Machine Learning 2 Homework 5

Group: DLBSP

May 23, 2019

Excercise 1

a)

$$
\begin{aligned}
&\ln p(X, Z \big| \theta) \\
&= \sum_{n=1}^{N} \left\{ \ln p(x_n \big| z_n; \theta) + \ln p(z_n \big| \theta)) \right\} \\
&= \sum_{n=1}^{N} \left\{ \ln \mathcal{N}_d(x_n \big| W z_n + \mu, \sigma^2 \mathbb{I}_d) + \ln p(z_n) \right\} \\
&= \sum_{n=1}^{N} \left\{ \ln \mathcal{N}_d(x_n \big| W z_n + \mu, \sigma^2 \mathbb{I}_d) + \ln \mathcal{N}_m(z_n \big| 0, \mathbb{I}_m) \right\} \\
&= -\frac{1}{2} \sum_{n=1}^{N} \left(d \ln 2\pi + \ln |\sigma^2 \mathbb{I}_d| + \operatorname{tr} \left\{ (\sigma^2 \mathbb{I}_d)^{-1} (x_n - W z_n - \mu)(x_n - W z_n - \mu)^\intercal \right\} + m \ln 2\pi + \ln |\mathbb{I}_m| + \operatorname{tr} \left\{ \mathbb{I}_m^{-1} z_n z_n^\intercal \right\} \right) \\
&= -\frac{1}{2} \sum_{n=1}^{N} \left(d \ln 2\pi + d\sigma^2 + \operatorname{tr} \left\{ \frac{1}{\sigma^2} (x_n - W z_n - \mu)(x_n - W z_n - \mu)^\intercal \right\} + m \ln 2\pi + \operatorname{tr} \left\{ z_n z_n^\intercal \right\} \right) \\
&= -\frac{1}{2} \sum_{n=1}^{N} \left((d+m) \ln 2\pi + d\sigma^2 + \operatorname{tr} \left\{ \frac{1}{\sigma^2} (x_n - W z_n - \mu)(x_n - W z_n - \mu)^\intercal + z_n z_n^\intercal \right\} \right)
\end{aligned}
$$

with

$$
\begin{aligned}
&\operatorname{tr} \left\{ (x_n - W z_n - \mu)(x_n - W z_n - \mu)^\intercal \right\} \\
&= \operatorname{tr} \left\{ x_n x_n^\intercal - x_n (W z_n)^\intercal - x_n \mu^\intercal - W z_n x_n^\intercal + W z_n (W z_n)^\intercal + W z_n \mu^\intercal - \mu x_n^\intercal + \mu (W z_n)^\intercal + \mu \mu^\intercal) + z_n z_n^\intercal \right\} \\
&= x_n^\intercal x_n - x_n^\intercal \mu - \mu^\intercal x_n + \mu^\intercal \mu + \operatorname{tr}\{ -x_n (W z_n)^\intercal - W z_n x_n^\intercal + W z_n (W z_n)^\intercal + W z_n \mu^\intercal + \mu (W z_n)^\intercal \} \\
&= (x_n - \mu)^\intercal (x_n - \mu) + \operatorname{tr}\{ -x_n (W z_n)^\intercal - W z_n x_n^\intercal + W z_n (W z_n)^\intercal + W z_n \mu^\intercal + \mu (W z_n)^\intercal \}
\end{aligned}
$$

Excercise 2

a)

Notation comment - $\mathbb{I}$ is an identity matrix.
Given that $p(z) = \mathcal{N}(z|0, \mathbb{I})$ and $p(x|z) = \mathcal{N}(x|Wz + \mu, \sigma^2 \mathbb{I})$ the joint distribution $p(X, Z)$ can be considered as

$$
\begin{aligned}
\ln p(x, z) &= \ln p(x|z) + \ln p(z) \\
= -\frac{1}{2}(x - Wz - \mu)^T \frac{1}{\sigma^2} \mathbb{I} (x - Wz - \mu) &- \frac{1}{2} z^T \mathbb{I} z + \text{const}
\end{aligned}
$$

