(A.6)の導出

[5.4.2]から
↓
$\frac{\partial J}{\partial \mu_k}=0$, $\frac{\partial J}{\partial \Sigma_k}=0$ の理由
↓
[(A.6)まで
↓
(A.6)の導出
↳ デルタ関数の(A.1)についての説明
↳ (A.5)=(A.6)
↓
更に(A.4)の導出 (p.265)
L (...) (A.9)
L $A=\Sigma^{-1}$なら 対称行列だから
L $\frac{\partial}{\partial z_k}(z^T A z) = (2Az)_k$
$\Leftrightarrow \frac{\partial}{\partial z}(z^T \Sigma^{-1} z) = 2\Sigma^{-1} z$
↓ (...)
(A.10)
↓
$\frac{\partial z}{\partial \mu} = -I$
↓
(A.4)
↓ Σ
$\partial\Sigma$の推定など N=1の場合
L (A.11), (A.12) はおいて
↓
(A.15)まで
↓
(3.12)を示す
↓
(A.11)の証明

$$\mathcal{N}(x;\mu,\Sigma) = \frac{1}{\sqrt{(2\pi)^D|\Sigma|}} \exp\left\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right\} \quad (A.1)$$

→ $-\infty \sim \infty$で積分すると1

$$L(\mu,\Sigma) = \log \prod_{n=1}^{N} \mathcal{N}(x^{(n)};\mu,\Sigma) \quad (A.2)$$

尤度＝「パラメータをθとして与えられたとき、サンプルDが得られる確率密度」

‖

「連続型確率分布」をとる場合、（xが連続型の確率変数の場合）、 p(x)は確率密度とよばれる

$$\log p(D;\mu,\sigma)$$
$$= \log \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x^{(n)}-\mu)^2}{2\sigma^2}\right\}$$
$$= \log \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}\sigma} + \underline{\log \prod_{n=1}^{N} \exp\left\{-\frac{(x^{(n)}-\mu)^2}{2\sigma^2}\right\}}$$

$\log\exp(\overset{(1)}{\sim}) + \log\exp(\overset{(2)}{\sim}) + \cdots$

$$= \log\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N + \underline{\sum_{n=1}^{N} \frac{-(x^{(n)}-\mu)^2}{2\sigma^2}}$$
$$= -\frac{N}{2}\log 2\pi\sigma^2 - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x^{(n)}-\mu)^2$$

← $f(\mu)$とみるとき、2次係数は $-\frac{1}{2\sigma^2} < 0$（上に凸）

$= g(\sigma)$とする。

$\boxed{\log\sigma = \alpha}$とおくと、

$\sigma = e^{\alpha}$, $\frac{1}{\sigma^2} = e^{-2\alpha}$

を代入して

今は無視

$$\log p(D;\mu,\sigma) = -\frac{N}{2}\log 2\pi - \frac{N}{2}\cdot 2\alpha - \frac{1}{2\sigma^2}\sum_{n=1}^{N}(x^{(n)}-\mu)^2$$
$$= -N\alpha - \frac{1}{2}e^{-2\alpha}\sum_{n=1}^{N}(x^{(n)}-\mu)^2$$
$$\frac{d}{d\alpha} = -N + e^{-2\alpha}\sum_{n=1}^{N}(x^{(n)}-\mu)^2$$
$$\frac{d^2}{d\alpha^2} = -2e^{-2\alpha}\sum_{n=1}^{N}(x^{(n)}-\mu)^2 \underset{\sim\sim}{< 0}$$

2階微分が負 ⟺ 上に凸

$\frac{\partial L}{\partial \Sigma} = 0$ を調べる段

さて $\frac{\partial L}{\partial \mu} = 0$, $\frac{\partial L}{\partial \Sigma} = 0$ を計算する。

$$L(\mu,\Sigma) = \log\prod_{n=1}^{N}\mathcal{N}(x^{(n)};\mu,\Sigma)$$
$$= \sum_{n=1}^{N}\log\mathcal{N}(x^{(n)};\mu,\Sigma)$$
$$= \sum_{n=1}^{N}\log\left(\frac{1}{(2\pi)^D|\Sigma|}\exp\left\{-\frac{1}{2}(x^{(n)}-\mu)^T\Sigma^{-1}(x^{(n)}-\mu)\right\}\right)$$
$$= \sum_{n=1}^{N}\log\frac{1}{\sqrt{(2\pi)^D|\Sigma|}} + \sum_{n=1}^{N}\ln\exp\left\{-\frac{1}{2}(x^{(n)}-\mu)^T\Sigma^{-1}(x^{(n)}-\mu)\right\}$$

$$= \underbrace{N \log (2\pi)^{-\frac{D}{2}}}_{= -\frac{N}{2}\log(2\pi)^D} + \underbrace{N \log |\Sigma|^{-\frac{1}{2}}}_{= -\frac{N}{2}\log|\Sigma|} - \frac{1}{2}\sum_{n=1}^{N} (x^{(n)}-\mu)^T \Sigma (x^{(n)}-\mu) \}$$ (A.3)

