Exercise sheet 6

Dienstag, 7. Juni 2022 12:41

Lars Kühmichel
Nicolas Wolf

Task 1. Kullback-Leibler divergence of two multivariate normal distributions
The Kullback-Leibler divergence of two continuous distributions $\mathcal{P}$ and $\mathcal{Q}$ is defined as:

$$KL(P \parallel Q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)}\, dx.$$

Compute it analytically in the case where $\mathcal{P}$ and $\mathcal{Q}$ are two multivariate normal distributions $\mathcal{P} = \mathcal{N}(\mu_1, \Sigma_1)$ and $\mathcal{Q} = \mathcal{N}(\mu_2, \Sigma_2)$: Derive the Kullback-Leibler-divergence for two distinct normal distribution.

1

$$p,q(x) = \frac{1}{\sqrt{(2\pi)^M |\Sigma_{1,2}|}} \exp\left[-\frac{1}{2}(x-\mu_{1,2})^T \Sigma_{1,2}^{-1} (x-\mu_{1,2})\right] \quad \text{with } x, \mu_1, \mu_2 \in \mathbb{R}^M,\ \Sigma_{1,2} \in \mathbb{R}^{M\times M}$$

$$KL[P \| Q] = \int p(x)\left[\log p(x) - \log q(x)\right] dx$$

$$= \int p(x)\Big[\cancel{-\frac{M}{2}\log 2\pi} - \frac{1}{2}\log|\Sigma_1| - \frac{1}{2}(x-\mu_1)^T \Sigma_1^{-1}(x-\mu_1)$$

$$+ \cancel{\frac{M}{2}\log 2\pi} + \frac{1}{2}\log|\Sigma_2| + \frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\Big]\, dx$$

$x^T A y = tr(A y x^T)$

$$= \int p(x)\left[\frac{1}{2}\log\frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2}(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\right] dx$$

$$= \frac{1}{2}\log\frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2}\int p(x)(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)\,dx + \frac{1}{2}\int p(x)(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\,dx$$

$$= \frac{1}{2}\log\frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2}\int p(x)\, tr\left[\Sigma_1^{-1}(x-\mu_1)(x-\mu_1)^T\right] dx + \frac{1}{2}\int p(x)(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\,dx$$

$$= \frac{1}{2}\log\frac{|\Sigma_2|}{|\Sigma_1|} - \frac{1}{2}\, tr\left(\Sigma_1^{-1}\,\mathbb{E}_p\left[(x-\mu_1)(x-\mu_1)^T\right]\right) + \frac{1}{2}\,\mathbb{E}_p\left[(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\right]$$

- $\mathbb{E}_p\left[(x-\mu_1)(x-\mu_1)^T\right] = \Sigma_1$
- $\mathbb{E}_p\left[(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)\right] \overset{(380)}{=} (\mu_1-\mu_2)^T\Sigma_2^{-1}(\mu_1-\mu_2) + tr(\Sigma_2^{-1}\Sigma_1)$ Matrix Cookbook

$$= \frac{1}{2}\left[\log\frac{|\Sigma_2|}{|\Sigma_1|} - \underbrace{tr(\Sigma_1^{-1}\cdot\Sigma_1)}_{=tr(\mathbb{1}_M) = M} + (\mu_1-\mu_2)^T\Sigma_2^{-1}(\mu_1-\mu_2) + \frac{1}{2}tr(\Sigma_2^{-1}\Sigma_1)\right]$$

$$= \frac{1}{2}\left[\log\frac{|\Sigma_2|}{|\Sigma_1|} - M + (\mu_1-\mu_2)^T\Sigma_2^{-1}(\mu_1-\mu_2) + tr(\Sigma_2^{-1}\Sigma_1)\right]$$