる
  - 尤度関数 $p(\mathcal{D}|\boldsymbol{\theta})$ を、パラメータ $\boldsymbol{\theta}$ の関数とみなして、$\boldsymbol{\theta}$ について最大化することにより、$\boldsymbol{\theta}$ を求めるという考え方

  - 尤度関数 $p(\mathcal{D}|\boldsymbol{\theta})$ は、パラメータ $\boldsymbol{\theta}$ が与えられたときの、データの条件付き確率である
  - パラメータを 1 つに決めたときに、データ $\mathcal{D}$ が得られる確率

混合ガウス分布

- 対数尤度関数 $\ln p(\mathcal{D}|\boldsymbol{\theta})$ は次のようになる

$$
\begin{aligned}
& \ln p(\mathcal{D}|\boldsymbol{\theta}) \\
= & \ln \prod_i p(\boldsymbol{x}_i|\boldsymbol{\theta}) \\
= & \ln \prod_i \left(\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \\
= & \sum_i \ln \left(\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \\
= & \sum_i \ln \left(\sum_k \pi_k \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \right. \\
& \quad \left. \exp\left\{ -\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right\} \right)
\end{aligned}
\tag{4}
$$

混合ガウス分布

- 上式の最初の変形では、各データ $\mathcal{D} = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ は、確率分布 $p(\boldsymbol{x})$ から、独立に得られているという仮定を用いた
- このようなデータ $\mathcal{D}$ を、i.i.d 標本という (independently and identically distributed)

- 対数 $\ln$ は単調増加関数であるため、対数を適用しても、関数の極値は変化しない
- 尤度関数 $p(\mathcal{D}|\boldsymbol{\theta})$ の最大化は、対数尤度関数 $\ln p(\mathcal{D}|\boldsymbol{\theta})$ の最大化と等価

- 重大な問題点
  - 対数関数 $\ln$ の内部に、総和 $(\sum)$ が入ってしまっている
  - log-sum の形状になっているため、これ以上式を簡単にできない!
  - クラスタ数 $K = 1$ であれば、対数 $\ln$ と、ガウス分布の指数 $\exp$ が打ち消し合って、式が簡潔になる

- しかしここでは、このまま最尤推定を続けてみる

混合ガウス分布

- パラメータ $\boldsymbol{\mu}_k$ の最尤推定
  - 対数尤度関数 $\ln p(\mathcal{D}|\boldsymbol{\theta})$ において、$\boldsymbol{\theta}$ はパラメータ (定数) で、$\mathcal{D}$ は変数であるが、実際は $\mathcal{D}$ にはデータが入っているので、パラメータ $\boldsymbol{\theta}$ を変数とみなす
  - $\ln p(\mathcal{D}|\boldsymbol{\theta})$ をパラメータ $\boldsymbol{\mu}_k$ で微分してみる
  - これを 0 と等置することで、最適な $\boldsymbol{\mu}_k$ が満たすべき式が得られる
  - その前に、関数 f の対数の微分について、以下が成立することを確認しておく

$$(\ln f)' = \frac{f'}{f}, \qquad f' = f \cdot (\ln f)' \tag{5}$$

- このとき次のようになる

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\mu}_k} \ln p(\mathcal{D}|\boldsymbol{\theta}) \\
=\ & \frac{\partial}{\partial \boldsymbol{\mu}_k} \sum_i \ln \left(\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \\
=\ & \sum_i \frac{\partial}{\partial \boldsymbol{\mu}_k} \ln \left(\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \\
=\ & \sum_i \frac{1}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{\partial}{\partial \boldsymbol{\mu}_k} \left(\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \\
=\ & \sum_i \frac{1}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{\partial}{\partial \boldsymbol{\mu}_k} \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \\
=\ & \sum_i \frac{1}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \left(\frac{\partial}{\partial \boldsymbol{\mu}_k} \pi_k \right.
\end{aligned}
$$

$$
\begin{aligned}
&\qquad \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right\}\Bigg) \\
= &\sum_i \frac{1}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \pi_k \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \\
&\qquad \frac{\partial}{\partial \boldsymbol{\mu}_k} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right\} \qquad (6)
\end{aligned}
$$

ここで、以下のようにできる

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\mu}_k} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right\} \\
= &\exp\left\{-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right\} \\
&\qquad \frac{\partial}{\partial \boldsymbol{\mu}_k} \left(-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right) \qquad (7)
\end{aligned}
$$

混合ガウス分布

更に、次が成立する

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\mu}_k}\left(-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)\right) \\
= \quad & -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\mu}_k}\left(\boldsymbol{x}_i^T \boldsymbol{\Sigma}_k^{-1}\boldsymbol{x}_i - \boldsymbol{x}_i^T \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k - {\color{red}\boldsymbol{\mu}_k^T \boldsymbol{\Sigma}_k^{-1}\boldsymbol{x}_i} + \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k\right) \\
= \quad & -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\mu}_k}\left(-\boldsymbol{x}_i^T \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k - {\color{red}\boldsymbol{x}_i^T \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k} + \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k\right) \\
= \quad & -\frac{1}{2}\left(-2\boldsymbol{x}_i^T \boldsymbol{\Sigma}_k^{-1} + 2\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k\right) \\
= \quad & \boldsymbol{\Sigma}_k^{-1}\boldsymbol{x}_i - \boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k \\
= \quad & \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)
\end{aligned}
\tag{8}
$$

ここで、以下を用いた

$$
\begin{aligned}
& \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}_k^{-1}\boldsymbol{x}_i \\
= \quad & \left(\boldsymbol{\mu}_k^T \boldsymbol{\Sigma}_k^{-1}\boldsymbol{x}_i\right)^T \quad (\because \text{スカラーであるため転置してもよい})
\end{aligned}
$$

