9.19

ノートのタイトル 2016/04/26

(確率変数 x の分布(9.85)について)

$x_{ij} \in \{0, 1\}$

$\sum_j x_{ij} = 1 \Longrightarrow$ ベクトルの成分 x_i は 1 of M で記されているというイミ

要素を並べてベクトルとして表現すると

$$x = (\underbrace{\overbrace{x_{11}\ x_{12}\ x_{13}}^{x_1\text{成分}}}_{M\text{個}}, \underbrace{\overbrace{x_{21}\ x_{22}\ x_{23}}^{x_2\text{成分}}}_{M\text{個}}, \underbrace{\overbrace{x_{31}\ x_{32}\ x_{33}}^{x_3\text{成分}}}_{M\text{個}}) \leftarrow D \times M \text{次元のベクトルになる。}$$

D組

M は状態数, D は次元数を表わす。x は3次元のデータで多次元を 1 of 3状態で表わすという感じ

x の分布は

$(x_{11}, x_{12}, x_{13}), (x_{21}, x_{22}, x_{23}), (x_{31}, x_{32}, x_{33})$ は独立なので

$$p(x) = p(x_{11}, x_{12}, x_{13}, x_{21}, x_{22}, x_{23}, x_{31}, x_{32}, x_{33}) = p(x_{11}, x_{12}, x_{13}) \cdot p(x_{21}, x_{22}, x_{23}) \cdot p(x_{31}, x_{32}, x_{33}) = \prod_{i=1}^{D} p(x_i) \quad \cdots Ⓐ$$

となる。

2.2章より多項分布は

$$(2.34)\quad \mathrm{Mult}(m_1, \cdots m_K \mid \mu, N) = \frac{N!}{m_1! \cdots m_K!} \prod_{k=1}^{K} \mu_k^{m_k}$$

$$(2.36)\quad \sum_{k=1}^{K} m_k = N$$

これを $p(x_i)$ の分布に使う

$$\sum_{j=1}^{M} x_{ij} = N = 1$$

なので

$x_{ij} = 0$ or 1 なので $x_{ij}! = 1$

$$p(x_i \mid \mu_{ki}) = \mathrm{Mult}(x_{i1}, \cdots x_{iM} \mid \mu_{ki}, N=1) = \frac{1!}{x_{i1}! \cdots x_{iM}!} \prod_{j=1}^{M} \mu_{kij}^{x_{ij}} = \prod_{j=1}^{M} \mu_{kij}^{x_{ij}}$$

となる。

Ⓐに入れて

$$p(x \mid \mu_k) = \prod_{i=1}^{D} \prod_{j=1}^{M} \mu_{kij}^{x_{ij}} \quad \cdots (9.85)$$

を得る。

2.2章より多項分布において μ_{kij} は $x_{ij} = 1$ となる確率である。

確率なので $\sum_j \mu_{kij} = 1, \mu_{kij} \geq 0$ という制約を満たす。($\sum_j \mu_{kij} = 1, \mu_{kij} \geq 0$ より $\mu_{kij} \leq 1$ も言える)

(EMアルゴリズム)

完全データの尤度は

$(x_1, z_1), (x_2, z_2) \cdots$ は独立なので

$$p(X, Z | \mu, \pi) = p(x_1, x_2 \cdots z_1, z_2 \cdots | \mu, \pi) = p(x_1, z_1 | \mu, \pi)\, p(x_2, z_2 | \mu, \pi) \cdots p(x_N, z_N | \mu, \pi)$$

$$= \prod_{n=1}^{N} p(x_n, z_n | \mu, \pi)$$

$$= \prod_{n=1}^{N} p(x_n | z_n, \mu)\, p(z_n | \pi)$$

$$= \prod_{n=1}^{N} \prod_{k=1}^{K} \{ p(x_n | \mu_k)\, \pi_k \}^{z_{nk}} \longleftarrow (9.52), (9.53)$$

$$= \prod_{n=1}^{N} \prod_{k=1}^{K} \prod_{i=1}^{D} \prod_{j=1}^{M} \left(\mu_{kij}^{x_{nij}}\, \pi_k \right)^{z_{nk}} \longleftarrow (9.85)$$

