Assignment V: Mixture Models and EM

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

We show that if we maximize

$$\mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})] = \sum_{n=1}^{N} \sum_{k=1}^{K} \gamma(z_{nk}) \{\ln \pi_k + \ln \mathcal{N}(\mathbf{x}_n|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)\} \tag{a}$$

with respect to $\boldsymbol{\mu}_k$ and $\boldsymbol{\Sigma}_k$ while keeping the responsibilities $\gamma(z_{nk})$ fixed, we obtain the closed form solutions given by

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \mathbf{x}_n \tag{b}$$

$$\boldsymbol{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}}. \tag{c}$$

If we differentiate Eq. (a) with respect to $\boldsymbol{\mu}_k$, while keeping the $\gamma(z_{nk})$ fixed, we get

$$\frac{\partial}{\partial \boldsymbol{\mu}_k} \mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})] = \boxed{\ (1)\ }.$$

Setting this equal to zero and rearranging, we obtain Eq. (b). Similarly, if we differentiate Eq. (a) with respect to $\boldsymbol{\Sigma}_k^{-1}$, while keeping the $\gamma(z_{nk})$ fixed, we get

$$\frac{\partial}{\partial \boldsymbol{\Sigma}_k^{-1}} \mathbb{E}_{\mathbf{Z}}[\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\pi})] = \boxed{\ (2)\ }.$$

Setting this equal to zero and rearranging, we obtain Eq. (c).

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

(A) $-\sum_{n=1}^{N} \gamma(z_{nk})(\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k - \boldsymbol{\Sigma}_k^{-1}\mathbf{x}_n)$

(B) $\sum_{n=1}^{N} \gamma(z_{nk})(\boldsymbol{\Sigma}_k^{-1}\boldsymbol{\mu}_k - \boldsymbol{\Sigma}_k^{-1}\mathbf{x}_n)$

(C) $-\sum_{n=1}^{N} \gamma(z_{nk})(\boldsymbol{\mu}_k\boldsymbol{\Sigma}_k^{-1} - \mathbf{x}_n\boldsymbol{\Sigma}_k^{-1})$

(D) $\sum_{n=1}^{N} \gamma(z_{nk})(\boldsymbol{\mu}_k\boldsymbol{\Sigma}_k^{-1} - \mathbf{x}_n\boldsymbol{\Sigma}_k^{-1})$

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

(A) $-\sum_{n=1}^{N} \gamma(z_{nk}) \left(\boldsymbol{\Sigma}_k - (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}} \right) / 2$

(B) $\sum_{n=1}^{N} \gamma(z_{nk}) \left(\boldsymbol{\Sigma}_k - (\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}} \right) / 2$

(C) $-\sum_{n=1}^{N} \gamma(z_{nk}) \left(\boldsymbol{\Sigma}_k^{-1} - ((\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}})^{-1} \right) / 2$

(D) $\sum_{n=1}^{N} \gamma(z_{nk}) \left(\boldsymbol{\Sigma}_k^{-1} - ((\mathbf{x}_n - \boldsymbol{\mu}_k)(\mathbf{x}_n - \boldsymbol{\mu}_k)^{\mathrm{T}})^{-1} \right) / 2$

Questions 3 and 4

Consider the incremental form of the EM algorithm for a mixture of Gaussians, in which responsibilities are recomputed only for a specific data point $\mathbf{x}_m$, and the corresponding old and new values of the responsibilities are denoted by $\gamma^{\mathrm{old}}(z_{mk})$ and $\gamma^{\mathrm{new}}(z_{mk})$. Here we assume

$$
\begin{aligned}
N_k^{\mathrm{old}} &= \sum_n \gamma^{\mathrm{old}}(z_{nk}) \\
\boldsymbol{\mu}_k^{\mathrm{old}} &= \frac{1}{N_k^{\mathrm{old}}} \sum_n \gamma^{\mathrm{old}}(z_{nk}) \mathbf{x}_n \\
\boldsymbol{\Sigma}_k^{\mathrm{old}} &= \frac{1}{N_k^{\mathrm{old}}} \sum_n \gamma^{\mathrm{old}}(z_{nk}) (\mathbf{x}_n - \boldsymbol{\mu}_k^{\mathrm{old}})(\mathbf{x}_n - \boldsymbol{\mu}_k^{\mathrm{old}})^{\mathrm{T}} \\
\pi_k^{\mathrm{old}} &= \frac{N_k^{\mathrm{old}}}{N} = \frac{1}{N} \sum_n \gamma^{\mathrm{old}}(z_{nk}) \\
N_k^{\mathrm{new}} &= N_k^{\mathrm{old}} + \gamma^{\mathrm{new}}(z_{mk}) - \gamma^{\mathrm{old}}(z_{mk}) \\
\boldsymbol{\mu}_k^{\mathrm{new}} &= \boldsymbol{\mu}_k^{\mathrm{old}} + \left(\frac{\gamma^{\mathrm{new}}(z_{mk}) - \gamma^{\mathrm{old}}(z_{mk})}{N_k^{\mathrm{new}}} \right) (\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{old}}).
\end{aligned}
$$

