9.9

ノートのタイトル 2016/04/18

(Σ_kについて)

(9.40) $E[\ln p(X,Z)] = \sum_{n=1}^{N}\sum_{j=1}^{K} \gamma(z_{nj})\{\ln\pi_j + \ln N(x_n|\mu_j,\Sigma_j)\}$

Σ_kについての微分を0として、Eを極大にする解を求める

$$\frac{\partial E}{\partial \Sigma_k} = \sum_{n=1}^{N}\gamma(z_{nk})\frac{\partial}{\partial\Sigma_k}\ln N(x_n|\mu_k\Sigma_k) = 0$$

ここで

$$\frac{\partial}{\partial\Sigma_k}\ln N(x_n|\mu_k\Sigma_k) = \frac{\frac{\partial}{\partial\Sigma_k}N(x_n|\mu_k\Sigma_k)}{N(x_n|\mu_k\Sigma_k)}$$

$$= -\frac{1}{2}\{\Sigma_k^{-1} - \Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\Sigma_k^{-1}\} \leftarrow \text{右記の③}$$

なので

$$\sum_{n=1}^{N}\gamma(z_{nk})\left(-\frac{1}{2}\right)\{\Sigma_k^{-1} - \Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\Sigma_k^{-1}\} = 0$$

左右からΣ_kをかけて

$\gamma(z_{nk})$はスカラーなので $\Sigma_k\gamma = \gamma\Sigma_k$と交換してよい

$$\sum_{n=1}^{N}\gamma(z_{nk})\{\Sigma_k - (x_n-\mu_k)(x_n-\mu_k)^T\} = 0$$

$$\therefore \Sigma_k = \frac{\sum_{n=1}^{N}\gamma(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T}{\sum_{n=1}^{N}\gamma(z_{nk})}$$

$$= \frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(x_n-\mu_k)(x_n-\mu_k)^T$$

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

を得る。これは、(9.19)と同じ式になっている。

$Y = |\Sigma_k|^{-\frac{1}{2}}$ を対数微分する

$$\ln Y = -\frac{1}{2}\ln|\Sigma_k|$$

$$\therefore \frac{Y'}{Y} = -\frac{1}{2}\mathrm{Tr}\left(\Sigma_k^{-1}\frac{\partial\Sigma_k}{\partial\Sigma_{kij}}\right) \leftarrow (C.22), \quad Y' = \frac{\partial Y}{\partial\Sigma_{kij}}$$

$$= -\frac{1}{2}\mathrm{Tr}(\Sigma_k^{-1}J_{ij}), \quad J_{ij}: \text{シングルエントリ行列}$$

$$= -\frac{1}{2}\Sigma_{kji}^{-1}$$

$$= -\frac{1}{2}\Sigma_{kij}^{-1} \leftarrow \Sigma_k\text{対称なので}\Sigma_k^{-1}\text{も対称}$$

よって

$$Y' = \frac{\partial|\Sigma_k|^{-\frac{1}{2}}}{\partial\Sigma_{kij}} = -\frac{1}{2}\Sigma_{kij}^{-1}|\Sigma_k|^{-\frac{1}{2}}$$

ij成分を行列にまとめて

$$\frac{\partial|\Sigma_k|^{-\frac{1}{2}}}{\partial\Sigma_k} = -\frac{1}{2}\Sigma_k^{-1}|\Sigma_k|^{-\frac{1}{2}} \quad \cdots ①$$

$$\frac{\partial}{\partial\Sigma_{kij}}(x_n-\mu_k)^T\Sigma_k^{-1}(x-\mu_k) = \frac{\partial}{\partial\Sigma_{kij}}\mathrm{Tr}[(x_n-\mu_k)^T\Sigma_k^{-1}(x-\mu_k)] \leftarrow \text{スカラーなので Trとっても同じ}$$

$$= \frac{\partial}{\partial\Sigma_{kij}}\mathrm{Tr}[\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T] \leftarrow (C.9)\ \text{Trの循環}$$

$$= \mathrm{Tr}\left[\frac{\partial\Sigma_k^{-1}}{\partial\Sigma_{kij}}(x_n-\mu_k)(x_n-\mu_k)^T\right] \leftarrow \text{Trと微分の交換は成分計算でOK}$$

