Assignment VI: Continuous Latent Variables

Select the equations that fit in the blanks in the question text. The selected answers ((A)–(D)) must be submitted via the link to the assignment that appears in the “General” channel of “機械学習 2024 KA240201-teams.”

Questions 1 and 2

Given a marginal Gaussian distribution for $\mathbf{y}$ and a conditional Gaussian distribution for $\mathbf{v}$ given $\mathbf{y}$ in the form

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

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

the conditional distribution of $\mathbf{y}$ given $\mathbf{v}$ is given by

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

where $\boldsymbol{\Sigma} = (\boldsymbol{\Lambda} + \mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A})^{-1}$. By making use of this general result, we show that the posterior distribution $p(\mathbf{z}|\mathbf{x})$ for the probabilistic PCA model is given by

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

where $\mathbf{M} = \mathbf{W}^{\mathrm{T}}\mathbf{W} + \sigma^2\mathbf{I}$.

By matching $p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$ with Eq. (a) and $p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mathbf{W}\mathbf{z} + \boldsymbol{\mu}, \sigma^2\mathbf{I})$ with Eq. (b), we have from Eq. (c) that

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

Question 1. Select the equation that fills in the blank (1).

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

Question 2. Select the equation that fills in the blank (2).

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

Questions 3 and 4

We derive the M-step equations for the probabilistic PCA model

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

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

by maximizing the expected complete-data log likelihood function given by

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

Using standard derivatives together with the rules for matrix differentiation, we can compute the derivatives of Eq. (f) with respect to $\mathbf{W}$ and σ^2 as follows:

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

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

Setting these equal to zero and re-arranging, and substituting $\bar{\mathbf{x}}$ for $\boldsymbol{\mu}$, $\mathbf{W}_{\text{new}}$ for $\mathbf{W}$, and σ^2_{new} for σ^2, we obtain Eqs. (d) and (e), respectively.

Question 3. Select the equation that fills in the blank (3).

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

Question 4. Select the equation that fills in the blank (4).

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