1. Given the prior $p(z) \sim N(0, I)$ and the posterior approximation $q(z|x;\theta) \sim N(\mu_\theta(x), \sum_\theta(x))$, prove that $KL(q(z|x;\theta)||p(z))$ is tractable; that is, it can be the functions of $\mu_\theta(x)$ and $\sum_\theta(x)$, expressed as a closed-form expression. Both dimensions of multivariate Gaussian are n where mean $\mu_\theta(x)$ and covariance matrix $\sum_\theta(x) = diag(\sigma_1^2, \ldots, \sigma_n^2)$ are functions of x and the parameters θ of a neural network.

$$q(z|x;\theta) = \frac{1}{\sqrt{2\pi^n|\Sigma|}} \exp\left(-\frac{1}{2}(z-\mu_\theta(x))^T \Sigma_\theta^{-1}(x)(z-\mu_\theta(x))\right) = N(\mu_\theta(x), \Sigma_\theta(x))$$

$$p(z) = \frac{1}{\sqrt{2\pi^n|\Sigma|}} \exp\left(-\frac{1}{2}(z-0)^T I^{-1}(z-0)\right) = N(0, I)$$

$$KL(q(z|x;\theta)||p(z)) = \int q(z|x;\theta) \log \frac{q(z|x;\theta)}{p(z)} dz$$

$$= \int q(z|x;\theta)\left[\log q(z|x;\theta) - \log p(z)\right] dz$$

$$= \int q(z|x;\theta)\left[-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|\Sigma|) - \frac{1}{2\Sigma}(z-\mu)^T(z-\mu) + \frac{n}{2}\log(2\pi) - \frac{1}{2}\log(|I|) + \frac{1}{2I}(z)^T(z)\right] dz$$

$$= \int q(z|x;\theta)\left[\frac{1}{2}\log\frac{|I|}{|\Sigma|} + \frac{1}{2}\left(\frac{z^Tz}{I} - \frac{(z-\mu)^T(z-\mu)}{\Sigma}\right)\right] dz$$

$$= E_q\left[\frac{1}{2}\log\frac{|I|}{|\Sigma|} + \frac{1}{2}\left(\frac{z^Tz}{I} - \frac{(z-\mu)^T(z-\mu)}{\Sigma}\right)\right]$$

$$= \frac{1}{2}\log\frac{|I|}{|\Sigma|} + \frac{1}{2I}E_q[z^Tz] - \frac{1}{2\Sigma}E_q[(z-\mu)^T(z-\mu)]$$

$$= \frac{1}{2}\log\frac{|I|}{|\Sigma|} + \frac{1}{2I}E_q[z^Tz] - \frac{1}{2\Sigma}\Sigma$$

$$= \frac{1}{2}\log\frac{|I|}{|\Sigma|} + \frac{1}{2I}E_q[(z-\mu+\mu+0)^T(z-\mu+\mu+0)] - \frac{1}{2}tr(I_n)$$

$$= \frac{1}{2}\log\frac{|I|}{|\Sigma|} + \frac{1}{2I}\left[E_q[(z-\mu)^T(z-\mu)] + 2\mu \underbrace{E_q[z-\mu]}_{=0} + \mu^T\mu\right] - \frac{1}{2}tr(I_n)$$

$$= \frac{1}{2}\log\frac{|I|}{|\Sigma|} + \frac{1}{2I}\left[\Sigma + \mu^T\mu\right] - \frac{1}{2}n$$

$$= \frac{1}{2}\left\{\log\frac{|I|}{|\Sigma|} + tr(I^{-1}\Sigma) + \mu^T I^{-1}\mu - n\right]$$

$\therefore KL(q(z|x;\theta)||p(z))$ 與 $\mu_\theta(x)$、$\Sigma_\theta(x)$ 有关