混合ガウス分布

$$
\begin{aligned}
&= \boldsymbol{x}_i^T \left(\boldsymbol{\Sigma}_k^{-1}\right)^T \left(\boldsymbol{\mu}_k^T\right)^T \\
&= \boldsymbol{x}_i^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k \qquad (9)
\end{aligned}
$$

共分散行列 $\boldsymbol{\Sigma}_k$ は対称行列 $(\boldsymbol{\Sigma}_k^T = \boldsymbol{\Sigma}_k)$ であるため、以下が成立

$$
\left(\boldsymbol{\Sigma}_k^{-1}\right)^T = \left(\boldsymbol{\Sigma}_k^T\right)^{-1} = \boldsymbol{\Sigma}_k^{-1} \qquad (10)
$$

各項の微分は次のようになる

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\mu}_k} \boldsymbol{x}_i^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k \\
&= \left(\boldsymbol{x}_i^T \boldsymbol{\Sigma}_k^{-1}\right)^T = \left(\boldsymbol{\Sigma}_k^{-1}\right)^T \left(\boldsymbol{x}_i^T\right)^T = \boldsymbol{\Sigma}_k^{-1} \boldsymbol{x}_i \qquad (11) \\
& \frac{\partial}{\partial \boldsymbol{\mu}_k} \boldsymbol{\mu}_k^T \boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k \\
&= \left(\boldsymbol{\Sigma}_k^{-1} + \left(\boldsymbol{\Sigma}_k^{-1}\right)^T\right) \boldsymbol{\mu}_k = 2\boldsymbol{\Sigma}_k^{-1} \boldsymbol{\mu}_k \qquad (12)
\end{aligned}
$$

混合ガウス分布

結局、対数尤度関数 $\ln p(\mathcal{D}|\boldsymbol{\theta})$ の $\boldsymbol{\mu}_k$ による微分は

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\mu}_k} \ln p(\mathcal{D}|\boldsymbol{\theta}) \\
= \quad & \sum_i \frac{1}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \pi_k \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \\
& \quad \frac{\partial}{\partial \boldsymbol{\mu}_k} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right\} \\
= \quad & \sum_i \frac{1}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \pi_k \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \\
& \quad \exp\left\{-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right\} \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \\
= \quad & \sum_i \frac{\pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \qquad (13) \\
= \quad & 0
\end{aligned}
$$

混合ガウス分布

- 未知のパラメータ $\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$ が、分母と分子の双方に出現する複雑な式
- 直接この連立方程式を解いて、パラメータの最尤推定量を求めるのは難しそうである

- 勾配 $\nabla_{\boldsymbol{\mu}_k} \ln p(\mathcal{D}|\boldsymbol{\theta})$ を利用した最適化も可能である
- この勾配の方向に、パラメータ $\boldsymbol{\mu}_k$ を少しだけ更新する
- ここでは、EM アルゴリズムという別の手法を導出しようとしている

- $\boldsymbol{x}$ のほかに、潜在変数という仮想的な変数 $\boldsymbol{z}$ を導入することで、簡単に解けるようになる

- 上と似たような式が、後ほど登場する

混合ガウス分布

- ここまでの話の流れ
  1. K-Means では、データを単一のクラスタに割り当てた (ハード割り当て)
  2. データが属するクラスタだけではなく、より多くの情報 (各クラスタに属する確率) を手に入れたい
  3. ソフト割り当てを実現するためには、クラスタリングを統計的機械学習 (確率分布) の観点から見直して、再定式化を行う必要があった
  4. 各クラスタをガウス分布として、データ全体を混合ガウス分布に当てはめることを考えた
  5. 混合ガウス分布のパラメータを、最尤推定により求めようとしたが、困難であることが分かった
  6. そこで、潜在変数を導入して、最尤推定を簡単に解こうと考えている

混合ガウス分布

- 潜在変数 $\boldsymbol{z}$ の導入
  - 各データ $\boldsymbol{x}_i$ につき、1 つのベクトル $\boldsymbol{z}_i \in \mathbb{R}^K$ が対応しているとする
  - $\boldsymbol{z}_i$ は、データ $\boldsymbol{x}_i$ が属するクラスタ を表現する

- 潜在変数 $\boldsymbol{z}$ の表現
  - $\boldsymbol{z}_i$ は、K 次元の二値確率変数 $\boldsymbol{z}$ の観測値である
  - $\boldsymbol{z}$ の k 番目の要素を、z_k と表すことにする

  - 確率変数 $\boldsymbol{z}$ は、1-of-K 符号化法により表現されるとする
  - 即ち、ある 1 つの $k \in \{1, \ldots, K\}$ について $z_k = 1$ で、$j \neq k$ に対し $z_j = 0$ となる
  - $z_k (k = 1, \ldots, K)$ は、$z_k \in \{0, 1\}$ かつ $\sum_k z_k = 1$ をみたす

  - ベクトル $\boldsymbol{z}$ は K 種類の状態を取る

混合ガウス分布

- 潜在変数 $\boldsymbol{z}$ の例
  - 例えば、データ点 $\boldsymbol{x}_i$ に対して $\boldsymbol{z}_i$ があるとする
  - $z_{i1} = 1$ $(\boldsymbol{z}_i = [1, 0, 0, \ldots, 0])$ ならば、$\boldsymbol{x}_i$ は 1 番目のクラスタ出身
  - $z_{i2} = 1$ $(\boldsymbol{z}_i = [0, 1, 0, \ldots, 0])$ ならば、$\boldsymbol{x}_i$ は 2 番目のクラスタ出身

- データ $\boldsymbol{x}_i$ が作られるまでの流れ
  - $\boldsymbol{z}$ に関する確率分布 $p(\boldsymbol{z})$ から、$\boldsymbol{z}_i$ がサンプルされる
  - $\boldsymbol{z}$ が与えられた下での条件付き分布 $p(\boldsymbol{x}|\boldsymbol{z})$ から、$\boldsymbol{x}_i$ がサンプルされる

