提出課題 VI：連続潜在変数

問題文中の空欄に入る数式を選択せよ．選択した解答（(A)–(D)）は「機械学習 2024 KA240201-teams」の「一般」チャネルに出現する課題へのリンクから提出すること．

問題 1，2

$\mathbf{y}$ のガウス周辺分布と，$\mathbf{y}$ が与えられた下での $\mathbf{v}$ の条件付きガウス分布が

$$p(\mathbf{y}) = \mathcal{N}\left(\mathbf{y}|\boldsymbol{\eta}, \boldsymbol{\Lambda}^{-1}\right) \tag{a}$$

$$p(\mathbf{v}|\mathbf{y}) = \mathcal{N}\left(\mathbf{v}|\mathbf{A}\mathbf{y} + \mathbf{b}, \mathbf{L}^{-1}\right) \tag{b}$$

と与えられたとき，$\mathbf{v}$ が与えられた下での $\mathbf{y}$ の条件付きガウス分布は

$$p(\mathbf{y}|\mathbf{v}) = \mathcal{N}\left(\mathbf{y}|\boxed{\ (1)\ }, \boldsymbol{\Sigma}\right) \tag{c}$$

となる．ただし，$\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} + \mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A})^{-1}$ である．この一般的な結果を用いて，確率的主成分分析モデルの事後分布 $p(\mathbf{z}|\mathbf{x})$ が

$$p(\mathbf{z}|\mathbf{x}) = \mathcal{N}\left(\mathbf{z}|\mathbf{M}^{-1}\mathbf{W}^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu}), \sigma^2\mathbf{M}^{-1}\right)$$

となることを示す（ただし，$\mathbf{M} = \mathbf{W}^{\mathrm{T}}\mathbf{W} + \sigma^2\mathbf{I}$）．

$p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$ と式 (a)，$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \sigma^2\mathbf{I})$ と式 (b) を対応付けることで，式 (c) から

$$\begin{aligned} p(\mathbf{z}|\mathbf{x}) &= \mathcal{N}\left(\mathbf{z}|\boxed{\ (2)\ }, (\mathbf{I} + \sigma^{-2}\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\right) \\ &= \mathcal{N}\left(\mathbf{z}|\mathbf{M}^{-1}\mathbf{W}^{\mathrm{T}}(\mathbf{x} - \boldsymbol{\mu}), \sigma^2\mathbf{M}^{-1}\right) \end{aligned}$$

が得られる．

問題 1. 空欄 (1) に入る数式を選択せよ．

(A) $\boldsymbol{\Sigma}\{\mathbf{A}^{\mathrm{T}}\mathbf{L}(\mathbf{v} - \mathbf{b})\}$

(B) $\boldsymbol{\Sigma}\{\mathbf{A}^{\mathrm{T}}\mathbf{L}^{-1}(\mathbf{v} - \mathbf{b})\}$

(C) $\boldsymbol{\Sigma}\{\mathbf{A}^{\mathrm{T}}\mathbf{L}(\mathbf{v} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\eta}\}$

(D) $\boldsymbol{\Sigma}\{\mathbf{A}^{\mathrm{T}}\mathbf{L}^{-1}(\mathbf{v} - \mathbf{b}) + \boldsymbol{\Lambda}\boldsymbol{\eta}\}$

問題 2. 空欄 (2) に入る数式を選択せよ．

(A) $(\sigma^2\mathbf{I} + \mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\sigma^{-2}\mathbf{I}(\mathbf{x} - \boldsymbol{\mu})$

(B) $(\sigma^2\mathbf{I} + \mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\sigma^{2}\mathbf{I}(\mathbf{x} - \boldsymbol{\mu})$

(C) $(\mathbf{I} + \sigma^{-2}\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\sigma^{-2}\mathbf{I}(\mathbf{x} - \boldsymbol{\mu})$

(D) $(\mathbf{I} + \sigma^{-2}\mathbf{W}^{\mathrm{T}}\mathbf{W})^{-1}\mathbf{W}^{\mathrm{T}}\sigma^{2}\mathbf{I}(\mathbf{x} - \boldsymbol{\mu})$

