混合分布

正田 備也

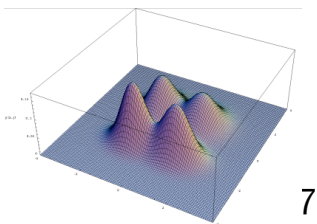
masada@rikkyo.ac.jp

Contents

これまでのモデリングの問題点

- これまでは、データ集合 $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N\}$ 全体に対して、一つの確率分布を使うモデリングだけについて議論していた
- しかし、データ集合は、たった一つの分布ではモデリングし切れない多様性を含んでいることが多い
  - 例えば、数値データの集合であれば、それに近い数値が頻繁に出現するという数値が、複数あったりする

例. 多峰性をもつ multimodal データ集合



混合分布

- これまでは、全てのデータ $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ が同じ分布から生成されると仮定していた
- 一方、混合分布によるモデリングでは、同じ種類の分布だがパラメータの値が違う分布を K 個用意する
  - これらの分布をコンポーネントと呼ぶことがある
- そして、各データ $\boldsymbol{x}_i$ について…

1. カテゴリカル分布 $\mathsf{Cat}(\boldsymbol{\theta})$ に従って K 個の分布から一つ選ぶ
   - $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$ は混合分布のパラメータ。
   - θ_k は k 番目の分布が選ばれる確率。$\sum_{k=1}^{K} \theta_k = 1$ が成り立つ
2. そして、$\boldsymbol{x}_i$ の値がその選ばれた分布に従うと考える。

Contents

混合正規分布

- 混合正規分布を使ったモデリングでは、各 $\boldsymbol{x}_i$ が以下のように生成されると仮定する

1. カテゴリカル分布 $\mathrm{Cat}(\boldsymbol{\theta})$ から、確率変数 z_i の値を draw する
   - 「$z_i = k$」は、$\boldsymbol{x}_i$ については k 番目のコンポーネントが選ばれた、ということを意味する
2. その z_i の値に対応する確率分布から、$\boldsymbol{x}_i$ を draw する
   - $\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k$ は k 番目のコンポーネントである正規分布のパラメータ

$$
\begin{aligned}
z_i &\sim \mathrm{Cat}(\boldsymbol{\theta}) \\
\boldsymbol{x}_i &\sim \mathcal{N}(\boldsymbol{\mu}_{z_i}, \boldsymbol{\Sigma}_{z_i})
\end{aligned}
\tag{1}
$$

単変量正規分布の混合分布の場合

- K 個のコンポーネントのなかから一つを選ぶ際に使われるカテゴリカル分布のパラメータは $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$
  - θ_k は k 番目のコンポーネントが選ばれる確率
  - $\sum_{k=1}^{K} \theta_k = 1$ が成り立つ
- K 個の単変量正規分布をコンポーネントとして用意する
- k 番目の分布のパラメータは、平均 μ_k と標準偏差 σ_k
  - つまり、k 番目のコンポーネントの確率密度関数は

$$p(x; \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(- \frac{(x - \mu_k)^2}{2\sigma_k^2} \right) \tag{2}$$

単変量正規分布の混合分布における同時分布

- 観測されたデータを表す確率変数 $\mathcal{X} = \{x_1, \ldots, x_N\}$ と、コンポーネントへの所属を表す確率変数 $\mathcal{Z} = \{z_1, \ldots, z_N\}$ との同時分布は

$$p(\mathcal{X}, \mathcal{Z}; \boldsymbol{\theta}, \{\mu_k\}, \{\sigma_k\}) = \prod_{i=1}^{N} p(z_i; \theta_{z_i}) p(x_i | z_i; \mu_{z_i}, \sigma_{z_i})$$

$$= \prod_{i=1}^{N} \left[\theta_{z_i} \times \frac{1}{\sqrt{2\pi\sigma_{z_i}^2}} \exp\left(- \frac{(x_i - \mu_{z_i})^2}{2\sigma_{z_i}^2} \right) \right] \tag{3}$$