  - 即ち、$\boldsymbol{x}, \boldsymbol{z}$ の同時分布は、周辺分布 $p(\boldsymbol{z})$ と、条件付き分布 $p(\boldsymbol{x}|\boldsymbol{z})$ を用いて次のように書ける

$$p(\boldsymbol{x}, \boldsymbol{z}) = p(\boldsymbol{z}) p(\boldsymbol{x}|\boldsymbol{z}) \tag{14}$$

  - 潜在変数 $\boldsymbol{z}_i$ が最初に決められ、その $\boldsymbol{z}_i$ に応じて $\boldsymbol{x}_i$ が決まると考える
  - $\boldsymbol{z}_i$ は実際には存在しない、仮想的なものである

  - $\boldsymbol{z}_i$ は、実際に観測される $\boldsymbol{x}_i$ の裏側に潜んでいる

混合ガウス分布

- $p(\boldsymbol{x})$ の表現 (予想)
  - $p(\boldsymbol{x})$ は混合ガウス分布になってほしい

$$p(\boldsymbol{x}) = \sum_k \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{15}$$

- 周辺分布 $p(\boldsymbol{z})$ の定義
  - $p(\boldsymbol{z})$ は次のように定める

$$p(\boldsymbol{z}) = \prod_k \pi_k^{z_k}, \quad p(z_k = 1) = \pi_k \tag{16}$$

  - 但し π_k は混合係数であり、$\sum_k \pi_k = 1, 0 \leq \pi_k \leq 1$ をみたす
  - $\boldsymbol{z}$ の表現には 1-of-K 符号化法を使うため、左側のようにも書ける

混合ガウス分布

- 条件付き分布 $p(\boldsymbol{x}|\boldsymbol{z})$ の定義
  - $p(\boldsymbol{x}|\boldsymbol{z})$ は次のように定める

$$p(\boldsymbol{x}|\boldsymbol{z}) = \prod_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)^{z_k} \tag{17}$$

$$p(\boldsymbol{x}|z_k = 1) = \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \tag{18}$$

- $p(\boldsymbol{x})$ の導出
  - $\sum_{\boldsymbol{z}}$ は、可能な全ての $\boldsymbol{z}$ についての総和を取るということ
  - $\boldsymbol{z} = [1, 0, \ldots, 0]^T, [0, 1, 0, \ldots, 0]^T, \ldots, [0, \ldots, 0, 1]^T$ についての和
  - これは、ベクトル $\boldsymbol{z}$ の中で、1 である要素のインデックス k についての総和 $\sum_k$ を取ることに相当

混合ガウス分布

- $p(\boldsymbol{x}, \boldsymbol{z}) = p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z})$ を、$\boldsymbol{z}$ について周辺化すればよい

$$
\begin{aligned}
p(\boldsymbol{x}) &= \sum_{\boldsymbol{z}} p(\boldsymbol{x}, \boldsymbol{z}) && (19) \\
&= \sum_{\boldsymbol{z}} p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z}) && (20) \\
&= \sum_{k} p(z_k = 1)p(\boldsymbol{x}|z_k = 1) && (21) \\
&= \sum_{k} \pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) && (22)
\end{aligned}
$$

- これは、混合ガウス分布と同じ形になっている

- 何が嬉しいのか?
  - 潜在変数を陽に含む表現 $p(\boldsymbol{x}, \boldsymbol{z}) = p(\boldsymbol{z})p(\boldsymbol{x}|\boldsymbol{z})$ を得たことで、この同時分布を使った議論が可能になった

混合ガウス分布

- $p(z_k = 1|\boldsymbol{x})$ の表現
  - $\boldsymbol{x}$ が与えられた下での、$\boldsymbol{z}$ の条件付き確率
  - 実は、$p(z_k = 1|\boldsymbol{x})$ は、データ $\boldsymbol{x}$ がクラスタ k に属する確率 を表す
  - 求めようとしているのは、この値である!

- $\gamma(z_k) = p(z_k = 1|\boldsymbol{x})$ とすると、ベイズの定理から次のように書ける

$$\gamma(z_k) \equiv p(z_k = 1|\boldsymbol{x}) \tag{23}$$

$$= \frac{p(z_k = 1)p(\boldsymbol{x}|z_k = 1)}{\sum_j p(z_j = 1)p(\boldsymbol{x}|z_j = 1)}$$

$$= \frac{\pi_k \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_j \pi_j \mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)} \tag{24}$$

- $\sum_k \gamma(z_k) = 1$ であることに注意

- 潜在変数を導入したので、最尤推定について再度考えてみる

混合ガウス分布

- 最尤推定を再挑戦
  - パラメータ $\boldsymbol{\theta}$ は、$\boldsymbol{\pi} \equiv \{\pi_1, \pi_2, \ldots, \pi_K\}$、$\boldsymbol{\mu} \equiv \{\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_K\}$、$\boldsymbol{\Sigma} \equiv \{\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \ldots, \boldsymbol{\Sigma}_K\}$ をまとめたもの
  - 対数尤度関数 $\ln p(\mathcal{D}|\boldsymbol{\theta})$ は以下に示す通りであった

$$
\begin{aligned}
&\ln p(\mathcal{D}|\boldsymbol{\theta}) \\
=\ &\sum_i \ln \left(\sum_k \pi_k \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \right. \\
&\qquad \left. \exp \left\{ -\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right\} \right)
\end{aligned}
\tag{25}
$$

  - 尤度関数を、$\boldsymbol{\pi}, \boldsymbol{\mu}, \boldsymbol{\Sigma}$ のそれぞれについて最大化する
  - ここでは、尤度関数を最大化する $\boldsymbol{\mu}_k$ が、満たすべき条件を考える

