宿題 1

入力 $x \in \mathbb{R}^d$，期待値 $\mu \in \mathbb{R}^d$，共分散行列 $\boldsymbol{\Sigma} \in \mathbb{R}^{d\times d}$ のガウスモデルは次のように表される。

$$q(\boldsymbol{x};\, \boldsymbol{\mu},\, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{d/2}\det(\boldsymbol{\Sigma})^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right) \tag{1}$$

標本 $\{x_i\}_{i=1}^n$ に対して，このモデルの最尤推定量 $\hat{\boldsymbol{\mu}}_{\mathrm{ML}}$, $\hat{\boldsymbol{\Sigma}}_{\mathrm{ML}}$ を求める。対数尤度は次のようになる。

$$\begin{aligned}
\log L(\boldsymbol{\mu},\, \boldsymbol{\Sigma}) &= \sum_{i=1}^n q(\boldsymbol{x}_i;\, \boldsymbol{\mu},\, \boldsymbol{\Sigma}) && (2)\\
&= \sum_{i=1}^n \left(-\frac{d}{2}\log(2\pi) - \frac{1}{2}\log(\det(\boldsymbol{\Sigma})) - \frac{1}{2}\sum_{i=1}^n (\boldsymbol{x}_i-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu})\right) && (3)\\
&= -\frac{nd}{2}\log(2\pi) - \frac{n}{2}\log(\det(\boldsymbol{\Sigma})) - \frac{1}{2}\sum_{i=1}^n (\boldsymbol{x}_i-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_i-\boldsymbol{\mu}) && (4)\\
& && (5)
\end{aligned}$$

ここで，各要素の分散が等しく，共分散が 0 となる，すなわち $\boldsymbol{\Sigma} = \sigma^2\boldsymbol{I}$ であるガウスモデルを考えると，

$$q(\boldsymbol{x};\, \boldsymbol{\mu},\, \boldsymbol{\Sigma}) = \frac{1}{(2\pi\sigma^2)^{d/2}} \exp\left(-\frac{(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{\mu})}{2\sigma^2}\right) \tag{6}$$

となる。このとき，$\det(\boldsymbol{\Sigma}) = \sigma^{2d}$，$\Sigma^{-1} = (1/\sigma^2)\boldsymbol{I}$ から，対数尤度は，

$$\begin{aligned}
\log L(\boldsymbol{\mu},\, \sigma) &= -\frac{nd}{2}\log(2\pi) - nd\log(\sigma) - \frac{1}{2\sigma^2}\sum_{i=1}^n (\boldsymbol{x}_i-\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{x}_i-\boldsymbol{\mu}) && (7)\\
& && (8)
\end{aligned}$$

$\boldsymbol{\mu}, \sigma$ でそれぞれ偏微分して，

$$\frac{\partial}{\partial\mu}(\log L(\boldsymbol{\mu},\, \sigma)) = \frac{1}{\sigma^2}\left(-n\boldsymbol{\mu} + \sum_{i=1}^n \boldsymbol{x}_i\right) \tag{9}$$

$$\frac{\partial}{\partial\sigma}(\log L(\boldsymbol{\mu},\, \sigma)) = -\frac{nd}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^n (\boldsymbol{x}_i-\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{x}_i-\boldsymbol{\mu}) \tag{10}$$

これらをそれぞれ $\boldsymbol{0}$, 0 と置くことで，L に最大値を与える $\boldsymbol{\mu}$, σ が得られる。つまり，

$$\hat{\boldsymbol{\mu}}_{\mathrm{ML}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{x}_i \tag{11}$$

$$\begin{aligned}
\hat{\sigma}^2_{\mathrm{ML}} &= \frac{1}{nd}\sum_{i=1}^n (\boldsymbol{x}_i-\boldsymbol{\mu})^{\mathrm{T}}(\boldsymbol{x}_i-\boldsymbol{\mu}) && (12)\\
&= \frac{1}{nd}\sum_{i=1}^n\sum_{j=1}^d \left(x_i^{(j)} - \hat{\mu}_{\mathrm{ML}}^{(j)}\right)^2 && (13)
\end{aligned}$$