  - z_i は x_i が属するコンポーネントを表す
  - $p(z_i)$ はカテゴリカル分布 $\mathrm{Cat}(\boldsymbol{\theta})$ の pmf から計算される z_i の尤度
  - $p(x_i|z_i)$ は正規分布 $\mathcal{N}(\mu_{z_i}, \sigma_{z_i})$ の pdf から計算される x_i の尤度

Contents

教師ありの設定の場合 (1/2)

- 教師ありの設定で、混合分布のパラメータを最尤推定する
- 教師ありの設定の場合、各データ x_i について、それがどのコンポーネントから生成されたかは、すでに分かっている
- 言い換えれば、z_i の値も観測データに含まれる
  - つまり、観測データを $\mathcal{D}$ で表すとすると、
    $\mathcal{D} = \{(x_1, z_1), \ldots, (x_N, z_N)\}$
  - x_i も z_i も、すでに観測されており、よって、固定されている
- このとき、式 (3) の同時分布が、そのまま、観測データ $\mathcal{D}$ の尤度を表すことになる

教師ありの設定の場合 (2/2)

- そして、観測データ $\mathcal{D} = \{(x_1, z_1), \ldots, (x_N, z_N)\}$ の尤度は、下のように書き直すことができる

$$p(\mathcal{D}; \boldsymbol{\theta}, \{\mu_k\}, \{\sigma_k\}) = \prod_{i=1}^{N} p(z_i; \theta_{z_i}) p(x_i | z_i; \mu_{z_i}, \sigma_{z_i})$$

$$= \prod_{i=1}^{N} \left[\theta_{z_i} \times \frac{1}{\sqrt{2\pi\sigma_{z_i}^2}} \exp\left(- \frac{(x_i - \mu_{z_i})^2}{2\sigma_{z_i}^2} \right) \right]$$

$$= \prod_{k=1}^{K} \left[\theta_k^{c_k} \times \frac{1}{(\sqrt{2\pi\sigma_k^2})^{c_k}} \exp\left(- \frac{\sum_{\{i:z_i=k\}} (x_i - \mu_k)^2}{2\sigma_k^2} \right) \right] \quad (4)$$

  - c_k は、k 番目のコンポーネントに属するデータの個数

教師ありの場合の混合正規分布の最尤推定

- 混合正規分布のパラメータを、式 (4) の観測データの尤度を最大化することによって推定する方法を、以下に示す

目的関数を $L(\boldsymbol{\theta}, \{\mu_k\}, \{\sigma_k\})$ とおくと、

$$
\begin{aligned}
L(\boldsymbol{\theta}, \{\mu_k\}, \{\sigma_k\}) &= \ln p(\mathcal{D}; \boldsymbol{\theta}, \{\mu_k\}, \{\sigma_k\}) + \lambda\left(1 - \sum_{k=1}^{K} \theta_k\right) \\
&= \sum_{k=1}^{K} c_k \ln \theta_k - \sum_{k=1}^{K} c_k \ln \sigma_k - \sum_{k=1}^{K} \frac{\sum_{\{i:z_i=k\}} (x_i - \mu_k)^2}{2\sigma_k^2} + \lambda\left(1 - \sum_{k=1}^{K} \theta_k\right) + const. \qquad (5)
\end{aligned}
$$

目的関数 L を、各パラメータで偏微分する。

$$
\frac{\partial L}{\partial \mu_k} = \frac{\sum_{\{i:z_i=k\}} (x_i - \mu_k)}{\sigma_k^2} = \frac{\sum_{\{i:z_i=k\}} x_i - c_k \mu_k}{\sigma_k^2} \qquad (6)
$$

$\frac{\partial L}{\partial \mu_k} = 0$ より、$\mu_k = \frac{\sum_{\{i:z_i=k\}} x_i}{c_k} = \bar{x}_k$ を得る。

$$
\frac{\partial L}{\partial \sigma_k} = -\frac{c_k}{\sigma_k} + \frac{\sum_{\{i:z_i=k\}} (x_i - \mu_k)^2}{\sigma_k^3} \qquad (7)
$$