混合ガウス分布

- $\ln p(\mathcal{D}|\boldsymbol{\theta})$ を $\boldsymbol{\mu}_k$ について偏微分して 0 と等置すると、以下を得る

$$\frac{\partial}{\partial \boldsymbol{\mu}_k} \ln p(\mathcal{D}|\boldsymbol{\theta})$$

$$= \sum_i \frac{\pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_k) \tag{26}$$

$$= \sum_i \gamma(z_{ik}) \boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i - \boldsymbol{\mu}_k) = 0 \tag{27}$$

- 途中までは、先程導出したものを利用
- 負担率 $\gamma(z_{ik})$ が現れていることに注意

混合ガウス分布

- 共分散行列 $\boldsymbol{\Sigma}_k$ が正則であると仮定して、両辺に左から掛けて整理する

$$
\begin{aligned}
& \sum_i \gamma(z_{ik})(\boldsymbol{x}_i - \boldsymbol{\mu}_k) = 0 \\
\Rightarrow \quad & \sum_i \gamma(z_{ik})\boldsymbol{\mu}_k = \sum_i \gamma(z_{ik})\boldsymbol{x}_i \\
\Rightarrow \quad & \boldsymbol{\mu}_k \sum_i \gamma(z_{ik}) = \sum_i \gamma(z_{ik})\boldsymbol{x}_i \\
\Rightarrow \quad & \boldsymbol{\mu}_k = \frac{1}{\sum_i \gamma(z_{ik})} \sum_i \gamma(z_{ik})\boldsymbol{x}_i
\end{aligned}
\tag{28}
$$

- これより、$\boldsymbol{\mu}_k$ を導出する式が得られた

混合ガウス分布

- K-Means 法との比較
  - K-Means 法における、平均ベクトル $\boldsymbol{\mu}_k$ の更新式と見比べてみる

$$\boldsymbol{\mu}_k = \frac{1}{\sum_i \gamma(z_{ik})} \sum_i \gamma(z_{ik}) \boldsymbol{x}_i$$
$$\boldsymbol{\mu}_k = \frac{1}{\sum_i r_{ik}} \sum_i r_{ik} \boldsymbol{x}_i \tag{29}$$

  - r_{ik} を、$\gamma(z_{ik})$ に置き換えたものとなっている
  - $\gamma(z_{ik})$ は、データ $\boldsymbol{x}_i$ が、クラスタ k に属する確率である
  - $\gamma(z_{ik})$ を、全てのデータ $\boldsymbol{x}_i$ について足し合わせたもの $\sum_i \gamma(z_{ik})$ は、実質的に、k 番目のクラスタに割り当てられるデータの数を表している (整数になるとは限らない)

混合ガウス分布

- そこで、K-Means 法のときと同じように、N_k を次のように定める

$$N_k = \sum_i \gamma(z_{ik}) \tag{30}$$

このとき、$\boldsymbol{\mu}_k$ の式は

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_i \gamma(z_{ik}) \boldsymbol{x}_i \tag{31}$$

- 例えば、$\gamma(z_{ik})$ が $0, 1$ のいずれかであれば、$\sum_i \gamma(z_{ik})$ は、k 番目のクラスタに属するデータの数と完全に一致
- クラスタ k に対応するガウス分布の平均 $\boldsymbol{\mu}_k$ は、各データ $\boldsymbol{x}_i$ の重み付き平均
- 重み因子は、事後確率 $p(z_k = 1|\boldsymbol{x}_i) \equiv \gamma(z_{ik})$ である

混合ガウス分布

- $\gamma(z_{ik})$ は、$\sum_k \gamma(z_{ik}) = 1$ となることからも分かるように、$\boldsymbol{x}_i$ を生成するために、k 番目のガウス分布が、どの程度貢献したかを表す
- 言い換えると、k 番目のガウス分布が、$\boldsymbol{x}_i$ の出現を説明する度合いである

- この意味で、$\gamma(z_{ik})$ のことを負担率 (Responsibility) という

混合ガウス分布

- 尤度関数を最大化する $\boldsymbol{\Sigma}_k$ の導出
  - $\boldsymbol{\mu}_k$ の場合と同様に、対数尤度関数 $\ln p(\mathcal{D}|\boldsymbol{\theta})$ を、$\boldsymbol{\Sigma}_k$ に関して微分して、0 と等置すればよい
  - かなり導出が長くなるので注意

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \ln p(\mathcal{D}|\boldsymbol{\theta}) \\
= & \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \sum_i \ln \left(\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \\
= & \sum_i \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \ln \left(\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \\
= & \sum_i \frac{1}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \left(\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \right) \\
= & \sum_i \frac{1}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \pi_k \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \qquad (32)
\end{aligned}
$$

混合ガウス分布

ここで、以下の部分を求める

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \pi_k \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \\
= \quad & \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \pi_k \frac{1}{(2\pi)^{\frac{D}{2}}} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right\} \\
= \quad & \pi_k \frac{1}{(2\pi)^{\frac{D}{2}}} \left(\frac{\partial}{\partial \boldsymbol{\Sigma}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \right. \\
& \qquad \left. \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right\} \right) \qquad (33)
\end{aligned}
$$

微分の中身は、$\boldsymbol{\Sigma}_k$ についての合成関数となっている

$$
\frac{\partial}{\partial \boldsymbol{\Sigma}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right\}
$$

混合ガウス分布

- 一般に、行列 $\boldsymbol{X}$ についてのスカラー関数 $f(\boldsymbol{X}), g(\boldsymbol{X})$ があるとき、以下の連鎖律が成立

$$\frac{\partial}{\partial \boldsymbol{X}} f(\boldsymbol{X}) g(\boldsymbol{X}) = f(\boldsymbol{X}) \frac{\partial g(\boldsymbol{X})}{\partial \boldsymbol{X}} + g(\boldsymbol{X}) \frac{\partial f(\boldsymbol{X})}{\partial \boldsymbol{X}} \tag{34}$$

