Cor Zuurmond
10580395
jczuurmond@gmail.com

Homework Assignment 7
Machine Learning 2, 17/18

2018-06-02

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