$\frac{\partial L}{\partial \sigma_k} = 0$ より、$\sigma_k^2 = \frac{\sum_{\{i:z_i=k\}} (x_i - \bar{x}_k)^2}{c_k}$ を得る。

$$\frac{\partial L}{\partial \theta_k} = \frac{c_k}{\theta_k} - \lambda \ , \ \frac{\partial L}{\partial \lambda} = 1 - \sum_{k=1}^{K} \theta_k \tag{8}$$

$\frac{\partial L}{\partial \theta_k} = 0$ より、$\theta_k = \frac{c_k}{\lambda}$ を得る。
$\frac{\partial L}{\partial \lambda} = 0$ より、$1 - \sum_{k=1}^{K} \frac{c_k}{\lambda} = 0$ を得る。
つまり、$\lambda = \sum_k c_k$ が言えるので、$\theta_k = \frac{c_k}{\sum_k c_k}$ を得る。

まとめると、

- θ_k は、k 番目のコンポーネントから生成されたデータの割合となる。
- μ_k と σ_k は、k 番目のコンポーネントから生成されたデータだけの尤度をもとに最尤推定した値となる。

Contents

教師なしの設定の場合

- 教師なしの設定の場合、各データ x_i について、それがどのコンポーネントから生成されたかは、分からない！
- すなわち、z_i は潜在変数 latent variables
  - つまり、$\mathcal{X} = \{x_1, \ldots, x_N\}$ だけが観測変数
  - 一方、潜在変数の集合を $\mathcal{Z} = \{z_1, \ldots, z_N\}$ とする
- 観測変数と潜在変数の同時分布 $p(\mathcal{X}, \mathcal{Z})$ は、式 (3) と同じで

$$p(\mathcal{X}, \mathcal{Z}; \boldsymbol{\theta}, \{\mu_k\}, \{\sigma_k\}) = \prod_{i=1}^{N} p(z_i; \theta_{z_i}) p(x_i | z_i; \mu_{z_i}, \sigma_{z_i})$$

$$= \prod_{i=1}^{N} \left[\theta_{z_i} \times \frac{1}{\sqrt{2\pi\sigma_{z_i}^2}} \exp\left(- \frac{(x_i - \mu_{z_i})^2}{2\sigma_{z_i}^2} \right) \right]$$

教師なしの場合の観測データの尤度

- 潜在変数を含むモデリングの場合、観測データの尤度 $p(\mathcal{X})$ は、潜在変数 $\mathcal{Z}$ を周辺化してはじめて得られる
  - 潜在変数を周辺化する＝潜在変数がとりうる値の全てを考慮する

$$
\begin{aligned}
p(\mathcal{X}) &= \sum_{\mathcal{Z}} p(\mathcal{X}, \mathcal{Z}) = \sum_{z_1=1}^{K} \sum_{z_2=1}^{K} \cdots \sum_{z_{N-1}=1}^{K} \sum_{z_N=1}^{K} p(\mathcal{X}, \mathcal{Z}) \\
&= \sum_{z_1=1}^{K} \sum_{z_2=1}^{K} \cdots \sum_{z_{N-1}=1}^{K} \sum_{z_N=1}^{K} \prod_{i=1}^{N} p(x_i, z_i) \\
&= \prod_{i=1}^{N} \left(\sum_{z_i=1}^{K} p(x_i, z_i) \right) = \prod_{i=1}^{N} \left(\sum_{z_i=1}^{K} p(z_i) p(x_i | z_i) \right) \quad (9)
\end{aligned}
$$

教師なしの場合の観測データの対数尤度

- よって、対数尤度は、式 (3) と式 (9) より

$$\ln p(\mathcal{X}) = \ln \sum_{\mathcal{Z}} p(\mathcal{X}, \mathcal{Z}) = \ln \prod_{i=1}^{N} \left(\sum_{z_i=1}^{K} p(x_i, z_i) \right)$$

$$= \ln \prod_{i=1}^{N} \left(\sum_{z_i=1}^{K} p(z_i) p(x_i|z_i) \right) = \sum_{i=1}^{N} \ln \left(\sum_{z_i=1}^{K} p(z_i) p(x_i|z_i) \right)$$

$$= \sum_{i=1}^{N} \ln \left(\sum_{z_i=1}^{K} \left[\theta_{z_i} \times \frac{1}{\sqrt{2\pi\sigma_{z_i}^2}} \exp \left(- \frac{(x_i - \mu_{z_i})^2}{2\sigma_{z_i}^2} \right) \right] \right) \quad (10)$$