これを利用して、先程の微分を求める

$$\begin{aligned} & \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right\} \\ = \quad & \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \exp\left\{-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right\} + \\ & \exp\left\{-\frac{1}{2}(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)\right\} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \end{aligned} \tag{35}$$

- 各項の微分を順番に求める

$$
\begin{aligned}
& \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \\
= \quad & \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \exp\left(\ln \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \right) \\
= \quad & \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \exp\left(-\frac{1}{2} \ln |\boldsymbol{\Sigma}_k| \right) \\
= \quad & \exp\left(-\frac{1}{2} \ln |\boldsymbol{\Sigma}_k| \right) \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \left(-\frac{1}{2} \ln |\boldsymbol{\Sigma}_k| \right) \\
= \quad & \exp\left(-\frac{1}{2} \ln |\boldsymbol{\Sigma}_k| \right) \left(-\frac{1}{2} \right) \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \ln |\boldsymbol{\Sigma}_k| \\
= \quad & -\frac{1}{2} \exp\left(-\frac{1}{2} \ln |\boldsymbol{\Sigma}_k| \right) \left(\boldsymbol{\Sigma}_k^{-1} \right)^T \\
= \quad & -\frac{1}{2} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \boldsymbol{\Sigma}_k^{-1}
\end{aligned}
\tag{36}
$$

また

$$\begin{aligned} & \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right\} \\ = \quad & \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right\} \\ & \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \left(-\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right) \end{aligned} \tag{37}$$

であって

$$\begin{aligned} & \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \left(-\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right) \\ = \quad & -\frac{1}{2} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \left((\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right) \\ = \quad & -\frac{1}{2} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \operatorname{Tr} \left((\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right) \end{aligned}$$

混合ガウス分布

$$
\begin{aligned}
&= -\frac{1}{2}\frac{\partial}{\partial \boldsymbol{\Sigma}_k}\mathrm{Tr}\left(\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\right) \\
&= -\frac{1}{2}\left(-\left(\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}\right)^T\right) \\
&= \frac{1}{2}\left(\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}\right)^T \\
&= \frac{1}{2}\left(\boldsymbol{\Sigma}_k^{-1}\right)^T\left((\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\right)^T\left(\boldsymbol{\Sigma}_k^{-1}\right)^T \\
&= \frac{1}{2}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1} \qquad (38)
\end{aligned}
$$

のように求まる

- 行列のトレースについて、一般に以下が成り立つことを利用している

$$\mathrm{Tr}\left(\boldsymbol{XY}\right) = \mathrm{Tr}\left(\boldsymbol{YX}\right) \qquad (39)$$

$$\mathrm{Tr}\left(\boldsymbol{XYZ}\right) = \mathrm{Tr}\left(\boldsymbol{YZX}\right) = \mathrm{Tr}\left(\boldsymbol{ZXY}\right) \qquad (40)$$

混合ガウス分布

- また先程の微分では以下の公式を用いている

$$\frac{\partial}{\partial \boldsymbol{X}} \mathrm{Tr}\left(\boldsymbol{X}^{-1}\boldsymbol{Y}\right) = -\left(\boldsymbol{X}^{-1}\boldsymbol{Y}\boldsymbol{X}^{-1}\right)^T \tag{41}$$

- この公式を証明するためには、いくつかの段階を踏む必要がある
- まずは以下の微分公式の導出から始める

$$\frac{\partial}{\partial x}\boldsymbol{AB}$$

上式の微分は、行列の積 $\boldsymbol{AB}$ の i, k 成分について考えれば

$$\begin{aligned} & \frac{\partial}{\partial x}\sum_j A_{ij}B_{jk} \\ = & \sum_j \frac{\partial}{\partial x} A_{ij}B_{jk} \end{aligned}$$

$$
\begin{aligned}
&= \sum_j \left(\frac{\partial A_{ij}}{\partial x} B_{jk} + A_{ij} \frac{\partial B_{jk}}{\partial x} \right) \\
&= \sum_j \frac{\partial A_{ij}}{\partial x} B_{jk} + \sum_j A_{ij} \frac{\partial B_{jk}}{\partial x} \qquad (42)
\end{aligned}
$$

であるから、結局

$$
\frac{\partial}{\partial x} \boldsymbol{AB} = \frac{\partial \boldsymbol{A}}{\partial x} \boldsymbol{B} + \boldsymbol{A} \frac{\partial \boldsymbol{B}}{\partial x} \qquad (43)
$$

となる

混合ガウス分布

- 上記の公式から、以下の公式を簡単に導ける

$$
\begin{aligned}
\frac{\partial}{\partial x}\boldsymbol{A}^{-1}\boldsymbol{A} &= \frac{\partial \boldsymbol{A}^{-1}}{\partial x}\boldsymbol{A} + \boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial x} \\
\frac{\partial}{\partial x}\boldsymbol{I} &= \frac{\partial \boldsymbol{A}^{-1}}{\partial x}\boldsymbol{A} + \boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial x} \\
0 &= \frac{\partial \boldsymbol{A}^{-1}}{\partial x}\boldsymbol{A} + \boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial x}
\end{aligned} \tag{44}
$$

これに右から $\boldsymbol{A}^{-1}$ を掛ければ

$$
\begin{aligned}
0 &= \frac{\partial \boldsymbol{A}^{-1}}{\partial x}\boldsymbol{A}\boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial x}\boldsymbol{A}^{-1} \\
0 &= \frac{\partial \boldsymbol{A}^{-1}}{\partial x}\boldsymbol{I} + \boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial x}\boldsymbol{A}^{-1} \\
0 &= \frac{\partial \boldsymbol{A}^{-1}}{\partial x} + \boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial x}\boldsymbol{A}^{-1}
\end{aligned}
$$