よって

$$\ln p(X, Z | \mu, \pi) = \sum_n \sum_k \sum_i \sum_j z_{nk} \{ \ln \pi_k + x_{nij} \ln \mu_{kij} \}$$

$$= \sum_n \sum_k z_{nk} \{ \ln \pi_k + \sum_i \sum_j x_{nij} \ln \mu_{kij} \} \longleftarrow ①$$

また

$$p(Z | X, \mu, \pi) \propto p(X, Z | \mu, \pi)$$

なので

$$\gamma(z_{nk}) = E_{Z|X}[z_{nk}] = \sum_Z \frac{p(X, Z | \mu, \pi)\, z_{nk}}{\sum_Z p(X, Z | \mu, \pi)}$$

$$= \frac{\sum_Z p(X, Z | \mu, \pi)\, z_{nk}}{\sum_Z p(X, Z | \mu, \pi)}$$

$\sum_Z (\cdots) = \sum_{z_1} \cdots \sum_{z_N} (\cdots)$

$$= \frac{\sum_{z_1} \cdots \sum_{z_n} \cdots \sum_{z_N} z_{nk} \prod_{m=1}^{N} p(x_m, z_m | \mu, \pi)}{\sum_{z_1} \cdots \sum_{z_n} \cdots \sum_{z_N} \prod_{m=1}^{N} p(x_m, z_m | \mu, \pi)} \longleftarrow p(X, Z | \mu, \pi) = \prod_{n=1}^{N} p(x_n, z_n | \mu, \pi)$$

うまく $\prod_m$ の各項が z_m でカりしてるので
$\sum_{z_n}$ と $\prod_m$ を交換できる

$$= \frac{\sum_{z_1} p(x_1 z_1 | \mu \pi) \cdots \sum_{z_n} z_{nk}\, p(x_n z_n | \mu \pi) \cdots \sum_{z_N} p(x_N z_N | \mu \pi)}{\sum_{z_1} p(x_1 z_1 | \mu \pi) \cdots \sum_{z_n} p(x_n z_n | \mu \pi) \cdots \sum_{z_N} p(x_N z_N | \mu \pi)}$$

$$= \frac{\sum_{z_n} z_{nk}\, p(x_n, z_n | \mu \pi)}{\sum_{z_n} p(x_n, z_n | \mu \pi)}$$

$$= \frac{\sum_{z_n} z_{nk}\, p(x_n | z_n, \mu)\, p(z_n | \pi)}{\sum_{z_n} p(x_n | z_n, \mu)\, p(z_n | \pi)}$$

$$= \frac{\sum_{z_n} z_{nk} \prod_{l=1}^{K} \{ p(x_n | \mu_l)\, \pi_l \}^{z_{nl}}}{\sum_{z_n} \prod_{l=1}^{K} \{ p(x_n | \mu_l)\, \pi_l \}^{z_{nl}}} \longleftarrow \begin{array}{l} (9.52)\ p(x | z, \mu) = \prod_{k=1}^{K} p(x | \mu_k)^{z_k} \\ (9.53)\ p(z | \pi) = \prod_{k=1}^{K} \pi_k^{z_k} \end{array}$$

$$= \frac{z_{nk} \{ p(x_n | \mu_k)\, \pi_k \}^{z_{nk}}}{\sum_{l=1}^{K} \{ p(x_n | \mu_l)\, \pi_l \}^{z_{nl}}}$$

$\sum_{z_n}$ は $z_n = (1, 0, \cdots), (0, 1, \cdots)$ の和なので
分子は $z_{nk} = 1$ となる項しか残らない
分母は $\prod_{l=1}^{K} (\cdots)^{z_{nl}}$ の総積の各項のうち $z_{nl} = 1$ の項のみ残る

$$= \frac{z_{nk} \{ \prod_{i=1}^{D} \prod_{j=1}^{M} \mu_{kij}^{x_{nij}}\, \pi_k \}^{z_{nk}}}{\sum_{l=1}^{K} \{ \prod_{i=1}^{D} \prod_{j=1}^{M} \mu_{lij}^{x_{nij}}\, \pi_l \}^{z_{nl}}} \longleftarrow (9.85)$$