観測データの対数尤度の最大化？

- あとは、対数尤度 $\ln p(\mathcal{X}) = \sum_{i=1}^{N} \ln \left(\sum_{z_i=1}^{K} p(z_i) p(x_i|z_i) \right)$ を最大にする $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$ や $\mu_1, \ldots, \mu_K$ や $\sigma_1, \ldots, \sigma_K$ を求めれば良い・・・？？？
  - 式 (10) をそのまま最大化することはしない

積の対数と和の対数

- 何かを掛け算したものの対数は、何かの対数の和に書き直せるので、扱いやすい

$$\log(a \times b) = \log(a) + \log(b) \tag{11}$$

- 何かを足し算したものの対数は、それ以上変形のしようがないので、扱いにくい

$$\log(a + b) = \ldots \tag{12}$$

イェンセンの不等式（対数関数の場合）

- $p_1, \ldots, p_K$ を、$\sum_{k=1}^{K} p_k = 1$ を満たす正の実数とする
- また、$x_1, \ldots, x_K$ を正の実数とする
- このとき、以下の不等式が成り立つ

$$\ln\left(\sum_{k=1}^{K} p_k x_k\right) \geq \sum_{k=1}^{K} p_k \ln(x_k) \tag{13}$$

  - 和の対数（扱いにくい！）の下界 lower bound を、対数の和（扱いやすい！）として得るため、イェンセンの不等式をよく使う
  - なお、対数関数に限らず、上に凸な関数なら、上の不等式は成立

観測データの対数尤度の下界

- イェンセンの不等式を利用して $\ln p(\mathcal{X})$ の下界を得たい
- そこで、各 x_i について $\boldsymbol{q}_i \equiv (q_{i,1}, \ldots, q_{i,K})$ という $\sum_k q_{i,k} = 1$ を満たす変数を用意すると

$$\ln p(\mathcal{X}) = \sum_{i=1}^{N} \ln \Big(\sum_{z_i=1}^{K} p(z_i) p(x_i|z_i) \Big)$$

$$= \sum_{i=1}^{N} \ln \Big(\sum_{z_i=1}^{K} q_{i,z_i} \frac{p(z_i) p(x_i|z_i)}{q_{i,z_i}} \Big) \geq \sum_{i=1}^{N} \sum_{z_i=1}^{K} q_{i,z_i} \ln \frac{p(z_i) p(x_i|z_i)}{q_{i,z_i}} \tag{14}$$

- この下界を $\mathcal{L}(\{\boldsymbol{q}_i\}, \boldsymbol{\theta}, \{\mu_k\}, \{\sigma_k\})$ と書くことにする

観測データの対数尤度の下界の最大化

- $\ln p(\mathcal{X})$ の代わりに $\mathcal{L}(\{\boldsymbol{q}_i\}, \boldsymbol{\theta}, \{\mu_k\}, \{\sigma_k\})$ を最大化することによって、次の未知量を推定する
  - 新たに導入した $\{\boldsymbol{q}_i\} \equiv \{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N\}$ where $\boldsymbol{q}_i = (q_{i,1}, \ldots, q_{i,K})$
  - モデルパラメータ $\boldsymbol{\Theta} \equiv \{\boldsymbol{\theta}, \{\mu_k\}, \{\sigma_k\}\}$
- いま考えている単変量正規分布の混合分布の場合

$$p(z_i = k) = \theta_k \tag{15}$$

$$p(x_i | z_i = k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(- \frac{(x_i - \mu_k)^2}{2\sigma_k^2} \right) \tag{16}$$