Q μ の最尤推定

$$\frac{\partial}{\partial \mu}\left((x-\mu)^T \Sigma^{-1} (x-\mu) \right) = -2\Sigma^{-1}(x-\mu) \qquad \text{(A.4)}$$

仮定

より、

$$\frac{\partial L}{\partial \mu} = \frac{\partial}{\partial \mu}\left(-\frac{1}{2}\sum_{n=1}^{N} (x^{(n)}-\mu)^T \Sigma (x^{(n)}-\mu) \right)$$

$$= -\frac{1}{2}\cdot(-2)\sum_{n=1}^{N} \Sigma^{-1}(x^{(n)}-\mu)$$

$$= \Sigma^{-1}\sum_{n=1}^{N}(x^{(n)}-\mu) \qquad \text{(A.5)}$$

$$= 0$$

$$\Leftrightarrow \Sigma\Sigma^{-1}\sum_{n=1}^{N}(x^{(n)}-\mu) = \Sigma 0$$

$$\Leftrightarrow \sum_{n=1}^{N}(x^{(n)}-\mu) = 0$$

$$\Leftrightarrow \sum_{n=1}^{N}\mu = \sum_{n=1}^{N}x^{(n)}$$

$$\Leftrightarrow N\mu = \sum_{n=1}^{N}x^{(n)}$$

$$\Leftrightarrow \boxed{\hat{\mu} = \frac{1}{N}\sum_{n=1}^{N}x^{(n)}}$$

$N=1$ のとき（1つのデータについて）、

$$\text{(A.5)} \Rightarrow \boxed{\frac{\partial}{\partial \mu}\log \mathcal{N}(x;\mu,\Sigma) = \Sigma^{-1}(x-\mu)} \qquad \text{(A.6)}$$

Q Σ の最尤推定

$$\text{(A.3)} \Rightarrow \frac{\partial L}{\partial \Sigma} = -\frac{N}{2}\frac{\partial}{\partial \Sigma}\log|\Sigma| - \frac{1}{2}\frac{\partial}{\partial \Sigma}\sum_{n=1}^{N}(x^{(n)}-\mu)^T\Sigma^{-1}(x^{(n)}-\mu)$$

ここで、

$$\frac{\partial}{\partial \Sigma}\log|\Sigma| = (\Sigma^{-1})^T \qquad \text{(A.11)}$$

を利用すると、

$$\frac{\partial}{\partial \Sigma}\sum_{n=1}^{N}(x^{(n)}-\mu)^T\Sigma^{-1}(x^{(n)}-\mu)$$

$$= -\left(\Sigma^{-1}\underbrace{\sum_{n=1}^{N}(x^{(n)}-\mu)(x^{(n)}-\mu)^T}_{S}\Sigma^{-1} \right)^T$$

$$= -(\Sigma^{-1}S\Sigma^{-1})^T \qquad \text{(A.12)}$$

※ (A.12) の証明

○トレース

$$\mathrm{Tr}(A) = a_{11} + a_{22} + \cdots + a_{DD} = \sum_{i=1}^{D} a_{ii}$$

また、

$$\mathrm{Tr}(A+B) = \mathrm{Tr}(A) + \mathrm{Tr}(B)$$

$$\mathrm{Tr}(AB) = \mathrm{Tr}(BA) \quad \leftarrow \quad AB = B^{-1}A^{-1}$$

$$\frac{\partial}{\partial a_{ij}} \mathrm{Tr}(AB) = b_{ji} \quad \leftarrow \quad AB = \begin{bmatrix} a_{11} & a_{12} & \\ a_{21} & & \\ & \cdots & a_{DD} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \\ b_{21} & & \\ & \cdots & b_{DD} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}b_{11} + a_{12}b_{21} \cdots + a_{1D}b_{D1} & \cdots \\ \vdots & \cdots \\ & \end{bmatrix}$$