混合ガウス分布

$$\frac{\partial}{\partial x}\boldsymbol{A}^{-1} = -\boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial x}\boldsymbol{A}^{-1} \tag{45}$$

より

$$\frac{\partial}{\partial x}\boldsymbol{A}^{-1} = -\boldsymbol{A}^{-1}\frac{\partial \boldsymbol{A}}{\partial x}\boldsymbol{A}^{-1} \tag{46}$$

を得る (逆行列の微分公式)

- 逆行列 $\boldsymbol{A}^{-1}$ の k, l 成分 $\left(\boldsymbol{A}^{-1}\right)_{kl}$ については、以下のように書ける

$$\begin{aligned}
&\frac{\partial}{\partial x}\left(\boldsymbol{A}^{-1}\right)_{kl} \\
&= -\sum_{m,n}\left(\boldsymbol{A}^{-1}\right)_{km}\left(\frac{\partial \boldsymbol{A}}{\partial x}\right)_{mn}\left(\boldsymbol{A}^{-1}\right)_{ml} \\
&= -\sum_{m,n}\left(\boldsymbol{A}^{-1}\right)_{km}\frac{\partial A_{mn}}{\partial x}\left(\boldsymbol{A}^{-1}\right)_{ml}
\end{aligned} \tag{47}$$

混合ガウス分布

- この逆行列の微分公式を使えば、以下の微分公式を導出できる

$$\frac{\partial}{\partial \boldsymbol{X}} \mathrm{Tr}\left(\boldsymbol{X}^{-1}\boldsymbol{Y}\right)$$

行列 $\boldsymbol{X}$ の i, j 要素による微分を考えれば

$$\begin{aligned}
& \frac{\partial}{\partial X_{ij}} \mathrm{Tr}\left(\boldsymbol{X}^{-1}\boldsymbol{Y}\right) \\
= \; & \frac{\partial}{\partial X_{ij}} \sum_{k,l} \left(\boldsymbol{X}^{-1}\right)_{kl} Y_{lk} \\
= \; & \sum_{k,l} \frac{\partial}{\partial X_{ij}} \left(\left(\boldsymbol{X}^{-1}\right)_{kl}\right) Y_{lk} \\
= \; & \sum_{k,l} \left(- \sum_{m,n} \left(\boldsymbol{X}^{-1}\right)_{km} \frac{\partial X_{mn}}{\partial X_{ij}} \left(\boldsymbol{X}^{-1}\right)_{ml} \right) Y_{lk} \\
= \; & \sum_{k,l} \left(- \left(\boldsymbol{X}^{-1}\right)_{ki} \left(\boldsymbol{X}^{-1}\right)_{jl} \right) Y_{lk}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{k,l} \left(- \left(\boldsymbol{X}^{-1} \right)_{jl} Y_{lk} \left(\boldsymbol{X}^{-1} \right)_{ki} \right) \\
&= - \left(\boldsymbol{X}^{-1} \boldsymbol{Y} \boldsymbol{X}^{-1} \right)_{ji} \\
&= - \left(\left(\boldsymbol{X}^{-1} \boldsymbol{Y} \boldsymbol{X}^{-1} \right)^T \right)_{ij} \qquad (48)
\end{aligned}
$$

であるから

$$
\frac{\partial}{\partial \boldsymbol{X}} \mathrm{Tr} \left(\boldsymbol{X}^{-1} \boldsymbol{Y} \right) = - \left(\boldsymbol{X}^{-1} \boldsymbol{Y} \boldsymbol{X}^{-1} \right)^T \qquad (49)
$$

を得られる

混合ガウス分布

- 尤度関数を最大化する $\boldsymbol{\Sigma}_k$ の導出
  - これより、結局次のようになる

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\Sigma}_k} \ln p(\mathcal{D}|\boldsymbol{\theta}) \\
=\ & \sum_i \frac{1}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k) \\
=\ & \sum_i \frac{1}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \pi_k \frac{1}{(2\pi)^{\frac{D}{2}}} \\
& \left(\frac{\partial}{\partial \boldsymbol{\Sigma}_k} \frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right\} \right) \\
=\ & \sum_i \frac{1}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \pi_k \frac{1}{(2\pi)^{\frac{D}{2}}} \\
& \left(\frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \frac{\partial}{\partial \boldsymbol{\Sigma}_k} \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \boldsymbol{\Sigma}_k^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu}_k) \right\} + \right.
\end{aligned}
$$

$$
\begin{aligned}
& \exp\left\{-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)\right\}\frac{\partial}{\partial\boldsymbol{\Sigma}_k}\frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}}\Bigg) \\
= & \sum_i \frac{1}{\sum_k \pi_k\mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}\pi_k\frac{1}{(2\pi)^{\frac{D}{2}}} \\
& \left(\frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)\right\}\right. \\
& \left(\frac{1}{2}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}\right)- \\
& \left.\frac{1}{2}\frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}}\boldsymbol{\Sigma}_k^{-1}\exp\left\{-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)\right\}\right) \\
= & \sum_i \frac{1}{\sum_k \pi_k\mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}\pi_k\frac{1}{(2\pi)^{\frac{D}{2}}}\frac{1}{|\boldsymbol{\Sigma}_k|^{\frac{1}{2}}} \\
& \exp\left\{-\frac{1}{2}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\left(\frac{1}{2}\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}-\frac{1}{2}\boldsymbol{\Sigma}_k^{-1}\right)\\
= &\sum_i \frac{\pi_k\mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}{\sum_k \pi_k\mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)}\\
&\frac{1}{2}\boldsymbol{\Sigma}_k^{-1}\left((\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}-\boldsymbol{I}\right)\\
= &\frac{1}{2}\sum_i \gamma(z_{ik})\boldsymbol{\Sigma}_k^{-1}\left((\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\boldsymbol{\Sigma}_k^{-1}-\boldsymbol{I}\right)=0 \qquad (50)
\end{aligned}
$$