- これらを式 (14) に当てはめることで、以下、推定計算を行う

$$
\begin{aligned}
&L(\{\boldsymbol{q}_i\}, \boldsymbol{\theta}, \{\mu_k\}, \{\sigma_k\}) \\
&= \mathcal{L}(\{\boldsymbol{q}_i\}, \boldsymbol{\theta}, \{\mu_k\}, \{\sigma_k\}) + \sum_{i=1}^{N} \lambda_i \left(1 - \sum_{k=1}^{K} q_{i,k}\right) + \lambda_0 \left(1 - \sum_{k=1}^{K} \theta_k\right) \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} q_{i,k} \ln \frac{\theta_k p(x_i | z_i = k)}{q_{i,k}} + \sum_{i=1}^{N} \lambda_i \left(1 - \sum_{k=1}^{K} q_{i,k}\right) + \lambda_0 \left(1 - \sum_{k=1}^{K} \theta_k\right) \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} q_{i,k} \ln \left(\theta_k p(x_i | z_i = k)\right) - \sum_{i=1}^{N} \sum_{k=1}^{K} q_{i,k} \ln q_{i,k} + \sum_{i=1}^{N} \lambda_i \left(1 - \sum_{k=1}^{K} q_{i,k}\right) + \lambda_0 \left(1 - \sum_{k=1}^{K} \theta_k\right)
\end{aligned}
\tag{17}
$$

$$
\frac{\partial L}{\partial q_{i,k}} = \ln \left(\theta_k p(x_i | z_i = k)\right) - \ln q_{i,k} - 1 - \lambda_i \tag{18}
$$

$\frac{\partial L}{\partial q_{i,k}} = 0$ と $\sum_k q_{i,k} = 1$ より $q_{i,k} = \frac{\theta_k p(x_i | z_i = k)}{\sum_k \theta_k p(x_i | z_i = k)}$ を得る。

$q_{i,k} \ln \left(\theta_k p(x_i|z_i = k)\right) = q_{i,k} \ln \theta_k + q_{i,k} \ln p(x_i|z_i = k)$ より、

$$\frac{\partial L}{\partial \theta_k} = \frac{\sum_{i=1}^{N} q_{i,k}}{\theta_k} - \lambda_0 \tag{19}$$

$\frac{\partial L}{\partial \theta_k} = 0$ と $\sum_k \theta_k = 1$ より $\theta_k = \frac{\sum_{i=1}^{N} q_{i,k}}{\sum_{k=1}^{K} \sum_{i=1}^{N} q_{i,k}} = \frac{\sum_{i=1}^{N} q_{i,k}}{N}$ を得る。

$\frac{\partial}{\partial \mu_k} \ln p(x_i|z_i = k) = \frac{x_i - \mu_k}{\sigma_k^2}$ と $\frac{\partial}{\partial \sigma_k} \ln p(x_i|z_i = k) = -\frac{1}{\sigma_k} + \frac{(x_i - \mu_k)^2}{\sigma_k^3}$ より

$$\frac{\partial L}{\partial \mu_k} = \frac{\sum_{i=1}^{N} q_{i,k}(x_i - \mu_k)}{\sigma_k^2} \tag{20}$$

$$\frac{\partial L}{\partial \sigma_k} = \frac{\sum_{i=1}^{N} q_{i,k}(-\sigma_k^2 + (x_i - \mu_k)^2)}{\sigma_k^3} \tag{21}$$

$\frac{\partial L}{\partial \mu_k} = 0$ より $\mu_k = \frac{\sum_{i=1}^{N} q_{i,k} x_i}{\sum_{i=1}^{N} q_{i,k}}$ を得る。また、$\frac{\partial L}{\partial \sigma_k} = 0$ より $\sigma_k^2 = \frac{\sum_{i=1}^{N} q_{i,k}(x_i - \mu_k)^2}{\sum_{i=1}^{N} q_{i,k}}$ を得る。

混合正規分布のためのEMアルゴリズム

- E step

$$q_{i,k} \leftarrow \frac{\theta_k p(x_i|z_i=k)}{\sum_k \theta_k p(x_i|z_i=k)} \tag{22}$$