また、

$$\underbrace{\frac{\partial}{\partial x}(\overset{I}{\overset{\parallel}{A^{-1}A}})}_{=0} = \frac{\partial A^{-1}}{\partial x} A + A^{-1} \frac{\partial}{\partial x} A$$

$$\Longleftrightarrow \quad \frac{\partial A^{-1}}{\partial x} A + A^{-1} \frac{\partial A}{\partial x} = 0$$

$$\Longleftrightarrow \quad \frac{\partial A^{-1}}{\partial x} \underbrace{A A^{-1}}_{=I} + A^{-1} \frac{\partial A}{\partial x} A^{-1} = 0$$

$$\Longleftrightarrow \quad \boxed{\frac{\partial A^{-1}}{\partial x} = -A^{-1} \frac{\partial A}{\partial x} A^{-1}} \quad (A.16)$$

以上を用いて、

2次形式 → スカラ

トレースをとれる。

$$\sum_{n=1}^{N} \underline{(x^{(n)} - \mu)^T \Sigma^{-1} (x^{(n)} - \mu)}$$

$$= \sum_{n=1}^{N} \mathrm{Tr}\left((x^{(n)} - \mu)^T \Sigma^{-1} (x^{(n)} - \mu) \right)$$

$$= \sum_{n=1}^{N} \mathrm{Tr}\left(\Sigma^{-1} (x^{(n)} - \mu)(x^{(n)} - \mu)^T \right) \quad (\because \mathrm{Tr}(AB) = \mathrm{Tr}(BA))$$

$$= \mathrm{Tr}\left[\sum_{n=1}^{N} \Sigma^{-1} (x^{(n)} - \mu)(x^{(n)} - \mu)^T \right] \quad (\because \mathrm{Tr}(A+B) = \mathrm{Tr}(A) + \mathrm{Tr}(B))$$

$$= \mathrm{Tr}\left[\Sigma^{-1} \sum_{n=1}^{N} (x^{(n)} - \mu)(x^{(n)} - \mu)^T \right]$$

$$= \mathrm{Tr}[\Sigma^{-1} S]$$

いま、

$$\frac{\partial}{\partial \sigma_{ij}} \sum_{n=1}^{N} (x^{(n)} - \mu) \Sigma^{-1} (x^{(n)} - \mu)$$

$$= \frac{\partial}{\partial \sigma_{ij}} \mathrm{Tr}[\Sigma^{-1} S]$$

$$= \mathrm{Tr}\left[\left(\frac{\partial}{\partial \sigma_{ij}} \Sigma^{-1} \right) S \right]$$

$$= \mathrm{Tr}\left[-\Sigma^{-1} \left(\frac{\partial}{\partial \sigma_{ij}} \Sigma \right) \Sigma^{-1} S \right]$$

$$= -\mathrm{Tr}\left[\left(\frac{\partial}{\partial \sigma_{ij}} \Sigma \right) \underbrace{\Sigma^{-1} S \Sigma^{-1}} \right]$$

$$= -\mathrm{Tr}\left[\left(\frac{\partial}{\partial \sigma_{ij}} \Sigma \right) \underline{C} \right]$$

よって

$$\boxed{\Sigma_k = \frac{\sum_{n=1}^{N} q^{(n)}(k)\,(x^{(n)}-\mu_k)(x^{(n)}-\mu_k)^T}{\sum_{n=1}^{N} q^{(n)}(k)}} \quad (5.12)$$

※(A.11)の証明

$$\frac{\partial}{\partial \Sigma}\ln|\Sigma| = (\Sigma^{-1})^T \quad \text{を示す.}$$

1. 余因子展開

$$|\Sigma| = \sum_{k=1}^{D} \Sigma_{ik} C_{ik} \Rightarrow \frac{\partial|\Sigma|}{\partial \Sigma_{ij}} = C_{ij} \quad (k \neq j \text{ などでは微分が} 0)$$

よって

$$\frac{\partial|\Sigma|}{\partial \Sigma} = C \ (\text{余因子行列})$$

2. 連鎖律

$$\frac{\partial}{\partial \Sigma_{ij}}\ln|\Sigma| = \frac{\partial \ln|\Sigma|}{\partial|\Sigma|}\frac{\partial|\Sigma|}{\partial \Sigma_{ij}} = \frac{1}{|\Sigma|}\frac{\partial|\Sigma|}{\partial \Sigma_{ij}} = \frac{1}{|\Sigma|}C_{ij}$$

よって

$$\frac{\partial}{\partial \Sigma}\ln|\Sigma| = \frac{1}{|\Sigma|}C$$

3. 逆行列の公式

$$\Sigma^{-1} = \frac{1}{|\Sigma|}C^T \iff |\Sigma|\Sigma^{-1} = C^T$$
$$\iff |\Sigma|(\Sigma^{-1})^T = C$$

よって

$$\frac{\partial}{\partial \Sigma}\ln|\Sigma| = (\Sigma^{-1})^T \quad \square$$