両辺に左右から $\boldsymbol{\Sigma}_k$ を掛けて、整理すれば

$$
\begin{aligned}
&\frac{1}{2}\sum_i \gamma(z_{ik})\left((\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T-\boldsymbol{\Sigma}_k\right)=0\\
\Rightarrow\ &\sum_i \gamma(z_{ik})\boldsymbol{\Sigma}_k=\sum_i \gamma(z_{ik})(\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T\\
\Rightarrow\ &\boldsymbol{\Sigma}_k=\frac{1}{\sum_i \gamma(z_{ik})}\sum_i \gamma(z_{ik})(\boldsymbol{x}_i-\boldsymbol{\mu}_k)(\boldsymbol{x}_i-\boldsymbol{\mu}_k)^T \qquad (51)
\end{aligned}
$$

混合ガウス分布

$$\Rightarrow \quad \boldsymbol{\Sigma}_k = \frac{1}{N_k}\sum_i \gamma(z_{ik})(\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \tag{52}$$

- これより、$\boldsymbol{\Sigma}_k$ を導出する式が得られた

混合ガウス分布

- 尤度関数を最大化する π_k の導出
  - $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$ の場合と同様に、対数尤度関数 $\ln p(\mathcal{D}|\boldsymbol{\theta})$ を、π_k に関して微分して、0 と等置すればよい
  - 但し、$\sum_k \pi_k = 1$ という制約条件を考慮しなければならない
  - そのため、ラグランジュの未定係数法を用いる

  - 以下を最大化する π_k を求める

$$\ln p(\mathcal{D}|\boldsymbol{\theta}) + \lambda \left(\sum_k \pi_k - 1 \right) \tag{53}$$

混合ガウス分布

- π_k で微分して 0 と等置すると、次のようになる

$$
\begin{aligned}
& \frac{\partial}{\partial \pi_k}\left(\ln p(\mathcal{D}|\boldsymbol{\theta}) + \lambda\left(\sum_k \pi_k - 1\right)\right) \\
= & \frac{\partial}{\partial \pi_k}\left(\sum_i \ln\left(\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\right) + \lambda\left(\sum_k \pi_k - 1\right)\right) \\
= & \sum_i \frac{\mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} + \lambda \qquad (54) \\
= & \sum_i \frac{\gamma(z_{ik})}{\pi_k} + \lambda = 0 \qquad (55)
\end{aligned}
$$

両辺に π_k を掛けて

$$\sum_i \gamma(z_{ik}) + \lambda \pi_k = 0 \tag{56}$$

混合ガウス分布

k についての和を取ると

$$\sum_k \left(\sum_i \gamma(z_{ik}) + \lambda \pi_k \right) = 0 \tag{57}$$

$$\sum_i \sum_k \gamma(z_{ik}) + \lambda \sum_k \pi_k = 0 \tag{58}$$

$$\sum_i 1 + \lambda = 0 \tag{59}$$

$$N + \lambda = 0 \tag{60}$$

$$\therefore \lambda = -N \tag{61}$$

これより

$$\sum_i \gamma(z_{ik}) + (-N)\pi_k = 0 \tag{62}$$

$$\pi_k = \frac{1}{N} \sum_i \gamma(z_{ik}) = \frac{N_k}{N} \tag{63}$$

混合ガウス分布

ここで、以下が成立することに注意

$$N_k = \sum_i \gamma(z_{ik}), \quad 1 = \sum_k \gamma(z_{ik})$$

- これより、混合係数 π_k は、全ての要素における、クラスタ k の負担率 $\gamma(z_{ik})$ の平均である
- ここまでで、対数尤度関数 $\ln p(\mathcal{D}|\boldsymbol{\theta})$ を最大化するような、$\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k$ の式が得られた

混合ガウス分布

$\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k$ の更新式

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_i \gamma(z_{ik}) \boldsymbol{x}_i \tag{64}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_i \gamma(z_{ik}) (\boldsymbol{x}_i - \boldsymbol{\mu}_k)(\boldsymbol{x}_i - \boldsymbol{\mu}_k)^T \tag{65}$$

$$\pi_k = \frac{N_k}{N} \tag{66}$$

$$N_k = \sum_i \gamma(z_{ik}) \tag{67}$$

混合ガウス分布

$\gamma(z_{ik})$ の更新式

$$\gamma(z_{ik}) = \frac{\pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)}{\sum_k \pi_k \mathcal{N}(\boldsymbol{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)} \tag{68}$$

- 注意点
  - $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k$ の更新式は、これらのパラメータについての、陽な解は与えていない
  - なぜなら、これらの更新式は全て、負担率 $\gamma(z_{ik})$ に依存しているため
  - そしてその負担率 $\gamma(z_{ik})$ は、$\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k$ の全てに依存する

混合ガウス分布

- これらの更新式の意味
  - 最尤推定の解を求めるための、繰り返し手続きの存在を示唆
  - 即ち、$\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k$ の初期化後に、(1) $\gamma(z_{ik})$ の更新と、(2) それを用いた $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k$ の更新という、2 段階の処理を繰り返す手続き
  - これは、混合ガウス分布を確率モデルとして使ったときの、EM アルゴリズムとなっている
  - 混合ガウス分布に対する EM アルゴリズムは重要なので、次にまとめる

