Cor Zuurmond
10580395
jczuurmond@gmail.com

Homework Assignment 7

Machine Learning 2, 17/18

2018-06-01

$$
\begin{aligned}
D_{KL}(q(\mathbf{z}) \parallel p(\mathbf{z})) &= \int q(\mathbf{z}) \log \left(\frac{p(\mathbf{z})}{q(\mathbf{z})} \right) d\mathbf{z} \\
&= \int q(\mathbf{z}) \log(p(\mathbf{z})) d\mathbf{z} - \int q(\mathbf{z}) \log(q(\mathbf{z})) d\mathbf{z} \\
\int q(\mathbf{z}) \log(p(\mathbf{z})) d\mathbf{z} &= \int q(\mathbf{z}) \log \left(\frac{1}{\sqrt{(2\pi)^K \mid \mathbf{I} \mid}} e^{-\frac{1}{2}\mathbf{z}^T \mathbf{I}^{-1} \mathbf{z})} \right) d\mathbf{z} \\
&= -\log \left(\sqrt{(2\pi)^K \mid \mathbf{I} \mid} \right) + \int q(\mathbf{z}) \left(-\frac{1}{2} \mathbf{z}^T \mathbf{I}^{-1} \boldsymbol{\mu} \right) d\mathbf{z} \\
&= -\frac{K}{2} \log(2\pi) - \frac{1}{2} \int \mathbf{z}^T \mathbf{z} q(\mathbf{z}) d\mathbf{z} \\
&= -\frac{K}{2} \log(2\pi) - \frac{1}{2} \int (\mathbf{z} + \boldsymbol{\mu})^T (\mathbf{z} + \boldsymbol{\mu}) q(\mathbf{z} + \boldsymbol{\mu}) d\mathbf{z} \\
&= -\frac{K}{2} \log(2\pi) - \frac{1}{2} \int (\mathbf{z}^T \mathbf{z} + \mathbf{z}^T \boldsymbol{\mu} + \boldsymbol{\mu}^T \mathbf{z} + \boldsymbol{\mu}^T \boldsymbol{\mu}) q(\mathbf{z} + \boldsymbol{\mu}) d\mathbf{z} \\
&= -\frac{K}{2} \log(2\pi) - \frac{1}{2} \left(\mathbb{E}_{q(\mathbf{z}+\boldsymbol{\mu})} \left[\mathbf{z}^T \mathbf{z} \right] + \mathbb{E}_{q(\mathbf{z}+\boldsymbol{\mu})} \left[\mathbf{z}^T \boldsymbol{\mu} \right] + \mathbb{E}_{q(\mathbf{z}+\boldsymbol{\mu})} \left[\boldsymbol{\mu}^T \mathbf{z} \right] + \mathbb{E}_{q(\mathbf{z}+\boldsymbol{\mu})} \left[\boldsymbol{\mu}^T \boldsymbol{\mu} \right] \right) \\
&= -\frac{K}{2} \log(2\pi) - \frac{1}{2} \left(\text{trace} \left(\text{diag} \left(\boldsymbol{\sigma} \right) \right) + \text{trace} \left(\boldsymbol{\mu}^T \boldsymbol{\mu} \right) \right) \\
&= -\frac{K}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{K} \left(\boldsymbol{\sigma}_i^2 + \boldsymbol{\mu}_i^2 \right) \\
\int q(\mathbf{z}) \log(q(\mathbf{z})) d\mathbf{z} &= \int q(\mathbf{z}) \log \left(\frac{1}{\sqrt{(2\pi)^K \mid \text{diag} \left(\boldsymbol{\sigma} \right) \mid}} e^{-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})^T (\text{diag}(\boldsymbol{\sigma}) \cdot \text{diag}(\boldsymbol{\sigma}))^{-1} (\mathbf{z}-\boldsymbol{\mu})} \right) d\mathbf{z} \\
