- $d \in \mathbb{N}^*$ (latent space dimension)
- $\mu \in \mathbb{R}^d$
- $\forall i \in [\![1, d]\!]$, $\sigma_i \in \mathbb{R}_+^*$
- $\Sigma = \operatorname{diag}(\sigma_1^2, \dots, \sigma_d^2)$
- $P \sim \mathcal{N}(\mu, \Sigma)$ (latent space distribution)
- $Q \sim \mathcal{N}(0_d, I_d)$ (target distribution)
- We denote by p (resp. q) the density associated with P (resp. Q)

$$D_{KL}(P \| Q) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{p(x)}{q(x)}\right) dx = \mathbb{E}_p\left\{\log\left(\frac{p(x)}{q(x)}\right)\right\}$$

$$= \mathbb{E}_p\left\{\log\left(\frac{\frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right]}{\frac{1}{(2\pi)^{d/2}|I_d|^{1/2}} \exp-\left[\frac{1}{2}(x-0_d)^t I_d^{-1}(x-0_d)\right]}\right)\right\}$$

$$= \mathbb{E}_p\left\{\log\left(\frac{|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right]}{\exp\left[-\frac{1}{2}x^t x\right]}\right)\right\}$$

$$= \mathbb{E}_p\left\{\log\left(|\Sigma|^{-1/2} \exp\left[-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu) + \frac{1}{2}x^t x\right]\right)\right\}$$

$$= \mathbb{E}_p\left\{-\frac{1}{2}\log(|\Sigma|) - \frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu) + \frac{1}{2}x^t x\right\}$$

$$= \frac{1}{2}(-\alpha - \beta + \gamma)$$

where $\alpha = \mathbb{E}_p\{\log(|\Sigma|)\}$, $\beta = \mathbb{E}_p\{(x-\mu)^t \Sigma^{-1}(x-\mu)\}$

and $\gamma = \mathbb{E}_p\{x^t x\}$.

$\cdot\ \alpha = \mathbb{E}_p\{\log(|\Sigma|)\} = \log(|\Sigma|)$ because $\log(|\Sigma|)$ is a constant.

$\alpha = \log(|\Sigma|) = \log\left(\prod_{i=1}^{d} \sigma_i^2\right)$ because Σ is diagonal.

$= \sum_{i=1}^{d} \log(\sigma_i^2)$.

$$\cdot\ \beta = \mathbb{E}_p\{\underbrace{(x-\mu)^t \Sigma^{-1}(x-\mu)}_{\in \mathbb{R}}\} = \mathbb{E}_p\{tr[(x-\mu)^t \Sigma^{-1}(x-\mu)]\}$$

$= \mathbb{E}_p\{tr[\underbrace{(x-\mu)(x-\mu)^t \Sigma^{-1}}_{=M}]\}$ because $tr(ABC) = tr(CAB)$

$$= \mathbb{E}_p\left\{\sum_{i=1}^{d} (M)_{i,i}\right\} = \sum_{i=1}^{d}\left(\mathbb{E}_p\{(M)_{i,i}\}\right)$$

$$= tr\begin{pmatrix} \mathbb{E}_p\{(M)_{1,1}\} & \mathbb{E}_p\{(M)_{1,2}\} & \cdots & \\ \mathbb{E}_p\{(M)_{2,1}\} & \mathbb{E}_p\{(M)_{2,2}\} & & \\ \vdots & & \ddots & \\ \vdots & & & \mathbb{E}_p\{(M)_{d,d}\} \end{pmatrix}$$

$$= tr[\mathbb{E}_p\{M\}] = tr[\mathbb{E}_p\{(x-\mu)(x-\mu)^t \Sigma^{-1}\}]$$

$= tr[\mathbb{E}_p\{(x-\mu)(x-\mu)^t\}\Sigma^{-1}]$ because Σ^{-1} is a constant.

$$= tr[Var(P)\Sigma^{-1}] = tr[\Sigma\Sigma^{-1}] = tr[I_d] = d.$$

$$\cdot\ \gamma = \mathbb{E}_p\{\overbrace{x^t x}^{\in \mathbb{R}}\} = \mathbb{E}_p\{tr[x^t x]\} = \mathbb{E}_p\{tr[\overbrace{x x^t}^{=N}]\}$$

$$= \mathbb{E}_p\left\{\sum_{i=1}^{d} (N)_{i,i}\right\} = \sum_{i=1}^{d}\left(\mathbb{E}_p\{(N)_{i,i}\}\right) = tr[\mathbb{E}_p\{N\}]$$

$$= tr\left[\mathbb{E}_p\{xx^t\}\right] = tr\left[\mathbb{E}_p\left\{xx^t + (\mu\mu^t - x\mu^t - \mu x^t) - (\mu\mu^t - x\mu^t - \mu x^t)\right\}\right]$$

$$= tr\left[\mathbb{E}_p\left\{(x-\mu)(x-\mu)^t - (\mu\mu^t - x\mu^t - \mu x^t)\right\}\right]$$

$$= tr\left[\mathbb{E}_p\left\{(x-\mu)(x-\mu)^t\right\} - \mathbb{E}_p\{\mu\mu^t\} + \mathbb{E}_p\{x\mu^t\} + \mathbb{E}_p\{\mu x^t\}\right]$$

$$= tr\left[Var(P) - \mu\mu^t + \mathbb{E}_p\{x\}\mu^t + \mu(\mathbb{E}_p\{x\})^t\right]$$ because μ is a constant

$$= tr\left[\Sigma - \mu\mu^t + \mu\mu^t + \mu\mu^t\right] = tr\left[\Sigma + \mu\mu^t\right] = tr[\Sigma] + tr[\mu\mu^t]$$

$$= \sum_{i=1}^{d} \sigma_i^2 + \sum_{i=1}^{d} \mu_i^2 = \sum_{i=1}^{d}(\mu_i^2 + \sigma_i^2)$$

$$D_{KL}(P\|Q) = \frac{1}{2}(-\alpha - \beta + \gamma) = \frac{1}{2}\left[-\sum_{i=1}^{d}\left(\log(\sigma_i^2)\right) - d + \sum_{i=1}^{d}(\mu_i^2 + \sigma_i^2)\right]$$

$$= \frac{1}{2}\sum_{i=1}^{d}\left[\mu_i^2 + \sigma_i^2 - 1 - \log(\sigma_i^2)\right]$$