where const is the term independent of z and x.
Expanding the terms we get

$$
\begin{aligned}
-\frac{1}{2}(z^T(\mathbb{I} + W^T \frac{1}{\sigma^2} \mathbb{I} W) z - x^T \frac{1}{\sigma^2} \mathbb{I} x + x^T \frac{1}{\sigma^2} W z + z^T W^T \frac{1}{\sigma^2} \mathbb{I} x) \\
= -\frac{1}{2} \begin{pmatrix} z \\ x \end{pmatrix}^T \begin{pmatrix} \mathbb{I} + W^T \frac{1}{\sigma^2} \mathbb{I} W & -W^T \frac{1}{\sigma^2} \mathbb{I} \\ -\frac{1}{\sigma^2} \mathbb{I} W & \frac{1}{\sigma^2} \mathbb{I} \end{pmatrix} \begin{pmatrix} z \\ x \end{pmatrix} \\
= -\frac{1}{2} \begin{pmatrix} z \\ x \end{pmatrix}^T R \begin{pmatrix} z \\ x \end{pmatrix}
\end{aligned}
$$

where R is an inverse covariance matrix. Covariance matrix can be found using the matrix inversion formula from Bishop 1995 (2.76)

$$
R = \begin{pmatrix} \mathbb{I} + W^T \frac{1}{\sigma^2} \mathbb{I} W & -W^T \frac{1}{\sigma^2} \mathbb{I} \\ -\frac{1}{\sigma^2} \mathbb{I} W & \frac{1}{\sigma^2} \mathbb{I} \end{pmatrix}
$$

$$R^{-1} = \begin{pmatrix} \mathbb{I} & \mathbb{I}W^T \\ W\mathbb{I} & \sigma^2\mathbb{I} + W\mathbb{I}W^T \end{pmatrix}$$

One can obtain the mean of the joint distribution by neglecting constant terms and considering only linear terms from the above expansion.

$$z^T\mathbb{I}\mu_z - z^TW^T\frac{1}{\sigma^2}\mathbb{I}\mu + x^T\frac{1}{\sigma^2}\mathbb{I}\mu = \begin{pmatrix} z \\ x \end{pmatrix}^T \begin{pmatrix} \mathbb{I}\mu_z - W^T\frac{1}{\sigma^2}\mathbb{I}\mu \\ \frac{1}{\sigma^2}\mathbb{I}\mu \end{pmatrix} = \begin{pmatrix} z \\ x \end{pmatrix}^T \begin{pmatrix} -W^T\frac{1}{\sigma^2}\mathbb{I}\mu \\ \frac{1}{\sigma^2}\mathbb{I}\mu \end{pmatrix}$$

since $\mu_z = 0$

By completing the square over the quadratic form of a multivariate Gaussian, we find that the mean of $p(x, z)$ is given by

$$\mathbb{E}_{x,z} = R^{-1}\begin{pmatrix} -W^T\frac{1}{\sigma^2}\mathbb{I}\mu \\ \frac{1}{\sigma^2}\mathbb{I}\mu \end{pmatrix} = \begin{pmatrix} 0 \\ \mu \end{pmatrix}$$

2.1

Then, using the known formulas (Bishop 1995 (2.73) and (2.75)) for inverse covariance matrix we can obtain

$$\mathbb{E}_{z_n|x_n} = (\mathbb{I} + \frac{1}{\sigma^2}W^TW)^{-1}(\frac{1}{\sigma^2}W^T(x-\mu))$$

and

$$\mathrm{cov}(z_nz_n^T|x_n) = (\mathbb{I} + \frac{1}{\sigma^2}W^TW)^{-1} = \sigma^2(W^TW + \sigma^2\mathbb{I})$$

Using these results we can write expression for $\mathbb{E}[z_nz_n^T|x_n]$.

$$\mathbb{E}[z_nz_n^T|x_n] = \mathrm{cov}(z_n) + \mathbb{E}[z_n]\mathbb{E}[z_n]^T = \sigma^2(W^TW + \sigma^2\mathbb{I}) + \mathbb{E}[z_n]\mathbb{E}[z_n]^T$$