&= -\frac{K}{2} \log(2\pi) - \frac{K}{2} \log \left(\mid \text{diag} \left(\boldsymbol{\sigma} \right) \cdot \text{diag} \left(\boldsymbol{\sigma} \right) \mid \right) - \frac{1}{2} \int (\mathbf{z} - \boldsymbol{\mu})^T (\text{diag} \left(\boldsymbol{\sigma} \right) \cdot \text{diag} \left(\boldsymbol{\sigma} \right))^{-1} (\mathbf{z} - \boldsymbol{\mu}) q(\mathbf{z}) d\mathbf{z} \\
&= -\frac{K}{2} \log(2\pi) - \frac{K}{2} \sum_{i=1}^{K} (\log(\boldsymbol{\sigma}_i^2)) - \frac{1}{2} \int \mathbf{z}^T (\text{diag} \left(\boldsymbol{\sigma} \right) \cdot \text{diag} \left(\boldsymbol{\sigma} \right))^{-1} \mathbf{z} q(\mathbf{z} + \boldsymbol{\mu}) d\mathbf{z} \\
&= -\frac{K}{2} \log(2\pi) - \frac{K}{2} \sum_{i=1}^{K} (\log(\boldsymbol{\sigma}_i^2)) - \frac{1}{2} \mathbb{E}_{q(\mathbf{z}+\boldsymbol{\mu})} \left[\mathbf{z}^T (\text{diag} \left(\boldsymbol{\sigma} \right) \cdot \text{diag} \left(\boldsymbol{\sigma} \right))^{-1} \mathbf{z} \right] \\
&= -\frac{K}{2} \log(2\pi) - \frac{K}{2} \sum_{i=1}^{K} (\log(\boldsymbol{\sigma}_i^2)) - \frac{1}{2} \text{trace} \left(\mathbb{E}_{q(\mathbf{z}+\boldsymbol{\mu})} \left[\mathbf{z}^T (\text{diag} \left(\boldsymbol{\sigma} \right) \cdot \text{diag} \left(\boldsymbol{\sigma} \right))^{-1} \mathbf{z} \right] \right) \\
&= -\frac{K}{2} \log(2\pi) - \frac{K}{2} \sum_{i=1}^{K} (\log(\boldsymbol{\sigma}_i^2)) - \frac{1}{2} \text{trace} \left((\text{diag} \left(\boldsymbol{\sigma} \right) \cdot \text{diag} \left(\boldsymbol{\sigma} \right))^{-1} (\text{diag} \left(\boldsymbol{\sigma} \right) \cdot \text{diag} \left(\boldsymbol{\sigma} \right)) \right) \\
&= -\frac{K}{2} \log(2\pi) - \frac{K}{2} \sum_{i=1}^{K} (\log(\boldsymbol{\sigma}_i^2)) - \frac{K}{2} \\
&= -\frac{K}{2} \log(2\pi) - \frac{K}{2} \sum_{i=1}^{K} (\log(\boldsymbol{\sigma}_i^2) + 1) \\
D_{KL}(q(\mathbf{z}) \parallel p(\mathbf{z})) &= \int q(\mathbf{z}) \log(p(\mathbf{z})) d\mathbf{z} - \int q(\mathbf{z}) \log(q(\mathbf{z})) d\mathbf{z} \\
&= -\frac{K}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^{K} \left(\boldsymbol{\sigma}_i^2 + \boldsymbol{\mu}_i^2 \right) - \left(-\frac{K}{2} \log(2\pi) - \frac{K}{2} \sum_{i=1}^{K} (\log(\boldsymbol{\sigma}_i^2) + 1) \right) \\
&= \frac{1}{2} \sum_{i=1}^{K} \left(\log(\boldsymbol{\sigma}_i^2) + 1 - \boldsymbol{\sigma}_i^2 - \boldsymbol{\mu}_i^2 \right)
\end{aligned}
$$