

We decided this derivation is too much work for a small part of an assignment. We provide the original question and the solution in this document.

This is a very good exercise to check your understanding of integration and Analysis, so we encourage you to try it yourself.

Todo: Compute the Entropy of the normal distribution.

$$\mathcal{N}(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right).$$

Solution:

The entropy of a Gaussian can be computed without performing any integrals directly:

$$\begin{aligned} H[\mathcal{N}] &= - \int_{-\infty}^{\infty} \log(N(x; \mu, \sigma^2)) \cdot N(x; \mu, \sigma^2) dx \\ &= - \int_{-\infty}^{\infty} \left[-\frac{(x - \mu)^2}{2\sigma^2} - \frac{1}{2} \log(2\pi\sigma^2) \right] \cdot N(x; \mu, \sigma^2) dx \\ &= \frac{1}{2\sigma^2} \underbrace{\int_{-\infty}^{\infty} (x - \mu)^2 \cdot N(x; \mu, \sigma^2) dx}_{=\sigma^2} + \frac{1}{2} \log(2\pi\sigma^2) \underbrace{\int_{-\infty}^{\infty} N(x; \mu, \sigma^2) dx}_{=1} \\ &= \frac{1}{2} + \frac{1}{2} \log(2\pi\sigma^2) = \frac{1}{2} \log(2\pi e\sigma^2) \end{aligned}$$

Todo: Find the Kullback-Leibler divergence between two Gaussian distributions.

$$D_{KL}(\mathcal{N}(\mu_1, \sigma_1^2) || \mathcal{N}(\mu_2, \sigma_2^2)) .$$

Solution:

Let's first simplify the expression for the Kullback-Leibler divergence, by introducing the notation:

$\mathcal{N}_1(x) = \mathcal{N}(x; \mu_1, \sigma_1^2)$ and $\mathcal{N}_2(x) = \mathcal{N}(x; \mu_2, \sigma_2^2)$:

$$\begin{aligned}
 D_{KL}(\mathcal{N}_1 || \mathcal{N}_2) &= \int_{-\infty}^{\infty} \mathcal{N}_1(x) \cdot \log \left(\frac{\mathcal{N}_1(x)}{\mathcal{N}_2(x)} \right) dx \\
 &= -H[\mathcal{N}_1] - \int_{-\infty}^{\infty} \log(\mathcal{N}_2(x)) \mathcal{N}_1(x) dx \\
 &= -\frac{\log(2\pi e \sigma_1^2)}{2} - \int_{-\infty}^{\infty} \left(-\frac{\log(2\pi \sigma_2^2)}{2} - \frac{(x - \mu_2)^2}{2\sigma_2^2} \right) \cdot \mathcal{N}_1(x) dx \\
 &= -\frac{\log(2\pi e \sigma_1^2)}{2} + \frac{\log(2\pi \sigma_2^2)}{2} + \int_{-\infty}^{\infty} \frac{(x - \mu_2)^2}{2\sigma_2^2} \cdot \mathcal{N}_1(x) dx \\
 &= \log \left(\frac{\sigma_2}{\sigma_1} \right) - \frac{1}{2} + \int_{-\infty}^{\infty} \frac{(x - \mu_2)^2}{2\sigma_2^2} \cdot \mathcal{N}_1(x) dx \tag{1}
 \end{aligned}$$

For space reasons, we continue to compute only the last integral

$$\begin{aligned}
 \int_{-\infty}^{\infty} (x - \mu_2)^2 \cdot \mathcal{N}(x; \mu_1, \sigma_1^2) dx &= \int_{-\infty}^{\infty} \underbrace{(x - \mu_1 + \mu_1 - \mu_2)^2}_{= (x - \mu_1)^2 + (\mu_1 - \mu_2)^2 + 2(x - \mu_1)(\mu_1 - \mu_2)} \cdot \mathcal{N}(x; \mu_1, \sigma_1^2) dx \\
 &= \underbrace{\int_{-\infty}^{\infty} (x - \mu_1)^2 \mathcal{N}(x; \mu_1, \sigma_1^2) dx}_{= \sigma_1^2} \\
 &\quad + (\mu_1 - \mu_2)^2 \underbrace{\int_{-\infty}^{\infty} \mathcal{N}(x; \mu_1, \sigma_1^2) dx}_{= 1} \\
 &\quad + 2(\mu_1 - \mu_2) \underbrace{\int_{-\infty}^{\infty} (x - \mu_1) \mathcal{N}(x; \mu_1, \sigma_1^2) dx}_{= 0} \\
 &= \sigma_1^2 + (\mu_1 - \mu_2)^2
 \end{aligned}$$

Plugging this back into the previous equation yields

$$D_{KL}(\mathcal{N}(\mu_1, \sigma_1^2) || \mathcal{N}(\mu_2, \sigma_2^2)) = \log \left(\frac{\sigma_2}{\sigma_1} \right) - \frac{1}{2} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2}.$$

Note that the divergence is 0, if $\mu_1 = \mu_2$ and $\sigma_1 = \sigma_2$.