ただし $p(x_i|z_i=k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}\right)$

- M step

$$\theta_k \leftarrow \frac{\sum_{i=1}^N q_{i,k}}{N} \tag{23}$$

$$\mu_k \leftarrow \frac{\sum_{i=1}^N q_{i,k} x_i}{\sum_{i=1}^N q_{i,k}} \tag{24}$$

$$\sigma_k^2 \leftarrow \frac{\sum_{i=1}^N q_{i,k}(x_i-\mu_k)^2}{\sum_{i=1}^N q_{i,k}} \tag{25}$$

$q_{i,k}$ とは何なのか (1/4)

- イェンセンの不等式を使って、次の下界を得たのだった

$$\sum_{i=1}^{N} \ln \Big(\sum_{z_i=1}^{K} p(z_i) p(x_i|z_i) \Big) \geq \sum_{i=1}^{N} \sum_{z_i=1}^{K} q_{i,z_i} \ln \frac{p(z_i) p(x_i|z_i)}{q_{i,z_i}}$$

- 左辺から右辺を引くと

$$\sum_{i=1}^{N} \ln \Big(\sum_{z_i=1}^{K} p(z_i) p(x_i|z_i) \Big) - \sum_{i=1}^{N} \sum_{z_i=1}^{K} q_{i,z_i} \ln \frac{p(z_i) p(x_i|z_i)}{q_{i,z_i}}$$

$$= \sum_{i=1}^{N} \sum_{z_i=1}^{K} q_{i,z_i} \ln \Big(\sum_{z_i=1}^{K} p(z_i) p(x_i|z_i) \Big) - \sum_{i=1}^{N} \sum_{z_i=1}^{K} q_{i,z_i} \ln \frac{p(z_i) p(x_i|z_i)}{q_{i,z_i}}$$

$q_{i,k}$ とは何なのか (2/4)

（続き）

$$
\begin{aligned}
&= \sum_{i=1}^{N}\sum_{z_i=1}^{K} q_{i,z_i} \ln \left(\sum_{z_i=1}^{K} p(x_i, z_i) \right) - \sum_{i=1}^{N}\sum_{z_i=1}^{K} q_{i,z_i} \ln \frac{p(z_i)p(x_i|z_i)}{q_{i,z_i}} \\
&= \sum_{i=1}^{N}\sum_{z_i=1}^{K} q_{i,z_i} \ln p(x_i) - \sum_{i=1}^{N}\sum_{z_i=1}^{K} q_{i,z_i} \ln \frac{p(x_i, z_i)}{q_{i,z_i}} \\
&= \sum_{i=1}^{N}\sum_{z_i=1}^{K} q_{i,z_i} \ln \frac{p(x_i)q_{i,k}}{p(x_i, z_i)} = \sum_{i=1}^{N}\sum_{z_i=1}^{K} q_{i,z_i} \ln \frac{q_{i,z_i}}{p(z_i|x_i)}
\end{aligned}
\tag{26}
$$

- $q_{i,k} = p(z_i = k|x_i)$ のとき、等号が成立する

カルバック・ライブラー情報量

- p, q を離散確率分布とすると、q から p への（p の q に対する）カルバック・ライブラー情報量 Kullback – Leibler divergence とは

$$D_{\mathrm{KL}}(p \parallel q) = \sum_x p(x) \ln \frac{p(x)}{q(x)} \tag{27}$$

- p, q が連続確率分布の場合は

$$D_{\mathrm{KL}}(p \parallel q) = \int p(x) \ln \left(\frac{p(x)}{q(x)} \right) dx \tag{28}$$

- $p = q$ ならば、またそのときに限り $D_{\mathrm{KL}}(p \parallel q) = 0$

注. $q(x) = 0$ なのに $p(x) \neq 0$ となる x があってはいけない！

$q_{i,k}$ とは何なのか (3/4)