Using the above definitions, M-step formulae for updating the covariance matrices $\boldsymbol{\Sigma}_k$ and mixing coefficients π_k in a Gaussian mixture model when the responsibilities are updated incrementally are $\boxed{\ (3)\ }$ and $\boxed{\ (4)\ }$, respectively.

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

(A)
$$
\begin{aligned}
\boldsymbol{\Sigma}_k^{\mathrm{new}} = \boldsymbol{\Sigma}_k^{\mathrm{old}} &+ \gamma^{\mathrm{new}}(z_{mk}) \left((\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{new}})(\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{new}})^{\mathrm{T}} - \boldsymbol{\Sigma}_k^{\mathrm{old}} \right) / N_k^{\mathrm{new}} \\
&- \gamma^{\mathrm{old}}(z_{mk}) \left((\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{old}})(\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{old}})^{\mathrm{T}} - \boldsymbol{\Sigma}_k^{\mathrm{old}} \right) / N_k^{\mathrm{old}}
\end{aligned}
$$

(B)
$$
\begin{aligned}
\boldsymbol{\Sigma}_k^{\mathrm{new}} = \boldsymbol{\Sigma}_k^{\mathrm{old}} &+ \gamma^{\mathrm{new}}(z_{mk}) \left((\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{new}})(\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{new}})^{\mathrm{T}} - \boldsymbol{\Sigma}_k^{\mathrm{old}} \right) / N_k^{\mathrm{old}} \\
&- \gamma^{\mathrm{old}}(z_{mk}) \left((\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{old}})(\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{old}})^{\mathrm{T}} - \boldsymbol{\Sigma}_k^{\mathrm{old}} \right) / N_k^{\mathrm{old}}
\end{aligned}
$$

(C)
$$
\begin{aligned}
\boldsymbol{\Sigma}_k^{\mathrm{new}} = \boldsymbol{\Sigma}_k^{\mathrm{old}} &+ \gamma^{\mathrm{new}}(z_{mk}) \left((\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{new}})(\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{new}})^{\mathrm{T}} - \boldsymbol{\Sigma}_k^{\mathrm{old}} \right) / N_k^{\mathrm{new}} \\
&- \gamma^{\mathrm{old}}(z_{mk}) \left((\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{old}})(\mathbf{x}_m - \boldsymbol{\mu}_k^{\mathrm{old}})^{\mathrm{T}} - \boldsymbol{\Sigma}_k^{\mathrm{old}} \right) / N_k^{\mathrm{new}}
\end{aligned}
$$

(D) $\boldsymbol{\Sigma}_k^{\text{new}} = \boldsymbol{\Sigma}_k^{\text{old}} + \gamma^{\text{new}}(z_{mk}) \left((\mathbf{x}_m - \boldsymbol{\mu}_k^{\text{new}})(\mathbf{x}_m - \boldsymbol{\mu}_k^{\text{new}})^{\mathrm{T}} - \boldsymbol{\Sigma}_k^{\text{old}} \right) / N$
$\quad - \gamma^{\text{old}}(z_{mk}) \left((\mathbf{x}_m - \boldsymbol{\mu}_k^{\text{old}})(\mathbf{x}_m - \boldsymbol{\mu}_k^{\text{old}})^{\mathrm{T}} - \boldsymbol{\Sigma}_k^{\text{old}} \right) / N$

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

(A) $\pi_k^{\text{new}} = \pi_k^{\text{old}} + \gamma^{\text{new}}(z_{mk})/N_k^{\text{new}} - \gamma^{\text{old}}(z_{mk})/N_k^{\text{old}}$

(B) $\pi_k^{\text{new}} = \pi_k^{\text{old}} + \gamma^{\text{new}}(z_{mk})/N_k^{\text{old}} - \gamma^{\text{old}}(z_{mk})/N_k^{\text{old}}$

(C) $\pi_k^{\text{new}} = \pi_k^{\text{old}} + \gamma^{\text{new}}(z_{mk})/N_k^{\text{new}} - \gamma^{\text{old}}(z_{mk})/N_k^{\text{new}}$

(D) $\pi_k^{\text{new}} = \pi_k^{\text{old}} + \gamma^{\text{new}}(z_{mk})/N - \gamma^{\text{old}}(z_{mk})/N$