Exercise 3

Some calculation in order to make the E-step function look nicer

$$\begin{aligned}
&\mathbb{E}_{Z\big|X,\theta^t}[\ln p(X, Z\big|\theta)] \\
&= -\frac{1}{2}\sum_{n=1}^{N}\left((d+m)\ln 2\pi + d\sigma^2 + \mathrm{tr}\left\{\frac{1}{\sigma^2}(x_n - W\mathbb{E}[z_n] - \mu)(x_n - W\mathbb{E}[z_n] - \mu)^\intercal + \mathbb{E}[z_nz_n]^\intercal\right\}\right) \\
&= -\frac{1}{2}\sum_{n=1}^{N}\left((d+m)\ln 2\pi + d\sigma^2 + \frac{1}{\sigma^2}(x_n - \mu)^\intercal(x_n - \mu) + \frac{1}{\sigma^2}\mathrm{tr}\left\{-x_n(W\mathbb{E}[z_n])^\intercal - W\mathbb{E}[z_n]x_n^\intercal + \right.\right. \\
&\left.\left. + WW^\intercal\mathbb{E}[z_nz_n]^\intercal] + W\mathbb{E}[z_n]\mu^\intercal + \mu(W\mathbb{E}[z_n])^\intercal\right\}\right) \\
&= -\frac{1}{2}\sum_{n=1}^{N}\left((d+m)\ln 2\pi + d\sigma^2 + \frac{1}{\sigma^2}(x_n - \mu)^\intercal(x_n - \mu) + \frac{1}{\sigma^2}\mathrm{tr}\left\{2W\mathbb{E}[z_n](\mu - x_n)^\intercal + WW^\intercal\mathbb{E}[z_nz_n^\intercal]\right\}\right) \\
&= -\frac{1}{2}\sum_{n=1}^{N}\left((d+m)\ln 2\pi + d\sigma^2 + \frac{1}{\sigma^2}(x_n - \mu)^\intercal(x_n - \mu) + \frac{2}{\sigma^2}(W\mathbb{E}[z_n])^\intercal(\mu - x_n) + \frac{1}{\sigma^2}\mathrm{tr}\left\{WW^\intercal\mathbb{E}[z_nz_n^\intercal]\right\}\right)
\end{aligned}$$

In order to get σ^2_{new} you need to derive the E-step function with respect to σ^2 and set it equal to zero:

$$\begin{aligned}
&\frac{\partial}{\partial(\sigma^2)}\mathbb{E}_{Z\big|X,\theta^t}[\ln p(X, Z\big|\theta)] = 0 \\
&\Rightarrow -\frac{1}{2}\sum_{n=1}^{N}\left(d - \frac{1}{\sigma^4}(x_n - \mu)^\intercal(x_n - \mu) - \frac{2}{\sigma^4}(W\mathbb{E}[z_n])^\intercal(\mu - x_n) - \frac{1}{\sigma^4}\mathrm{tr}\left\{WW^\intercal\mathbb{E}[z_nz_n^\intercal]\right\}\right) = 0 \\
&\Rightarrow \sum_{n=1}^{N}\left(d - \frac{1}{\sigma^4}(x_n - \mu)^\intercal(x_n - \mu) - \frac{2}{\sigma^4}(W\mathbb{E}[z_n])^\intercal(\mu - x_n) - \frac{1}{\sigma^4}\mathrm{tr}\left\{WW^\intercal\mathbb{E}[z_nz_n^\intercal]\right\}\right) = 0 \\
&\Rightarrow Nd\sigma^4 = \sum_{n=1}^{N}((x_n + \mu)^\intercal(x_n - \mu) + 2(W\mathbb{E}[z_n])^\intercal(\mu - x_n)\,\mathrm{tr}\left\{WW^\intercal\mathbb{E}[z_nz_n^\intercal]\right\}) \\
&\Rightarrow \sigma^2_{\text{new}} = \sqrt{\frac{1}{Nd}\sum_{n=1}^{N}((x_n + \mu)^\intercal(x_n - \mu) + 2(W_{\text{new}}\mathbb{E}[z_n])^\intercal(\mu - x_n)\,\mathrm{tr}\left\{W_{\text{new}}W_{\text{new}}^\intercal\mathbb{E}[z_nz_n^\intercal]\right\})}
\end{aligned}$$

Index of comments

1.1 Dimensionality mismatch: Left summand is dxd, right summand is mxm

2.1 Exercise does not state that you may use all these