- $q_i(z_i)$ を、K 個のアイテム $\{1, \ldots, K\}$ 上に定義されたカテゴリカル分布とし、$q_i(z_i = k) \equiv q_{i,k}$ と設定する
- 式 (26) は $q_i(z_i)$ の $p(z_i|x_i)$ に対するカルバック・ライブラー情報量になっている
- ところで、式 (22) より、

$$q_{i,k} = \frac{\theta_k p(x_i|z_i = k)}{\sum_k \theta_k p(x_i|z_i = k)} = \frac{p(z_i = k)p(x_i|z_i = k)}{\sum_k p(z_i = k)p(x_i|z_i = k)}$$
$$= \frac{p(x_i, z_i = k)}{\sum_k p(x_i, z_i = k)} = \frac{p(x_i, z_i = k)}{p(x_i)} = p(z_i = k|x_i) \quad (29)$$

$q_{i,k}$ とは何なのか (4/4)

- つまり、式 (22) は $q_{i,k} = p(z_i = k|x_i)$ を意味している
- このとき、式 (26) のカルバック・ライブラー情報量はゼロ！
- ただし、$q_{i,k}$ は、モデルパラメータ $\boldsymbol{\theta}, \{\mu_k\}, \{\sigma_k\}$ の値を固定し、$\ln p(x_i)$ の下界を最大化することで求めたものである
  - つまり、モデルパラメータの特定の値について式 (26) をゼロにできるだけで、$\ln p(x_i)$ が最大化できているわけではないが…
- E ステップでは、その固定されたパラメータ値に対しては、各 $q_{i,k}$ について最善の答えを得ていることは確か
- M ステップでは、$\{\boldsymbol{q}_i\}$ の値を固定し、$\ln p(x_i)$ の下界をできるだけ大きくすべく、パラメータの値を求めている

E step で何をしているか

- モデルのパラメータ $\boldsymbol{\Theta} \equiv \{\boldsymbol{\theta}, \mu_1, \ldots, \mu_K, \sigma_1, \ldots, \sigma_K\}$ の値を固定した状態で、対数尤度の下界 $\mathcal{L}(\boldsymbol{\Theta})$ を最大化する $\{\boldsymbol{q}_i\}$ を求めているのが、E step
- パラメータ $\boldsymbol{\Theta}$ の、固定された値を、$\boldsymbol{\Theta}_{\text{old}}$ と書くことにする
- その最大化によって得られる答えは

$$q_{i,k} = p(z_i = k | x_i; \boldsymbol{\Theta}_{\text{old}}) \tag{30}$$

- つまり、モデルパラメータ $\boldsymbol{\Theta}$ の値を固定したうえで、観測データを所与とする潜在変数の条件付き分布を求めている

Contents

多項分布の混合分布

- 第 k コンポーネントのパラメータを $\boldsymbol{\phi}_k$ とする
  - $\boldsymbol{\phi}_k = (\phi_{k,1}, \ldots, \phi_{k,W})$
  - $\phi_{k,w}$ は第 k コンポーネントに属する文書での単語 v_w の出現確率
- $\ln p(\mathcal{X})$ の代わりに $\mathcal{L}(\{q_{i,k}\}, \boldsymbol{\theta}, \{\boldsymbol{\phi}_k\})$ を最大化
- 多項分布の混合分布の場合

$$p(z_i = k) = \theta_k \tag{31}$$

$$p(\mathcal{D}_i | z_i = k) = \frac{n_i!}{\prod_w c_{i,w}!} \prod_{w=1}^{W} \phi_{k,w}^{c_{i,w}} \tag{32}$$