混合ガウス分布

混合ガウス分布に対する EM アルゴリズム

- 目的は、混合ガウスモデルが与えられているとき、そのパラメータ (各ガウス分布の平均、分散、そして混合係数) について、尤度関数を最大化することである

1. 平均 $\boldsymbol{\mu}_k^{\mathrm{old}}$、分散 $\boldsymbol{\Sigma}_k^{\mathrm{old}}$、そして混合係数 π_k^{old} を初期化し、対数尤度 $\ln p(\mathcal{D}|\boldsymbol{\theta})$ の初期値を計算
2. E ステップ: 現在のパラメータを用いて、負担率 $\gamma(z_{ik})$ を計算

$$\gamma(z_{ik}) \leftarrow \frac{\pi_k^{\mathrm{old}} \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k^{\mathrm{old}}, \boldsymbol{\Sigma}_k^{\mathrm{old}})}{\sum_k \pi_k^{\mathrm{old}} \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k^{\mathrm{old}}, \boldsymbol{\Sigma}_k^{\mathrm{old}})} \tag{69}$$

3 M ステップ: 現在の負担率 $\gamma(z_{ik})$ を用いて、パラメータを更新

$$\boldsymbol{\mu}_k^{\mathrm{new}} \leftarrow \frac{1}{N_k}\sum_i \gamma(z_{ik})\boldsymbol{x}_i \tag{70}$$

$$\boldsymbol{\Sigma}_k^{\mathrm{new}} \leftarrow \frac{1}{N_k}\sum_i \gamma(z_{ik})(\boldsymbol{x}_i - \boldsymbol{\mu}_k^{\mathrm{new}})(\boldsymbol{x}_i - \boldsymbol{\mu}_k^{\mathrm{new}})^T \tag{71}$$

$$\pi_k^{\mathrm{new}} \leftarrow \frac{N_k}{N} \tag{72}$$

但し

$$N_k = \sum_i \gamma(z_{ik}) \tag{73}$$

混合ガウス分布

4. 対数尤度 $\ln p(\mathcal{D}|\boldsymbol{\theta})$ を計算

$$\ln p(\mathcal{D}|\boldsymbol{\theta}) = \sum_i \ln \left(\sum_k \pi_k^{\text{new}} \mathcal{N}(\boldsymbol{x}_i | \boldsymbol{\mu}_k^{\text{new}}, \boldsymbol{\Sigma}_k^{\text{new}}) \right) \tag{74}$$

   パラメータの変化量、あるいは対数尤度の変化量をみて、収束性を判定

5. 収束基準を満たしていなければ、(2) に戻る

$$\boldsymbol{\mu}_k^{\text{old}} \leftarrow \boldsymbol{\mu}_k^{\text{new}}, \quad \boldsymbol{\Sigma}_k^{\text{old}} \leftarrow \boldsymbol{\Sigma}_k^{\text{new}}, \quad \pi_k^{\text{old}} \leftarrow \pi_k^{\text{new}} \tag{75}$$

混合ガウス分布

- EM アルゴリズムの概要
  - E ステップ (Expectation step) では、事後確率 $p(z_k = 1|\boldsymbol{x}_i)$、即ち負担率 $\gamma(z_{ik})$ を計算
  - M ステップ (Maximization step) では、事後確率を使って、各パラメータ $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k$ を再計算

- EM アルゴリズムでの注意点
  - 上記 (3) の M ステップにおける、各パラメータの計算順序に注意
  - 最初に新しい平均値 $\boldsymbol{\mu}_k^{\mathrm{new}}$ を計算し、その新しい平均値を使って、新しい共分散行列 $\boldsymbol{\Sigma}_k^{\mathrm{new}}$ を計算する
  - E ステップと M ステップは、対数尤度関数 $\ln p(\mathcal{D}|\boldsymbol{\theta})$ を増加させることが保証されている
  - EM アルゴリズムは、K-Means 法と比べて、収束までに必要な繰り返し回数と、各ステップでの計算量が非常に多くなる

混合ガウス分布

- 混合ガウス分布の良い初期値を見つけるために、最初に K-Means 法を実行し、その後に EM アルゴリズムを利用する、という方法がある

- K-Means 法により得られた平均ベクトル $\boldsymbol{\mu}_k$ を、各ガウス分布の平均 $\boldsymbol{\mu}_k$ の初期値とする
- 各クラスタに属するデータ点の標本分散を、共分散行列 $\boldsymbol{\Sigma}_k$ の初期値とする
- 各クラスタに属するデータ点の割合を、混合係数 π_k の初期値とする

- 一般に対数尤度には、多数の局所解が存在するため、その中で最大のもの (大域的最適解) に収束するとは限らない

混合ガウス分布のまとめ

- 混合ガウス分布を導入する理由
  - 曖昧さを含んだ、各データ点のクラスタへの割り当て (ソフト割り当て) を実現するため
  - 各データ点について、各クラスタに属する確率が分かるようにするため

- 混合ガウス分布による表現
  - 各クラスタがガウス分布 $\mathcal{N}(\boldsymbol{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$ に従うと仮定
  - 各ガウス分布を、混合係数 π_k で重み付けして足し合わせることで、混合ガウス分布を作り、データ全体を表現する
  - パラメータ $\boldsymbol{\theta}$ は、各ガウス要素の平均 $\boldsymbol{\mu}_k$ と共分散行列 $\boldsymbol{\Sigma}_k$、そして混合係数 π_k である

混合ガウス分布のまとめ

- ここまでの話の流れ
  - 最尤推定 (対数尤度 $\ln p(\mathcal{D}|\boldsymbol{\theta})$ の最大化) によって、パラメータ $\boldsymbol{\theta}$ を求める試みは失敗した
  - 対数の中に総和が入っているせいで、対数尤度の式が複雑になっていた

  - 潜在変数 $\boldsymbol{z}$ を導入し、$\boldsymbol{z}$ に関する分布を考えたことで、混合ガウス分布に対する EM アルゴリズムを自然に導出した

  - 事後確率 $p(z_k = 1|\boldsymbol{x}_i)$ 即ち負担率 $\gamma(z_{ik})$ の計算と、パラメータ $\boldsymbol{\theta}$ の更新という 2 段階の処理を、交互に繰り返していくアルゴリズムであった