$$= \mathrm{Tr}\left[-\Sigma_k^{-1}\frac{\partial\Sigma_k}{\Sigma_{kij}}\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\right] \leftarrow (C.11)$$

$$= -\mathrm{Tr}[J_{ij}\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\Sigma_k^{-1}]$$

$$= -[\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\Sigma_k^{-1}]_{ji}$$

$$= -[\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\Sigma_k^{-1}]_{ij} \leftarrow [\cdots]\text{の中が対称なので}$$

ij成分を行列にまとめて

$$\frac{\partial}{\partial\Sigma_k}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k) = -\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\Sigma_k^{-1} \quad \cdots ②$$

$$\frac{\partial}{\partial\Sigma_k}N(x_n|\mu_k\Sigma_k) = \frac{\partial}{\partial\Sigma_k}\frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma_k|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)\right\}$$

$$= \frac{1}{(2\pi)^{D/2}}\left\{\frac{\partial}{\partial\Sigma_k}|\Sigma_k|^{-\frac{1}{2}}\right\}\exp\{\cdots\} + \frac{1}{(2\pi)^{D/2}}\frac{1}{|\Sigma_k|^{\frac{1}{2}}}\exp\{\cdots\}\frac{\partial}{\partial\Sigma_k}\left\{-\frac{1}{2}(x_n-\mu_k)^T\Sigma_k^{-1}(x_n-\mu_k)\right\}$$

$$= -\frac{1}{2}\Sigma_k^{-1}N(x_n|\mu_k\Sigma_k) + N(x_n|\mu_k\Sigma_k)\left(-\frac{1}{2}\right)(-1)\Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\Sigma_k^{-1}$$

$$= -\frac{1}{2}N(x_n|\mu_k\Sigma_k)\{\Sigma_k^{-1} - \Sigma_k^{-1}(x_n-\mu_k)(x_n-\mu_k)^T\Sigma_k^{-1}\} \quad \cdots ③$$

(次は π_k について.)

$\sum_i \pi_i = 1$ という制約の下で (9.40) を最大化させる。

ラグランジアン

$$L = E[\ln p(X, Z)] + \lambda\left(\sum_i \pi_i - 1\right)$$

を微分して 0 とおけば、所望の π_k が得られる。

$$\frac{\partial L}{\partial \pi_k} = \frac{\partial}{\partial \pi_k}\left[\sum_{n=1}^{N}\sum_{j=1}^{K}\gamma(z_{nj})\{\ln \pi_j + \ln N(x_n|\mu_j, \Sigma_j)\} + \lambda\left(\sum_j \pi_j - 1\right)\right]$$

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

なので

$$\frac{\sum_{n=1}^{N}\gamma(z_{nk})}{\pi_k} + \lambda = 0$$

両辺に π_k をかけて k について総和をとると

$$\sum_{k=1}^{K}\sum_{n=1}^{N}\gamma(z_{nk}) + \lambda\underbrace{\sum_{k=1}^{K}\pi_k}_{1} = 0$$

総和の順番は交換しても同じ

$$\sum_{k=1}^{K}\sum_{n=1}^{N}\gamma(z_{nk}) = \sum_{k=1}^{K}\sum_{n=1}^{N}\frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_n|\mu_j, \Sigma_j)} = \sum_{n=1}^{N}\sum_{k=1}^{K}\frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_n|\mu_j, \Sigma_j)}$$

$$= \sum_{n=1}^{N}\frac{\sum_{k=1}^{K}\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{j=1}^{K}\pi_j N(x_n|\mu_j, \Sigma_j)} = \sum_{n=1}^{N} 1 = N$$

分母は k に依らないので $\sum_k$ の外に出して良い

$$\therefore\ N + \lambda = 0$$

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

これを元の式に戻して

$$\frac{\sum_{n=1}^{N}\gamma(z_{nk})}{\pi_k} - N = 0$$

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

$$= \frac{N_k}{N}$$

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

を得る。これは、(9.22) と同じ式になっている。