問題 3，4

確率的主成分分析モデルの M ステップの式

$$\mathbf{W}_{\mathrm{new}} = \left[\sum_{n=1}^{N}(\mathbf{x}_n - \bar{\mathbf{x}})\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}}\right]\left[\sum_{n=1}^{N}\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\right]^{-1} \tag{d}$$

$$\begin{aligned}\sigma^2_{\mathrm{new}} =& \frac{1}{ND}\sum_{n=1}^{N}\{||\mathbf{x}_n - \bar{\mathbf{x}}||^2 - 2\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}}\mathbf{W}_{\mathrm{new}}^{\mathrm{T}}(\mathbf{x}_n - \bar{\mathbf{x}}) \\ &+ \mathrm{Tr}\left(\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\mathbf{W}_{\mathrm{new}}^{\mathrm{T}}\mathbf{W}_{\mathrm{new}}\right)\}\end{aligned} \tag{e}$$

を，以下の期待完全データ対数尤度関数

$$\begin{aligned}\mathbb{E}[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \mathbf{W}, \sigma^2)] = -\sum_{n=1}^{N}\Bigg\{&\frac{D}{2}\ln(2\pi\sigma^2) + \frac{1}{2}\mathrm{Tr}\left(\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\right) \\ &+ \frac{1}{2\sigma^2}||\mathbf{x}_n - \boldsymbol{\mu}||^2 - \frac{1}{\sigma^2}\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}(\mathbf{x}_n - \boldsymbol{\mu}) \\ &+ \frac{1}{2\sigma^2}\mathrm{Tr}\left(\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\mathbf{W}^{\mathrm{T}}\mathbf{W}\right) + \frac{M}{2}\ln(2\pi)\Bigg\}\end{aligned} \tag{f}$$

を最大化することによって導出する．

標準的な導関数と行列微分のルールを用いると，式 (f) の導関数は以下のように計算できる．

$$\frac{\partial}{\partial \mathbf{W}}\mathbb{E}[\ln p\left(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \mathbf{W}, \sigma^2\right)] = \sum_{n=1}^{N}\boxed{\quad (3) \quad}$$

$$\frac{\partial}{\partial \sigma^2}\mathbb{E}[\ln p\left(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \mathbf{W}, \sigma^2\right)] = \sum_{n=1}^{N}\boxed{\quad (4) \quad}.$$

これらをゼロとして整理し，$\boldsymbol{\mu}$ に $\bar{\mathbf{x}}$，$\mathbf{W}$ に $\mathbf{W}_{\mathrm{new}}$，$\sigma^2$ に σ^2_{new} を代入すると，式 (d) と (e) がそれぞれ得られる．

問題 3. 空欄 (3) に入る数式を選択せよ．

(A) $((\mathbf{x}_n - \boldsymbol{\mu})\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}} - \mathbf{W}\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}])/\sigma^2$

(B) $((\mathbf{x}_n - \boldsymbol{\mu})\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}} - \mathbf{W}\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}])/(2\sigma^2)$

(C) $((\mathbf{x}_n - \boldsymbol{\mu})\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}} - \mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\mathbf{W})/\sigma^2$

(D) $((\mathbf{x}_n - \boldsymbol{\mu})\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}} - \mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\mathbf{W})/(2\sigma^2)$

問題 4. 空欄 (4) に入る数式を選択せよ．

(A) $(\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\mathbf{W}^{\mathrm{T}}\mathbf{W} + ||\mathbf{x}_n - \boldsymbol{\mu}||^2 - 2\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}(\mathbf{x}_n - \boldsymbol{\mu}) - D\sigma^2)/(2\sigma^4)$

(B) $(\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\mathbf{W}^{\mathrm{T}}\mathbf{W} + ||\mathbf{x}_n - \boldsymbol{\mu}||^2 - 2\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}(\mathbf{x}_n - \boldsymbol{\mu}) - ND\sigma^2)/(2\sigma^4)$

(C) $(\mathrm{Tr}\,(\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\mathbf{W}^{\mathrm{T}}\mathbf{W}) + ||\mathbf{x}_n - \boldsymbol{\mu}||^2 - 2\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}(\mathbf{x}_n - \boldsymbol{\mu}) - D\sigma^2)/(2\sigma^4)$

(D) $(\mathrm{Tr}\,(\mathbb{E}[\mathbf{z}_n\mathbf{z}_n^{\mathrm{T}}]\mathbf{W}^{\mathrm{T}}\mathbf{W}) + ||\mathbf{x}_n - \boldsymbol{\mu}||^2 - 2\mathbb{E}[\mathbf{z}_n]^{\mathrm{T}}\mathbf{W}^{\mathrm{T}}(\mathbf{x}_n - \boldsymbol{\mu}) - ND\sigma^2)/(2\sigma^4)$