となる。

また

$$E_{Z|X}[\ln p(X,Z|\mu,\pi)] = \sum_Z \frac{p(X,Z|\mu,\pi)\ln p(X,Z|\mu,\pi)}{\sum_Z p(X,Z|\mu,\pi)}$$

$$= \frac{\sum_Z p(X,Z|\mu,\pi)\ln p(X,Z|\mu,\pi)}{\sum_Z p(X,Z|\mu,\pi)}$$

①

$$= \frac{\sum_Z p(X,Z|\mu,\pi)\sum_n\sum_k z_{nk}\{\ln \pi_k + \sum_i\sum_j x_{nij}\ln \mu_{kij}\}}{\sum_Z p(X,Z|\mu,\pi)}$$

$$= \sum_n\sum_k \left\{\frac{\sum_Z p(X,Z|\mu,\pi)\, z_{nk}}{\sum_Z p(X,Z|\mu,\pi)}\right\}\{\ln \pi_k + \sum_i\sum_j x_{nij}\ln \mu_{kij}\}$$

$$= \sum_n\sum_k E_{Z|X}[z_{nk}]\{\ln \pi_k + \sum_i\sum_j x_{nij}\ln \mu_{kij}\}$$

$$= \sum_n\sum_k \gamma(z_{nk})\{\ln \pi_k + \sum_i\sum_j x_{nij}\ln \mu_{kij}\}$$

となる。

(μ_kについて $E_{Z|X}[\ln p(X,Z|\mu,\pi)]$ を最大化する)

$$\sum_j \mu_{kij} = 1,\quad k=1\sim K,\ i=1\sim D$$

という制約があるので、μ_kについての最大化する条件は以下のようになる

$$\frac{\partial}{\partial \mu_{kij}}\left\{E_{Z|X}[\ln p(X,Z|\mu,\pi)] + \sum_k^K\sum_i^D \lambda_{ki}\left(\sum_j \mu_{kij}-1\right)\right\}=0$$

$$\therefore\quad \sum_n \gamma(z_{nk})\, x_{nij}\frac{1}{\mu_{kij}} + \lambda_{ki} = 0$$

両辺にμ_{kij}をかけて$\sum_j$すると

$$\sum_n \gamma(z_{nk})\underbrace{\sum_j x_{nij}}_{=1} + \lambda_{ki}\underbrace{\sum_j \mu_{kij}}_{=1} = 0$$

$$\therefore\quad \lambda_{ki} = -\sum_n \gamma(z_{nk}) = -N_k \quad \longleftarrow (9.57)$$

元の式に戻して

$$\mu_{kij} = \frac{1}{N_k}\sum_n \gamma(z_{nk})x_{nij}$$

ij成分表記をベクトル表記にすると

$$\mu_k = \frac{1}{N_k}\sum_n \gamma(z_{nk})x_n = \bar{x}_k \quad \longleftarrow (9.58)$$

となる。

(π_kについて $E_{Z|X}[\ln p(X,Z|\mu,\pi)]$ を最大化する)

$$\sum_k \pi_k = 1$$

という制約があるので、π_kについて最大化する条件は以下のようになる

$$\frac{\partial}{\partial \pi_k}\left\{E_{Z|X}[\ln p(X,Z|\mu,\pi)] + \lambda\left(\sum_k \pi_k - 1\right)\right\}=0$$

$$\therefore\quad \sum_n \gamma(z_{nk})\frac{1}{\pi_k} + \lambda = 0$$

両辺にπ_kをかけて$\sum_k$すると

$$\sum_k\sum_n \gamma(z_{nk}) + \lambda\sum_k \pi_k = 0$$

$$\therefore\quad \lambda = -N$$

元の式に戻して

$$\pi_k = \frac{1}{N}\sum_n \gamma(z_{nk}) = \frac{N_k}{N} \quad \longleftarrow (9.57)$$

を得る。