- 以下、推定計算を行う

$$
\begin{aligned}
&L(\{q_{i,k}\}, \boldsymbol{\theta}, \{\boldsymbol{\phi}_k\}) \\
&= \mathcal{L}(\{q_{i,k}\}, \boldsymbol{\theta}, \{\boldsymbol{\phi}_k\}) + \sum_{i=1}^{N} \lambda_i \left(1 - \sum_{k=1}^{K} q_{i,k}\right) + \lambda_0 \left(1 - \sum_{k=1}^{K} \theta_k\right) + \sum_{k=1}^{K} \nu_k \left(1 - \sum_{w=1}^{W} \phi_{k,w}\right) \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} q_{i,k} \ln \frac{\theta_k p(\mathcal{D}_i | z_i = k)}{q_{i,k}} + \sum_{i=1}^{N} \lambda_i \left(1 - \sum_{k=1}^{K} q_{i,k}\right) + \lambda_0 \left(1 - \sum_{k=1}^{K} \theta_k\right) + \sum_{k=1}^{K} \nu_k \left(1 - \sum_{w=1}^{W} \phi_{k,w}\right) \\
&= \sum_{i=1}^{N} \sum_{k=1}^{K} q_{i,k} \ln \left(\theta_k p(\mathcal{D}_i | z_i = k)\right) - \sum_{i=1}^{N} \sum_{k=1}^{K} q_{i,k} \ln q_{i,k} \\
&\quad + \sum_{i=1}^{N} \lambda_i \left(1 - \sum_{k=1}^{K} q_{i,k}\right) + \lambda_0 \left(1 - \sum_{k=1}^{K} \theta_k\right) + \sum_{k=1}^{K} \nu_k \left(1 - \sum_{w=1}^{W} \phi_{k,w}\right) \qquad (33)
\end{aligned}
$$

$$
\frac{\partial L}{\partial q_{i,k}} = \ln \left(\theta_k p(\mathcal{D}_i | z_i = k)\right) - \ln q_{i,k} - 1 - \lambda_i \qquad (34)
$$

$\frac{\partial L}{\partial q_{i,k}} = 0$ と $\sum_k q_{i,k} = 1$ より $q_{i,k} = \frac{\theta_k p(\mathcal{D}_i | z_i = k)}{\sum_k \theta_k p(\mathcal{D}_i | z_i = k)}$ を得る。

$q_{i,k} \ln \left(\theta_k p(\mathcal{D}_i | z_i = k) \right) = q_{i,k} \ln \theta_k + q_{i,k} \ln p(\mathcal{D}_i | z_i = k)$ より、

$$\frac{\partial L}{\partial \theta_k} = \frac{\sum_{i=1}^{N} q_{i,k}}{\theta_k} - \lambda_0 \tag{35}$$

$\frac{\partial L}{\partial \theta_k} = 0$ と $\sum_k \theta_k = 1$ より $\theta_k = \frac{\sum_{i=1}^{N} q_{i,k}}{\sum_{k=1}^{K} \sum_{i=1}^{N} q_{i,k}} = \frac{\sum_{i=1}^{N} q_{i,k}}{N}$ を得る。

$\frac{\partial}{\partial \phi_{k,w}} \ln p(\mathcal{D}_i | z_i = k) = \frac{c_{i,w}}{\phi_{k,w}}$ より

$$\frac{\partial L}{\partial \phi_{k,w}} = \frac{\sum_{i=1}^{N} q_{i,k} c_{i,w}}{\phi_{k,w}} - \nu_k \tag{36}$$

$\frac{\partial L}{\partial \phi_{k,w}} = 0$ より $\phi_{k,w} = \frac{\sum_{i=1}^{N} q_{i,k} c_{i,w}}{\sum_{w=1}^{W} \sum_{i=1}^{N} q_{i,k} c_{i,w}}$ を得る。

混合多項分布のためのEMアルゴリズム

- E step

$$q_{i,k} \leftarrow \frac{\theta_k p(x_i|z_i = k)}{\sum_k \theta_k p(x_i|z_i = k)} \tag{37}$$

ただし $p(x_i|z_i = k) = \frac{n_i!}{\prod_w c_{i,w}!} \prod_{w=1}^{W} \phi_{k,w}^{c_{i,w}}$、つまり

$$q_{i,k} \leftarrow \frac{\theta_k \prod_w \phi_{k,w}^{c_{i,w}}}{\sum_k \theta_k \prod_w \phi_{k,w}^{c_{i,w}}}$$

- M step

$$\theta_k \leftarrow \frac{\sum_{i=1}^{N} q_{i,k}}{N} \tag{38}$$

$$\phi_{k,w} \leftarrow \frac{\sum_{i=1}^{N} q_{i,k} c_{i,w}}{\sum_{w=1}^{W} \sum_{i=1}^{N} q_{i,k} c_{i,w}} \tag{39}$$